Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 22 (2020), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Technology Acceptance in Mobile Health: Scoping Review of Definitions, Models, and Measurement (e17256)
Camille Nadal, Corina Sas, Gavin Doherty. . . . 13

The Acceptability and Usability of Digital Health Interventions for Adults With Depression, Anxiety, and Somatoform Disorders: Qualitative Systematic Review and Meta-Synthesis (e16228)
Shireen Patel, Athfah Akhtar, Sam Malins, Nicola Wright, Emma Rowley, Emma Young, Stephanie Sampson, Richard Morriss. . . . 30

Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review (e18480)
Dillys Larbi, Pietro Randine, Eirik Årsand, Konstantinos Antypas, Meghan Bradway, Elia Gabarron. . . . 45

Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials (e18386)
Hengfen Dai, Caiyun Zheng, Chun Lin, Yan Zhang, Hong Zhang, Fan Chen, Yunchun Liu, Jingwen Xiao, Chaoxin Chen. . . . 58

Digital Self-Management Interventions for People With Osteoarthritis: Systematic Review With Meta-Analysis (e15365)
Reza Safari, Jessica Jackson, David Sheffield. . . . 71

Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review (e18477)
Siqi Liu, Kay See, Kee Ngiam, Leo Celi, Xingzhi Sun, Mengling Feng. . . . 95

Implementing Web-Based Therapy in Routine Mental Health Care: Systematic Review of Health Professionals' Perspectives (e17362)
Fiona Davies, Heather Shepherd, Lisa Beatty, Brodie Clark, Phyllis Butow, Joanne Shaw. . . . 111

Information Needs of Breast Cancer Patients: Theory-Generating Meta-Synthesis (e17907)
Hongru Lu, Juan Xie, Lynette Gerido, Ying Cheng, Ya Chen, Lizhu Sun. . . . 144

Evidence Regarding Automatic Processing Computerized Tasks Designed For Health Interventions in Real-World Settings Among Adults: Systematic Scoping Review (e17915)
Harshani Jayasinghe, Camille Short, Annette Braunack-Mayer, Ashley Merkin, Clare Hume. . . . 162

Design, Use, and Effects of Sex Dolls and Sex Robots: Scoping Review (e18551)
Nicola Döring, M Mohseni, Roberto Walter. . . . 174

XSL•FO
RenderX

Computer-Controlled Virtual Humans in Patient-Facing Systems: Systematic Review and Meta-Analysis (e18839)
Debaleena Chattopadhyay, Tengteng Ma, Hasti Sharifi, Pamela Martyn-Nemeth. . . . 202

Effectiveness of Technology-Enabled Knowledge Translation Strategies in Improving the Use of Research in Public Health: Systematic Review (e17274)
Alison Brown, Courtney Barnes, Judith Byaruhanga, Matthew McLaughlin, Rebecca Hodder, Debbie Booth, Nicole Nathan, Rachel Sutherland, Luke Wolfenden. . . . 223

Digital Behavior Change Interventions for Younger Children With Chronic Health Conditions: Systematic Review (e16924)
Amberly Brigden, Emma Anderson, Catherine Linney, Richard Morris, Roxanne Parslow, Teona Serafimova, Lucie Smith, Emily Briggs, Maria Loades, Esther Crawley. . . . 239

Implementation Strategies for Web-Based Apps for Screening: Scoping Review (e15591)
Chor Ooi, Chirk Ng, Anne Sales, Hooi Lim. . . . 468

Assessing the Credibility and Authenticity of Social Media Content for Applications in Health Communication: Scoping Review (e17296)
Eva Jenkins, Jasmina Ilicic, Amy Barklamb, Tracy McCaffrey. . . . 756

Requirements of Health Data Management Systems for Biomedical Care and Research: Scoping Review (e17508)
Leila Ismail, Huned Materwala, Achim Karduck, Abdu Adem. . . . 1621

The Use of Social Media to Increase the Impact of Health Research: Systematic Review (e15607)
Marco Bardus, Rola El Rassi, Mohamad Chahrour, Elie Akl, Abdul Raslan, Lokman Meho, Elie Akl. . . . 1882

Original Papers

Using Digital Communication Technology to Increase HIV Testing Among Men Who Have Sex With Men and Transgender Women: Systematic Review and Meta-Analysis (e14230)
Vanessa Veronese, Kathleen Ryan, Chad Hughes, Megan Lim, Alisa Pedrana, Mark Stoové. . . . 123

Predictors of Self-Determined Module Choice in a Web-Based Computer-Tailored Diet and Physical Activity Intervention: Secondary Analysis of Data From a Randomized Controlled Trial (e15024)
Juul Coumans, Catherine Bolman, Anke Oenema, Lilian Lechner. . . . 391

Use of an Unguided, Web-Based Distress Self-Management Program After Breast Cancer Diagnosis: Sub-Analysis of CaringGuidance Pilot Study (e19734)
Robin Lally, Kevin Kupzyk, Steve Gallo, Donna Berry. . . . 406

Adherence to Established Treatment Guidelines Among Unguided Digital Interventions for Depression: Quality Evaluation of 28 Web-Based Programs and Mobile Apps (e16136)
Stefan Bubolz, Gwendolyn Mayer, Nadine Gronewold, Thomas Hilbel, Jobst-Hendrik Schultz. . . . 419

Effects of the e-Motivate4Change Program on Metabolic Syndrome in Young Adults Using Health Apps and Wearable Devices: Quasi-Experimental Study (e17031)
Ji-Soo Lee, Min-Ah Kang, Soo-Kyoung Lee. . . . 436

XSL•FO
RenderX

Dose-Response Relationship of a Web-Based Tailored Intervention Promoting Human Papillomavirus Vaccination: Process Evaluation of a Randomized Controlled Trial (e14822)
Mirjam Pot, Theo Paulussen, Robert Ruiter, Liesbeth Mollema, Miranda Hofstra, Hilde Van Keulen. 450

Internet-Based Self-Management Support After High-Altitude Climate Treatment for Severe Asthma: Randomized Controlled Trial (e13145)
Thijs Beerthuizen, Lucia Rijssenbeek-Nouwens, Sophia van Koppen, Rishi Khusial, Jiska Snoeck-Stroband, Jacob Sont. 479

Novel Stepped Care Approach to Provide Education and Exercise Therapy for Patellofemoral Pain: Feasibility Study (e18584)
Danilo De Oliveira Silva, Marcella Pazzinatto, Kay Crossley, Fabio Azevedo, Christian Barton. 494

Scheduled Telephone Support for Internet Cognitive Behavioral Therapy for Depression in Patients at Risk for Dropout: Pragmatic Randomized Controlled Trial (e15732)
Satu Pihlaja, Jari Lahti, Jari Lipsanen, Ville Ritola, Eero-Matti Gummerus, Jan-Henry Stenberg, Grigori Joffe. 507

Evaluation of a Blended Physical Activity Intervention for Older Adults: Mixed Methods Study (e16380)
Sumit Mehra, Jantine van den Helder, Bart Visser, Raoul Engelbert, Peter Weijs, Ben Kröse. 519

Evaluation of a Web-Based Self-Management Program for Patients With Cardiovascular Disease: Explorative Randomized Controlled Trial (e17422)
Marscha Engelen, Sandra van Dulmen, Saskia Puijk-Hekman, Hester Vermeulen, Maria Nijhuis-van der Sanden, Sebastian Bredie, Betsie van Gaal. 532

An mHealth Intervention (ReZone) to Help Young People Self-Manage Overwhelming Feelings: Cluster-Randomized Controlled Trial (e14223)
Chloe Edridge, Miranda Wolpert, Jessica Deighton, Julian Edbrooke-Childs. 546

Impact of an Electronic Health Service on Child Participation in Pediatric Oncology Care: Quasiexperimental Study (e17673)
Britt-Mari Gilljam, Jens Nygren, Petra Svedberg, Susann Arvidsson. 552

Teens Taking Charge: A Randomized Controlled Trial of a Web-Based Self-Management Program With Telephone Support for Adolescents With Juvenile Idiopathic Arthritis (e16234)
Jennifer Stinson, Chitra Lalloo, Amos Hundert, Sarah Campillo, Tania Cellucci, Paul Dancey, Ciaran Duffy, Janet Ellsworth, Brian Feldman, Adam Huber, Nicole Johnson, Geert't Jong, Kiem Oen, Alan Rosenberg, Natalie Shiff, Lynn Spiegel, Shirley Tse, Lori Tucker, Joseph Victor. 568

Designing a Personalized Digital Patient Support Program for Patients Treated With Growth Hormone: Key Design Considerations (e18157)
Sumaira Malik, Clare Moloney, Ekaterina Koledova, Jonathan Reston, John Weinman. 584

A Mobile Health Approach for Improving Outcomes in Suicide Prevention (SafePlan) (e17481)
Conor O'Grady, Ruth Melia, John Bogue, Mary O'Sullivan, Karen Young, Jim Duggan. 593

Participatory Interventions for Sexual Health Promotion for Adolescents and Young Adults on the Internet: Systematic Review (e15378)
Philippe Martin, Lorraine Cousin, Serge Gottot, Aurelie Bourmaud, Elise de La Rochebrochard, Corinne Alberti. 607

Perceived Effectiveness and Motivations for the Use of Web-Based Mental Health Programs: Qualitative Study (e16961)
Heidi Eccles, Molly Nannarone, Bonnie Lashewicz, Mark Attridge, Alain Marchand, Alice Aiken, Kendall Ho, Jianli Wang. 624

A Virtual Reality–Supported Intervention for Pulmonary Rehabilitation of Patients With Chronic Obstructive Pulmonary Disease: Mixed Methods Study (e14178)
Timothy Jung, Natasha Moorhouse, Xin Shi, Muhammad Amin. 637

XSL•FO
RenderX

A Dietary Mobile App for Patients Undergoing Hemodialysis: Prospective Pilot Study to Improve Dietary Intakes (e17817)
Cosette Fakih El Khoury, Rik Crutzen, Jos Schols, Ruud Halfens, Mirey Karavetian. 650

Experiences With Wearable Activity Data During Self-Care by Chronic Heart Patients: Qualitative Study (e15873)
Tariq Andersen, Henriette Langstrup, Stine Lomborg. 661

Perceptions and Acceptability of Digital Interventions Among Tuberculosis Patients in Cambodia: Qualitative Study of Video-Based Directly Observed Therapy (e16856)
Lila Rabinovich, James Molton, Wei Ooi, Nicholas Paton, Shelly Batra, Joanne Yoong. 675

Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis (e17758)
Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Velazquez, Josep Gonfaus, Jordi Gonzàlez. 685

An Instagram-Based Study to Understand Betel Nut Use Culture in Micronesia: Exploratory Content Analysis (e13954)
Wayne Buente, Francis Dalisay, Pallav Pokhrel, Hanae Kramer, Ian Pagano. 701

Characterizing Social Media Messages Related to Underage JUUL E-Cigarette Buying and Selling: Cross-Sectional Analysis of Reddit Subreddits (e16962)
Hejing Liu, Qiudan Li, Yongcheng Zhan, Zhu Zhang, Daniel Zeng, Scott Leischow. 714

The Development of a Smart Health Awareness Message Framework Based on the Use of Social Media: Quantitative Study (e16212)
Elaf Alsisi, Ahmed Al-Ashaab, Wadhah Abualfaraa. 727

Sudden Infant Death Syndrome on Facebook: Qualitative Descriptive Content Analysis to Guide Prevention Efforts (e18474)
Kelly Pretorius, Eunju Choi, Sookja Kang, Michael Mackert. 744

Influence of Social Media on Sexualized Drug Use and Chemsex Among Chinese Men Who Have Sex With Men: Observational Prospective Cohort Study (e17894)
Zixin Wang, Xue Yang, Phoenix Mo, Yuan Fang, Tsun Ip, Joseph Lau. 771

Blended Care-Cognitive Behavioral Therapy for Depression and Anxiety in Real-World Settings: Pragmatic Retrospective Study (e18723)
Anita Lungu, Janie Jun, Okhtay Azarmanesh, Yan Leykin, Connie Chen. 790

Effectiveness of a Transdiagnostic Guided Internet-Delivered Protocol for Emotional Disorders Versus Treatment as Usual in Specialized Care: Randomized Controlled Trial (e18220)
Alberto González-Robles, Amanda Díaz-García, Azucena García-Palacios, Pablo Roca, Josep Ramos-Quiroga, Cristina Botella. 802

Habits and Attitudes of Video Gaming and Information Technology Use in People with Schizophrenia: Cross-Sectional Survey (e14865)
William Choi, Dan Yu, Terry Wong, Tella Lantta, Min Yang, Maritta Välimäki. 825

Mobile Mental Health Apps in China: Systematic App Store Search (e14915)
Huifang Yin, Klaas Wardenaar, Yuhao Wang, Nan Wang, Wenjin Chen, Yan Zhang, Guangming Xu, Robert Schoevers. 843

XSL•FO
RenderX

Surveilling Influenza Incidence With Centers for Disease Control and Prevention Web Traffic Data: Demonstration Using a Novel Dataset (e14337)
Wendy Caldwell, Geoffrey Fairchild, Sara Del Valle. 856

Evaluation of Volume of News Reporting and Opioid-Related Deaths in the United States: Comparative Analysis Study of Geographic and Socioeconomic Differences (e17693)
Yulin Hswen, Amanda Zhang, Clark Freifeld, John Brownstein. 871

Eating Disorder Awareness Campaigns: Thematic and Quantitative Analysis Using Twitter (e17626)
Iranzu Viguria, Miguel Alvarez-Mon, Maria Llavero-Valero, Angel Asunsolo del Barco, Felipe Ortuño, Melchor Alvarez-Mon. 879

Conversations and Misconceptions About Chemotherapy in Arabic Tweets: Content Analysis (e13979)
Abdulrahman Alghamdi, Khalid Abumelha, Jawad Allarakia, Ahmed Al-Shehri. 890

Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study (e17633)
Muhammad Syamsuddin, Muhammad Fakhruddin, Jane Sahetapy-Engel, Edy Soewono. 899

Investigation of Geographic and Macrolevel Variations in LGBTQ Patient Experiences: Longitudinal Social Media Analysis (e17087)
Yulin Hswen, Amanda Zhang, Kara Sewalk, Gaurav Tuli, John Brownstein, Jared Hawkins. 909

The Effects of Telemonitoring on Patient Compliance With Self-Management Recommendations and Outcomes of the Innovative Telemonitoring Enhanced Care Program for Chronic Heart Failure: Randomized Controlled Trial (e17559)
Hang Ding, Rajiv Jayasena, Sheau Chen, Andrew Maiorana, Alison Dowling, Jamie Layland, Norm Good, Mohanraj Karunanithi, Iain Edwards.
9 1 8

Wearable Motion Sensor Device to Facilitate Rehabilitation in Patients With Shoulder Adhesive Capsulitis: Pilot Study to Assess Feasibility (e17032)
Yu-Pin Chen, Chung-Ying Lin, Ming-Jr Tsai, Tai-Yuan Chuang, Oscar Lee. 930

Health Care Professionals' Views on Using Remote Measurement Technology in Managing Central Nervous System Disorders: Qualitative Interview Study (e17414)
Jacob Andrews, Michael Craven, Jennifer Jamnadas-Khoda, Alexandra Lang, Richard Morriss, Chris Hollis, The RADAR-CNS Consortium. .
9 4 5

Comprehensive Telestroke Network to Optimize Health Care Delivery for Cerebrovascular Diseases: Algorithm Development (e18058)
Hernán Bayona, Brenda Ropero, Antonio Salazar, Juan Pérez, Manuel Granja, Carlos Martínez, Juan Useche. 957

The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis (e17995)
Yangyang Han, Reidar Lie, Rui Guo. 976

Technology-Enabled Self-Management of Chronic Obstructive Pulmonary Disease With or Without Asynchronous Remote Monitoring: Randomized Controlled Trial (e18598)
Vess Stamenova, Kyle Liang, Rebecca Yang, Katrina Engel, Florence van Lieshout, Elizabeth Lalingo, Angelica Cheung, Adam Erwood, Maria Radina, Allen Greenwald, Payal Agarwal, Aman Sidhu, R Bhatia, James Shaw, Roshan Shafai, Onil Bhattacharyya. 985

Clinical Characteristics of an Internet-Based Cohort of Patient-Reported Diagnosis of Granulomatosis With Polyangiitis and Microscopic Polyangiitis: Observational Study (e17231)
Jason Springer, Tanaz Kermani, Antoine Sreih, Dianne Shaw, Kalen Young, Cristina Burroughs, Peter Merkel. 999

Worldwide Prevalence of Hearing Loss Among Smartphone Users: Cross-Sectional Study Using a Mobile-Based App (e17238)
Marcin Masalski, Krzysztof Morawski. 1008

XSL•FO
RenderX

Digital Divide in Perceived Benefits of Online Health Care and Social Welfare Services: National Cross-Sectional Survey Study (e17616)
Tarja Heponiemi, Vesa Jormanainen, Lars Leemann, Kristiina Manderbacka, Anna-Mari Aalto, Hannele Hyppönen. 1021

Relationships Between the Usage of Televisions, Computers, and Mobile Phones and the Quality of Sleep in a Chinese Population: Community-Based Cross-Sectional Study (e18095)
Yao Xie, Daphne Cheung, Alice Loke, Bernice Nogueira, Karry Liu, Angela Leung, Alice Tsang, Cindy Leong, Alex Molassiotis. 1033

Exploring the Relationship Between Internet Use and Mental Health Among Older Adults in England: Longitudinal Observational Study (e15683)
Sabrina Lam, Stephen Jivraj, Shaun Scholes. 1046

Understanding the Dimensions of Medical Crowdfunding: A Visual Analytics Approach (e18813)
Jie Ren, Viju Raghupathi, Wullianallur Raghupathi. 1059

The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study (e17451)
Edmund Lee, Mesfin Bekalu, Rachel McCloud, Donna Vallone, Monisha Arya, Nathaniel Osgood, Xiaoyan Li, Sara Minsky, Kasisomayajula Viswanath. 1078

Diagnosing Parkinson Disease Through Facial Expression Recognition: Video Analysis (e18697)
Bo Jin, Yue Qu, Liang Zhang, Zhan Gao. 1090

Characterization and Correction of Bias Due to Nonparticipation and the Degree of Loyalty in Large-Scale Finnish Loyalty Card Data on Grocery Purchases: Cohort Study (e18059)
Anna-Leena Vuorinen, Maijaliisa Erkkola, Mikael Fogelholm, Satu Kinnunen, Hannu Saarijärvi, Liisa Uusitalo, Turkka Näppilä, Jaakko Nevalainen. 1102

Ontological Organization and Bioinformatic Analysis of Adverse Drug Reactions From Package Inserts: Development and Usability Study (e20443)
Xiaoying Li, Xin Lin, Huiling Ren, Jinjing Guo. 1117

QardioArm Blood Pressure Monitoring in a Population With Type 2 Diabetes: Validation Study (e19781)
Victoria Mazoteras-Pardo, Ricardo Becerro-De-Bengoa-Vallejo, Marta Losa-Iglesias, Eva Martínez-Jiménez, César Calvo-Lobo, Carlos Romero-Morales, Daniel López-López, Patricia Palomo-López. 1129

Understanding the Acceptability of Subdermal Implants as a Possible New HIV Prevention Method: Multi-Stage Mixed Methods Study (e16904)
Christine Rael, Cody Lentz, Alex Carballo-Diéguez, Rebecca Giguere, Curtis Dolezal, Daniel Feller, Richard D'Aquila, Thomas Hope. 1139

A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open MS BioScreen): Human-Centered Design and Development (e15605)
Erica Schleimer, Jennifer Pearce, Andrew Barnecut, William Rowles, Antoine Lizee, Arno Klein, Valerie Block, Adam Santaniello, Adam Renschen, Refujia Gomez, Anisha Keshavan, Jeffrey Gelfand, Roland Henry, Stephen Hauser, Riley Bove. 1153

The Promise of Patient Portals for Individuals Living With Chronic Illness: Qualitative Study Identifying Pathways of Patient Engagement (e17744)
Maureen Stewart, Timothy Hogan, Jeff Nicklas, Stephanie Robinson, Carolyn Purington, Christopher Miller, Varsha Vimalananda, Samantha Connolly, Hill Wolfe, Kim Nazi, Dane Netherton, Stephanie Shimada. 1171

Features of a Patient Portal for Blood Test Results and Patient Health Engagement: Web-Based Pre-Post Experiment (e15798)
Bas Struikman, Nadine Bol, Annelijn Goedhart, Julia van Weert, Esther Talboom-Kamp, Sanne van Delft, Anne Brabers, Liset van Dijk. 1181

XSL•FO
RenderX

An Untapped Potential in Primary Care: Semi-Structured Interviews with Clinicians on How Patient Portals Will Work for Caregivers in the Safety Net (e18466)
Alejandra Casillas, Anupama Cemballi, Anshu Abhat, Miya Lemberg, Jennifer Portz, Shobha Sadasivaiah, Neda Ratanawongsa, Wagahta Semere, Arleen Brown, Courtney Lyles. . . . 1191

Patient Portal Implementation and Uptake: Qualitative Comparative Case Study (e18973)
Melita Avdagovska, Mark Ballermann, Karin Olson, Timothy Graham, Devidas Menon, Tania Stafinski. . . . 1202

The Assessment of Supportive Accountability in Adults Seeking Obesity Treatment: Psychometric Validation Study (e17967)
Karishma Chhabria, Kathryn Ross, Shane Sacco, Tricia Leahey. . . . 1219

Benefits of Massive Open Online Course Participation: Deductive Thematic Analysis (e17318)
Elizabeth Blum, Terese Stenfors, Per Palmgren. . . . 1231

Use of a Low-Cost Portable 3D Virtual Reality Gesture-Mediated Simulator for Training and Learning Basic Psychomotor Skills in Minimally Invasive Surgery: Development and Content Validity Study (e17491)
Fernando Alvarez-Lopez, Marcelo Maina, Francesc Saigí-Rubió. . . . 1242

Evaluation of Adaptive Feedback in a Smartphone-Based Game on Health Care Providers' Learning Gain: Randomized Controlled Trial (e17100)
Timothy Tuti, Niall Winters, Hilary Edgcombe, Naomi Muinga, Conrad Wanyama, Mike English, Chris Paton. . . . 1279

An Internet-Based Psychological Intervention With a Serious Game to Improve Vitality, Psychological and Physical Condition, and Immune Function in Healthy Male Adults: Randomized Controlled Trial (e14861)
Lemmy Schakel, Dieuwke Veldhuijzen, Henriët van Middendorp, Corine Prins, Anne Drittij, Frank Vrieling, Leo Visser, Tom Ottenhoff, Simone Joosten, Andrea Evers. . . . 1295

Diabetes-Related Topics in an Online Forum for Caregivers of Individuals Living With Alzheimer Disease and Related Dementias: Qualitative Inquiry (e17851)
Yan Du, Kristi Paiva, Adrian Cebula, Seon Kim, Katrina Lopez, Chengdong Li, Carole White, Sahiti Myneni, Sudha Seshadri, Jing Wang. . . .
1 3 1 2

Assessment of Adaptive Engagement and Support Model for People With Chronic Health Conditions in Online Health Communities: Combined Content Analysis (e17338)
Brian Green, Katelyn Van Horn, Ketki Gupte, Megan Evans, Sara Hayes, Amrita Bhowmick. . . . 1325

Characteristics, Symptom Severity, and Experiences of Patients Reporting Chronic Kidney Disease in the PatientsLikeMe Online Health Community: Retrospective and Qualitative Study (e18548)
Glen James, Elisabeth Nyman, Marcy Fitz-Randolph, Anna Niklasson, Katarina Hedman, Jonatan Hedberg, Eric Wittbrodt, Jennie Medin, Carol Moreno Quinn, Alaster Allum, Cathy Emmas. . . . 1340

Evaluating the Impact of the Grading and Assessment of Predictive Tools Framework on Clinicians and Health Care Professionals' Decisions in Selecting Clinical Predictive Tools: Randomized Controlled Trial (e15770)
Mohamed Khalifa, Farah Magrabi, Blanca Gallego Luxan. . . . 1356

Paving the Way for the Implementation of a Decision Support System for Antibiotic Prescribing in Primary Care in West Africa: Preimplementation and Co-Design Workshop With Physicians (e17940)
Nathan Peiffer-Smadja, Armel Poda, Abdoul-Salam Ouedraogo, Jean-Baptiste Guiard-Schmid, Tristan Delory, Josselin Le Bel, Elisabeth Bouvet, Sylvie Lariven, Pauline Jeanmougin, Raheelah Ahmad, François-Xavier Lescure. . . . 1375

Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study (e16981)
Yang Xiang, Hangyu Ji, Yujia Zhou, Fang Li, Jingcheng Du, Laila Rasmy, Stephen Wu, W Zheng, Hua Xu, Degui Zhi, Yaoyun Zhang, Cui Tao.
1 3 9 1

XSL•FO
RenderX

Development and Usability Evaluation of a Facebook-Based Intervention Program for Childhood Cancer Patients: Mixed Methods Study (e18779)
Bu Park, Ji Kim, Valerie Rogers. . . . 1408

Enabling Better Use of Person-Generated Health Data in Stroke Rehabilitation Systems: Systematic Development of Design Heuristics (e17132)
Gerardo Dimaguila, Kathleen Gray, Mark Merolli. . . . 1422

Recruiting and Retaining Youth and Young Adults in the Policy and Communication Evaluation (PACE) Vermont Study: Randomized Controlled Trial of Participant Compensation (e18446)
Andrea Villanti, Christie Vallencourt, Julia West, Catherine Peasley-Miklus, S LePine, Caitlin McCluskey, Elias Klemperer, Jeffrey Priest, Alison Logan, Bill Patton, Nancy Erickson, Jennifer Hicks, Kathleen Horton, Shayla Livingston, Maria Roemhildt, Erin Singer, Megan Trutor, Rhonda Williams. . . . 1440

Effectiveness of Social Media Approaches to Recruiting Young Adult Cigarillo Smokers: Cross-Sectional Study (e12619)
David Cavallo, Rock Lim, Karen Ishler, Maria Pagano, Rachel Perovsek, Elizabeth Albert, Sarah Koopman Gonzalez, Erika Trapl, Susan Flocke. . . . 1453

Exploring the Privacy-Preserving Properties of Word Embeddings: Algorithmic Validation Study (e18055)
Mohamed Abdalla, Moustafa Abdalla, Graeme Hirst, Frank Rudzicz. . . . 1463

Surveillance of Noncommunicable Disease Epidemic Through the Integrated Noncommunicable Disease Collaborative Management System: Feasibility Pilot Study Conducted in the City of Ningbo, China (e17340)
Sixuan Li, Liang Zhang, Shiwei Liu, Richard Hubbard, Hui Li. . . . 1475

Considerations for Health Researchers Using Social Media for Knowledge Translation: Multiple Case Study (e15121)
Sarah Elliott, Michele Dyson, Gilbert Wilkes, Gabrielle Zimmermann, Christine Chambers, Kristy Wittmeier, Dianne Russell, Shannon Scott, Denise Thomson, Lisa Hartling. . . . 1486

Adherence to Blended or Face-to-Face Smoking Cessation Treatment and Predictors of Adherence: Randomized Controlled Trial (e17207)
Lutz Siemer, Marjolein Brusse-Keizer, Marloes Postel, Somaya Ben Allouch, Robbert Sanderman, Marcel Pieterse. . . . 1501

Effects of Incentives on Adherence to a Web-Based Intervention Promoting Physical Activity: Naturalistic Study (e18338)
Ramona Wurst, Anja Maliezefski, Christina Ramsenthaler, Judith Brame, Reinhard Fuchs. . . . 1519

Synthesizing Multiple Stakeholder Perspectives on Using Virtual Reality to Improve the Periprocedural Experience in Children and Adolescents: Survey Study (e19752)
Naseem Ahmadpour, Andrew Weatherall, Minal Menezes, Soojeong Yoo, Hanyang Hong, Gail Wong. . . . 1530

Combination of 3-Dimensional Virtual Reality and Hands-On Aromatherapy in Improving Institutionalized Older Adults' Psychological Health: Quasi-Experimental Study (e17096)
Vivian Cheng, Chiu-Mieh Huang, Jung-Yu Liao, Hsiao-Pei Hsu, Shih-Wen Wang, Su-Fei Huang, Jong-Long Guo. . . . 1542

Virtual Reality Smartphone-Based Intervention for Smoking Cessation: Pilot Randomized Controlled Trial on Initial Clinical Efficacy and Adherence (e17571)
Emilio Goldenhersch, Johannes Thrul, Joaquín Ungaretti, Nicolas Rosencovich, Cristian Waitman, Marcelo Ceberio. . . . 1554

Biased Sampling and Causal Estimation of Health-Related Information: Laboratory-Based Experimental Research (e17502)
María Moreno-Fernández, Helena Matute. . . . 1569

XSL•FO
RenderX

Crawling the German Health Web: Exploratory Study and Graph Analysis (e17853)
Richard Zowalla, Thomas Wetter, Daniel Pfeifer. 1583

Psychometric Properties of the Norwegian Version of the Electronic Health Literacy Scale (eHEALS) Among Patients After Percutaneous Coronary Intervention: Cross-Sectional Validation Study (e17312)
Gunhild Brørs, Tore Wentzel-Larsen, Håvard Dalen, Tina Hansen, Cameron Norman, Astrid Wahl, Tone Norekvål, The CONCARD Investigators. 1605

The Influence of Electronic Health Record Use on Physician Burnout: Cross-Sectional Survey (e19274)
Tania Tajirian, Vicky Stergiopoulos, Gillian Strudwick, Lydia Sequeira, Marcos Sanches, Jessica Kemp, Karishini Ramamoorthi, Timothy Zhang, Damian Jankowicz. 1637

Adoption and Performance of Complementary Clinical Information Technologies: Analysis of a Survey of General Practitioners (e16300)
Placide Poba-Nzaou, Sylvestre Uwizeyemungu, Xuecheng Liu. 1650

Information Overload in Emergency Medicine Physicians: A Multisite Case Study Exploring the Causes, Impact, and Solutions in Four North England National Health Service Trusts (e19126)
Laura Sbaffi, James Walton, John Blenkinsopp, Graham Walton. 1663

Assessing Patient Experience and Healthcare Quality of Dental Care Using Patient Online Reviews in the United States: Mixed Methods Study (e18652)
Ye Lin, Y Hong, Bradley Henson, Robert Stevenson, Simon Hong, Tianchu Lyu, Chen Liang. 1681

Causal Effect of Honorary Titles on Physicians' Service Volumes in Online Health Communities: Retrospective Study (e18527)
Haiyan Yu, Yali Wang, Jying-Nan Wang, Ya-Ling Chiu, Hang Qiu, Mingyue Gao. 1694

Gender, Soft Skills, and Patient Experience in Online Physician Reviews: A Large-Scale Text Analysis (e14455)
Zackary Dunivin, Lindsay Zadunayski, Ujjwal Baskota, Katie Siek, Jennifer Mankoff. 1708

Changing the Health Behavior of Patients With Cardiovascular Disease Through an Electronic Health Intervention in Three Different Countries: Cost-Effectiveness Study in the Do Cardiac Health: Advanced New Generation Ecosystem (Do CHANGE) 2 Randomized Controlled Trial (e17351)
Jordi Piera-Jiménez, Marjolein Winters, Eva Broers, Damià Valero-Bover, Mirela Habibovic, Jos Widdershoven, Frans Folkvord, Francisco Lupiáñez-Villanueva. 1724

The Cyclic Value-Context Reinforcement Model of Problematic Internet Use: Empirical Validation Using a Thematic Analysis of Children's Counseling Data (e17996)
Young Doh, Bugeun Kim, Seul Lee, Gahgene Gweon. 1742

Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media (e16649)
Shuqing Gao, Lingnan He, Yue Chen, Dan Li, Kaisheng Lai. 1755

Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study (e18082)
Filipe Barata, Peter Tinschert, Frank Rassouli, Claudia Steurer-Stey, Elgar Fleisch, Milo Puhan, Martin Brutsche, David Kotz, Tobias Kowatsch.
1 7 6 6

Artificial Intelligence in Health Care: Bibliometric Analysis (e18228)
Yuqi Guo, Zhichao Hao, Shichong Zhao, Jiaqi Gong, Fan Yang. 1781

Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis (e16021)
Alaa Abd-Alrazaq, Asma Rababeh, Mohannad Alajlani, Bridgette Bewick, Mowafa Househ. 1793

XSL•FO
RenderX

Acceptance and Commitment Therapy Delivered via a Mobile Phone Messaging Robot to Decrease Postoperative Opioid Use in Patients With Orthopedic Trauma: Randomized Controlled Trial (e17750)
Chris Anthony, Edward Rojas, Valerie Keffala, Natalie Glass, Apurva Shah, Benjamin Miller, Matthew Hogue, Michael Willey, Matthew Karam, John Marsh. . . . 1810

Reactions to an Online Demonstration of the Effect of Increased Fruit and Vegetable Consumption on Appearance: Survey Study (e15726)
Patrick Cairns, Gozde Ozakinci, David Perrett. . . . 1823

The Current State of Research, Challenges, and Future Research Directions of Blockchain Technology in Patient Care: Systematic Review (e18619)
Polina Durneva, Karlene Cousins, Min Chen. . . . 1832

Integrating Health Technologies in Health Services for Syrian Refugees in Lebanon: Qualitative Study (e14283)
Reem Talhouk, Chaza Akik, Vera Araujo-Soares, Balsam Ahmad, Sandra Mesmar, Patrick Olivier, Madeline Balaam, Kyle Montague, Andrew Garbett, Hala Ghattas. . . . 1848

Internet-Based Sharing Nurse Program and Nurses' Perceptions in China: Cross-Sectional Survey (e16644)
Rendong Huang, Mei Xu, Xiuting Li, Yinping Wang, Bin Wang, Naixue Cui. . . . 1856

Consequences of Gift Giving in Online Health Communities on Physician Service Quality: Empirical Text Mining Study (e18569)
Li Peng, Yanan Wang, Jing Chen. . . . 1866

Information Loss in Harmonizing Granular Race and Ethnicity Data: Descriptive Study of Standards (e14591)
Karen Wang, Holly Grossetta Nardini, Lori Post, Todd Edwards, Marcella Nunez-Smith, Cynthia Brandt. . . . 1902

Efficacy of a Guided Web-Based Self-Management Intervention for Depression or Dysthymia: Randomized Controlled Trial With a 12-Month Follow-Up Using an Active Control Condition (e15361)
Caroline Oehler, Frauke Görges, Mandy Rogalla, Christine Rummel-Kluge, Ulrich Hegerl. . . . 1907

Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation (e19029)
Yan Zhuang, Yin-Wu Chen, Zon-Yin Shae, Chi-Ren Shyu. . . . 1926

Mapping of Health Literacy and Social Panic Via Web Search Data During the COVID-19 Public Health Emergency: Infodemiological Study (e18831)
Chenjie Xu, Xinyu Zhang, Yaogang Wang. . . . 1948

Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study (e19514)
Hui Xu, Sufang Huang, Chun Qiu, Shangkun Liu, Juan Deng, Bo Jiao, Xi Tan, Ling Ai, Yaru Xiao, Mirko Belliato, Li Yan. . . . 1956

Racial and Ethnic Digital Divides in Posting COVID-19 Content on Social Media Among US Adults: Secondary Survey Analysis (e20472)
Celeste Campos-Castillo, Linnea Laestadius. . . . 1970

Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis (e19322)
Peter Lonergan, Samuel Washington III, Linda Branagan, Nathaniel Gleason, Raj Pruthi, Peter Carroll, Anobel Odisho. . . . 1984

XSL•FO
RenderX

Similarities and Differences in COVID-19 Awareness, Concern, and Symptoms by Race and Ethnicity in the United States: Cross-Sectional Survey (e20001)
Jeb Jones, Patrick Sullivan, Travis Sanchez, Jodie Guest, Eric Hall, Nicole Luisi, Maria Zlotorzynska, Gretchen Wilde, Heather Bradley, Aaron Siegler. . . . 1991

Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study (e19982)
Yulan Lin, Zhijian Hu, Haridah Alias, Li Wong. . . . 1998

Impact of National Containment Measures on Decelerating the Increase in Daily New Cases of COVID-19 in 54 Countries and 4 Epicenters of the Pandemic: Comparative Observational Study (e19904)
Carlos Wong, Janet Wong, Eric Tang, Chi Au, Kristy Lau, Abraham Wai. . . . 2011

Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns (e19483)
Henry Cousins, Clara Cousins, Alon Harris, Louis Pasquale. . . . 2026

Multiple Epidemic Wave Model of the COVID-19 Pandemic: Modeling Study (e20912)
Efthimios Kaxiras, Georgios Neofotistos. . . . 2033

Viewpoints

Artificial Intelligence and Health Technology Assessment: Anticipating a New Level of Complexity (e17707)
Hassane Alami, Pascale Lehoux, Yannick Auclair, Michèle de Guise, Marie-Pierre Gagnon, James Shaw, Denis Roy, Richard Fleet, Mohamed Ag Ahmed, Jean-Paul Fortin. . . . 261

Blockchain in Health Care: Hope or Hype? (e17199)
Rania El-Gazzar, Karen Stendal. . . . 273

The Tango of Problem Formulation: A Patient's/Researcher's Reflection on an Action Design Research Journey (e16916)
Michael Twomey, David Sammon, Tadhg Nagle. . . . 281

Codeveloping a Virtual Patient Simulation to Foster Nurses' Relational Skills Consistent With Motivational Interviewing: A Situation of Antiretroviral Therapy Nonadherence (e18225)
Geneviève Rouleau, Jérôme Pelletier, José Côté, Marie-Pierre Gagnon, Valérie Martel-Laferrière, Rock Lévesque, SimforHealth, Guillaume Fontaine. . . . 304

Unmuting Medical Students' Education: Utilizing Telemedicine During the COVID-19 Pandemic and Beyond (e19667)
Ariella Iancu, Michael Kemp, Hasan Alam. . . . 322

The Story of #GetMePPE and GetUsPPE.org to Mobilize Health Care Response to COVID-19 : Rapidly Deploying Digital Tools for Better Health Care (e20469)
Shuhan He, Ayotomiwa Ojo, Adam Beckman, Suhas Gondi, Megan Ranney, Marian Betz, Jeremy Faust, Esther Choo, Dara Kass, Ali Raja. . 3 2 8

Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges (e19916)
Huailiang Wu, Weiwei Sun, Xinyu Huang, Shinning Yu, Hao Wang, Xiaoyu Bi, Jie Sheng, Sihan Chen, Babatunde Akinwunmi, Casper Zhang, Wai-Kit Ming. . . . 334

An Integrative Model for the Effectiveness of Biofeedback Interventions for Anxiety Regulation: Viewpoint (e14958)
Joanneke Weerdmeester, Marieke van Rooij, Rutger Engels, Isabela Granic. . . . 339

XSL•FO
RenderX

Using Digital Technology to Reduce the Prevalence of Mental Health Disorders in Populations: Time for a New Approach (e17493)
C Taylor, Josef Ruzek, Ellen Fitzsimmons-Craft, Shiri Sadeh-Sharvit, Naira Topooco, Ruth Weissman, Daniel Eisenberg, David Mohr, Andrea Graham, Corinna Jacobi, Brian Oldenburg. 359

Social Medical Capital: How Patients and Caregivers Can Benefit From Online Social Interactions (e16337)
Pietro Panzarasa, Christopher Griffiths, Nishanth Sastry, Anna De Simoni. 369

Bringing Code to Data: Do Not Forget Governance (e18087)
Christine Suver, Adrian Thorogood, Megan Doerr, John Wilbanks, Bartha Knoppers. 373

Seeking Meaningful Innovation: Lessons Learned Developing, Evaluating, and Implementing the Electronic Patient-Reported Outcome Tool (e17987)
Carolyn Steele Gray. 384

Building a Successful Massive Open Online Course About Multiple Sclerosis: A Process Description (e16687)
Suzi Claflin, Rachael Gates, Maree Maher, Bruce Taylor. 1267

Corrigenda and Addendas

Correction: Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment (e21027)
Amanda Graham, George Papandonatos, Megan Jacobs, Michael Amato, Sarah Cha, Amy Cohn, Lorien Abroms, Robyn Whittaker. 1939

Correction: A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework (e21381)
Claire Reidy, Claire Foster, Anne Rogers. 1941

Authorship Correction: The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology (e19961)
Mario Alonso Puig, Mercedes Alonso-Prieto, Jordi Miró, Raquel Torres-Luna, Diego Plaza López de Sabando, Francisco Reinoso-Barbero. . .
1 9 4 3

Correction: Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach (e21753)
Cheng-Sheng Yu, Yu-Jiun Lin, Chang-Hsien Lin, Shiyng-Yu Lin, Jenny Wu, Shy-Shin Chang. 1945

Correction: Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of Invited Girls: Randomized Controlled Trial (e22565)
Mirjam Pot, Theo Paulussen, Robert Ruiter, Iris Eekhout, Hester de Melker, Maxine Spoelstra, Hilde van Keulen. 1946

XSL•FO
RenderX

Review

Technology Acceptance in Mobile Health: Scoping Review of Definitions, Models, and Measurement

Camille Nadal[1], MSc; Corina Sas[2], PhD; Gavin Doherty[1], DPhil

[1]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

[2]School of Computing and Communications, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Camille Nadal, MSc
School of Computer Science and Statistics
Trinity College Dublin
College Green
Dublin, D2
Ireland
Phone: 353 18963858
Email: nadalc@tcd.ie

Abstract

Background: Designing technologies that users will be interested in, start using, and keep using has long been a challenge. In the health domain, the question of technology acceptance is even more important, as the possible intrusiveness of technologies could lead to patients refusing to even try them. Developers and researchers must address this question not only in the design and evaluation of new health care technologies but also across the different stages of the user's journey. Although a range of definitions for these stages exists, many researchers conflate related terms, and the field would benefit from a coherent set of definitions and associated measurement approaches.

Objective: This review aims to explore how technology acceptance is interpreted and measured in mobile health (mHealth) literature. We seek to compare the treatment of acceptance in mHealth research with existing definitions and models, identify potential gaps, and contribute to the clarification of the process of technology acceptance.

Methods: We searched the PubMed database for publications indexed under the Medical Subject Headings terms "Patient Acceptance of Health Care" and "Mobile Applications." We included publications that (1) contained at least one of the terms "acceptability," "acceptance," "adoption," "accept," or "adopt"; and (2) defined the term. The final corpus included 68 relevant studies.

Results: Several interpretations are associated with technology acceptance, few consistent with existing definitions. Although the literature has influenced the interpretation of the concept, usage is not homogeneous, and models are not adapted to populations with particular needs. The prevalence of measurement by custom surveys suggests a lack of standardized measurement tools.

Conclusions: Definitions from the literature were published separately, which may contribute to inconsistent usage. A definition framework would bring coherence to the reporting of results, facilitating the replication and comparison of studies. We propose the Technology Acceptance Lifecycle, consolidating existing definitions, articulating the different stages of technology acceptance, and providing an explicit terminology. Our findings illustrate the need for a common definition and measurement framework and the importance of viewing technology acceptance as a staged process, with adapted measurement methods for each stage.

(J Med Internet Res 2020;22(7):e17256) doi:10.2196/17256

KEYWORDS

Technology Acceptance Lifecycle; patient acceptance; mobile applications; mHealth; mobile phone

Introduction

Background

Technology acceptance is a major challenge faced by designers of new technologies. In health care, patients are a vulnerable population, and their data are considered sensitive, especially in the case of stigmatized conditions such as those concerning mental health [1,2]. As mobile technology becomes increasingly pervasive in health care [3], the expanding use of potentially invasive technologies such as sensing and machine learning is

XSL•FO
RenderX

likely to lead to greater concerns among users, exacerbating existing problems with attrition, and willingness to use new technologies. In addition, health care providers' perception of a technology is likely to affect treatment delivery, especially if it is not considered sufficiently acceptable [4].

The last decade has seen an increasing number of studies addressing this issue in the mobile health (mHealth) domain. A recent systematic review by Wozney et al [5] revealed that acceptability was the most frequently measured outcome in studies on mental health technologies for anxiety and depression. Designing for acceptance is not straightforward, as the user journey is complex and often nonlinear. Patients who decide to try an application will not necessarily use it in the long run, and similarly, someone who has stopped using a system might go back to it later [6,7]. Conditions such as depression might also impact patients' perception of their experience with technology [8] and thus affect their acceptance. Different stages punctuate the users' journey with technology where they, consciously or not, repeatedly make the decision to keep using or to abandon it [9]. In addition, the extent to which users have appropriated and integrated technology into their lives may impact their decision to maintain use.

Terminology and Definitions

An evolving terminology and range of definitions can be found in the literature discussing technology acceptance. Terms such as acceptability, acceptance, and adoption are often employed, sometimes interchangeably. For instance, Al-Youssef [10] refers to acceptability as users' willingness to use a system while citing the definition of acceptance given in Dillon and Morris [11]. These terms are sometimes equated to other human-computer interaction (HCI) concepts such as user satisfaction [12]. Yet, a part of the literature tries to differentiate the process of technology acceptance from existing concepts. For instance, Schade and Schlag [13] describe acceptability as "a prospective judgement of measures to be introduced in the future," which they detail as "the target group will not yet have experienced the new measures." Close to this interpretation, Adell [14] equates acceptance to "the degree to which an individual intends to use a system and, when available, incorporates the system in his/her [driving]."

The Cambridge Dictionary defines acceptability as "the quality of being satisfactory and able to be agreed to or approved of" [15] and acceptance as a "general agreement that something is satisfactory or right" [16]. This suggests that acceptability is a *quality* of an object from the perspective of a stakeholder, whereas acceptance is more of a process relating to a stakeholder's *interaction* with this object. In the context of HCI, it translates into a user-system dyad.

In the same line, Proctor et al [17] define acceptability as "the perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory."

With the literature highlighting the importance of temporality in user experience research [18,19], some authors attempt to integrate a temporal dimension into the process of technology acceptance. In that respect, a group of authors supported the idea of a multistage process. Martin et al [20] and Distler et al [21] define technology acceptability as one's perception of a system before use. They distinguish acceptability from acceptance by equating acceptance to the users' perception of the system after use. Terrade et al [22] and Garces et al [23] go a step further, stating that acceptance refers to the initial use of a system in controlled settings. Reerink-Boulanger [24] introduces the continuum acceptability-acceptance-appropriation, describing acceptability as the subjective perception of the use of a system, acceptance as the first interactions with the system and appropriation as the use of the system by an individual in interaction with other individuals (translated from the original paper in French). Other definitions of appropriation include "a process of social construction in which the actions and thoughts of the user are shaped by the technology, while the meaning and effects of the technology are shaped through the users' actions" [25] and "the process through which technology users go beyond mere adoption to make technology their own and to embed it within their social, economic, and political practices" [26].

Whereas a consistent framework seems to emerge from these definitions, another term, adoption, comes into the picture. Some authors such as Carroll et al [27] equate the process of adoption with the entire user journey: "a multi-phase process starting with deciding to adopt (selecting, purchasing, or committing to use it) and then achieving persistent use." In contrast, Karahanna et al [28] describe "preadoption" and "postadoption (continued use)" stages and present adoption as an event allowing progress from one stage to the other. Similarly, Rogers [29] defines adoption as the user's decision to "make full use of an innovation as the best course of action available."

Models

Another strand of work has attempted to identify the factors influencing technology acceptance, the most well-known of which is the technology acceptance model (TAM) of Davis [30]. In the health domain, Kim and Park [31] have proposed a health information technology acceptance model (HITAM). This model integrates the TAM by Davis, along with antecedents and health-related constructs (health status, health beliefs and concerns, behavioral beliefs, and perceived health threat). Other models, such as the pervasive technology acceptance model [32], integrate the influence of demographics and trust on user acceptance. Dou et al [33] introduced the constructs *relationship with doctor* and *resistance to change*. Finally, Cheung et al [34] introduced constructs related to privacy and consumer innovativeness.

However, the temporal dimension is missing from these models. This raises the question of whether acceptance is motivated by the same factors and in the same manner, regardless of how long the system has been used. As stated in a review on user engagement, describing a concept as a process, rather than a discrete state, "enables the analysis of change over time" [35]. In line with this, some authors propose models that incorporate temporality. Karapanos et al [36] showed that, in their study, qualities that satisfied participants at initial use did not necessarily motivate prolonged use. Building on this, Karapanos et al [37] explored factors influencing user experience over

XSL•FO
RenderX

time; they proposed a temporal framework that identifies 4 stages: anticipation (formation of expectations before any experience of use), orientation (users' initial experiences), incorporation (how the product becomes meaningful in the user's life), and identification (how the product participates in users' social interactions). Although the work addresses user experience rather than acceptance, the authors claim that different qualities contribute to user experience over time, and thus, time alters the way users experience technology. As the extent to which users accept a technology is undeniably linked to their usage experience—or lack of—it is relevant to ask how user acceptance evolves over time and how the influencing factors vary accordingly.

Further exploring the temporal dimension of acceptance, Greenhalgh et al [38] proposed a framework for nonadoption, abandonment, scale-up, spread, and sustainability (NASSS). Directed at health care technologies, the NASSS framework describes different domains influencing technology adoption. It adds a temporal dimension targeting technology's "continuous embedding and adaptation over time." The framework is designed to be used at different points in time (at early design, after deployment, and after abandonment), which distinguishes it from the other models.

Measurement

Both definitions and more detailed models of acceptance may find expression in the form of approaches to measurement. Examples of acceptance studies at different stages of the technology lifecycle can be found in the literature, in some cases employing measurement tools based on theoretical models. This includes gathering of qualitative data via focus groups and interviews based on the TAM [39], a survey based on the unified theory of acceptance and use of technology [40], and interviews based on the fast form TAM [12].

Proctor et al [17] defined eight outcomes for the measurement of implementation of health interventions, two of which were acceptability and adoption. In a recent paper, Hermes et al [41] proposed to adapt the characterization of these outcomes for behavioral intervention technologies. The authors explicitly link the adoption outcome to the use (or intention to use) of a system. In addition, they state that "usability clearly overlaps with acceptability", and thus, usability measurement tools (such as the system usability scale) could be used to assess acceptability. It would be interesting to see whether researchers chose to measure a system's usability to make inferences regarding its acceptability.

Although there seems to be an effort in the research community to ground assessment methods in existing theoretical frameworks, custom measurement tools such as ad hoc surveys are also used to investigate technology acceptance (eg, Allen and Seaman [42]). The use of ad hoc tools allows explorations of acceptance to be tailored to a particular context but makes it more difficult to compare results across projects.

Objective

A number of definitions and models of technology acceptance are available in the literature, several of which make important distinctions between the different stages of the process. Omitting that distinction in the terminology, and using the terms acceptability, acceptance, and adoption interchangeably can create ambiguities about what is actually being measured, making the replication of interventions and comparison of results difficult. Such ambiguity is highlighted in a review [4] that showed the confusion around the concept of technology acceptability in the health care literature. Furthermore, a review of definitions for electronic health [43] stressed the importance of common terminology for interdisciplinary collaboration. Technology acceptance is a particularly significant challenge for the design of mHealth care technologies; hence we focus on that particular context. In this paper, we present the results of a scoping review of the mHealth literature addressing the following questions:

- How do researchers define technology acceptance?
- What terminology is used to refer to technology acceptance?
- How do researchers measure technology acceptance?
- How do researchers make use of existing models of acceptance?

In addition to outlining research practices to evaluate acceptance, this review reveals the potential limitations and areas for improvement of existing models. A better understanding of these elements contributes toward the development (or improvement) of methodologies and measurement tools for addressing technology acceptance within the development of mHealth applications. Informed by this analysis, a further contribution is made by integrating and disambiguating existing definitions. To this end, we present a lifecycle of the process of acceptance, providing researchers with a common terminology to report results, and help them to measure the evolution of user acceptance over time.

Methods

We performed a scoping review to map relevant literature in the field of mHealth. In contrast to systematic and narrative reviews [44], scoping reviews allow for a broad but structured exploration that permitted us to examine the range of definitions and measurements of technology acceptance, and to identify gaps and inconsistencies in the existing literature. We followed the framework developed by Arksey and O'Malley [45], which consists of the stages outlined below.

Identifying the Research Question

We were interested in how technology acceptance was understood, measured, and reported in mHealth studies.

Searching for Relevant Studies

We searched the PubMed database for papers addressing technology acceptance. Following preliminary searches to assess the relevance of search criteria, a search of the Medical Subject Headings (MeSH) terms "Patient Acceptance of Health Care" and "Mobile Applications," without time restriction, resulted in a corpus of 287 articles starting from 2013.

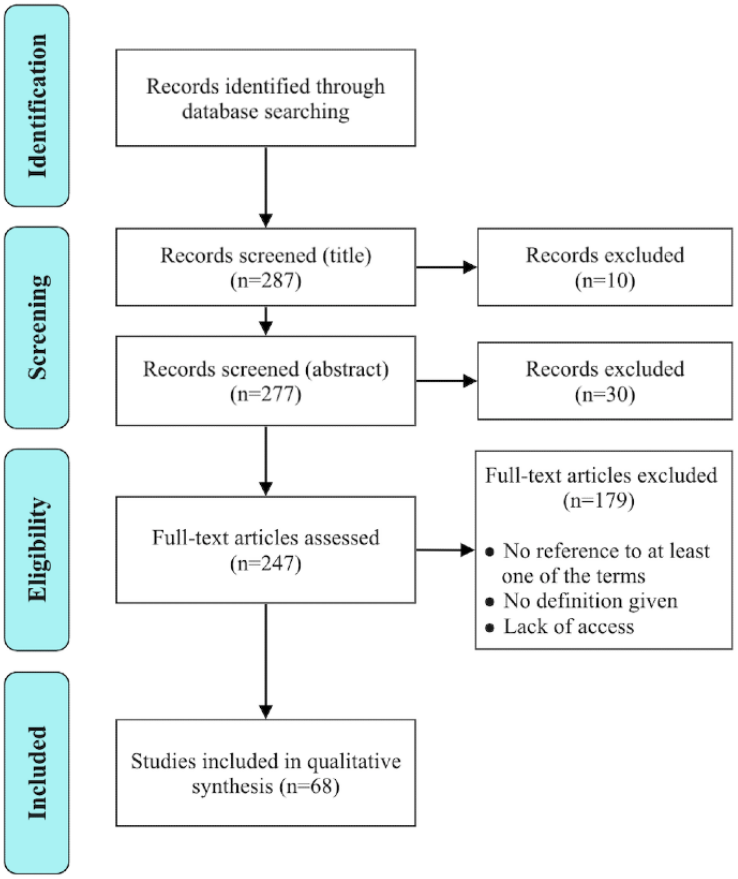
Selecting Studies to Include

Inclusion criteria were subject to discussion between the 3 authors; 2 random samples of 10 papers were independently assessed for inclusion by 2 pairs of authors (first and second,

XSL•FO
RenderX

first and third). Disagreements concerning the definitions of inclusion criteria were resolved by explicitly stating these criteria as following: a publication was considered relevant if (1) it contained at least one of the terms "acceptability," "acceptance," "adoption," "accept," or "adopt"; and (2) defined the concept in question with (1) a full definition; (2) a synonym; or (3) an operationalized definition (ie, means used to measure the concept). Some articles were associated with the specified MeSH terms but did not directly discuss technology acceptance. Articles were excluded if they did not contain any of the 5 terms above, did not provide a definition, or if their web version was not accessible. Among those, 40 were excluded at the screening stage (Figure 1) because they did not address acceptance in the sense of technology acceptance (eg, Acceptance and Commitment Therapy). A total of 68 relevant publications were included in the review.

Figure 1. Flow diagram.



Charting the Data

The codes used for analysis aimed to expose how technology acceptance was referred to, defined, and measured, depending on the context (Table 1). Each pair of authors coded 2 samples of 10 papers independently. This resulted in no disagreement, the codes were straightforward to interpret.

XSL•FO
RenderX

Table 1. Distribution of the main codes (some studies performed multiple measurements).

Themes and codes	Studies, n (%)
Given definition	68 (100)
Nature of given definition	
Operationalized	39 (58)
Synonym	22 (32)
Full definition	7 (10)
Cited reference for given definition	
Yes	13 (20)
No	55 (80)
Intervention domain	
Mental health	55 (80)
Health	12 (18)
Both	1 (2)
Goal of assessing acceptance	
Inform design	13 (19)
Evaluate a system	48 (71)
Review the literature	7 (10)
Acceptance measurement	
Customized survey	39 (54)
Standardized survey	13 (18)
Usage	7 (10)
Interviews (qualitative)	6 (9)
Focus groups (qualitative)	5 (7)
Adherence	1 (1)
Completion of task	1 (1)
Measurement timeline	
Preuse	16 (19)
Initial use	7 (10)
Sustained use	42 (61)
Unclear	1 (1)
No measurement	6 (9)

Collating, Summarizing, and Reporting the Findings

Definitions from the literature mentioned earlier distinguish between the different stages of the process of acceptance. We assessed whether the definitions found in our corpus also differentiate between the stages of acceptance and the different terms *acceptability*, *acceptance*, and *adoption*. We carried out 2 classifications of the papers, one with respect to the terminology used (ie, *acceptability*, *acceptance*, or *adoption*) and one following the measurement stage (ie, preuse, initial use, and sustained use). We then grouped papers with similar interpretations of technology acceptance and extracted the measurements used. Finally, we identified possible gaps and suggested ways to improve the exploration of acceptance in mHealth.

Results

All 3 terms (acceptance, acceptability, and adoption) were used in the mHealth literature sample and were associated with various meanings and measurements.

Definitions

Classification of the papers reveals that one-fifth of them cited definitions from the literature. Two-third of the articles reported on a mental health intervention, which shows that the question of technology acceptance is particularly studied in this context. Figure 2 compares the terminology used in the papers with a classification following the measurement stage. A significant difference can be seen between the 2 distributions, marking the

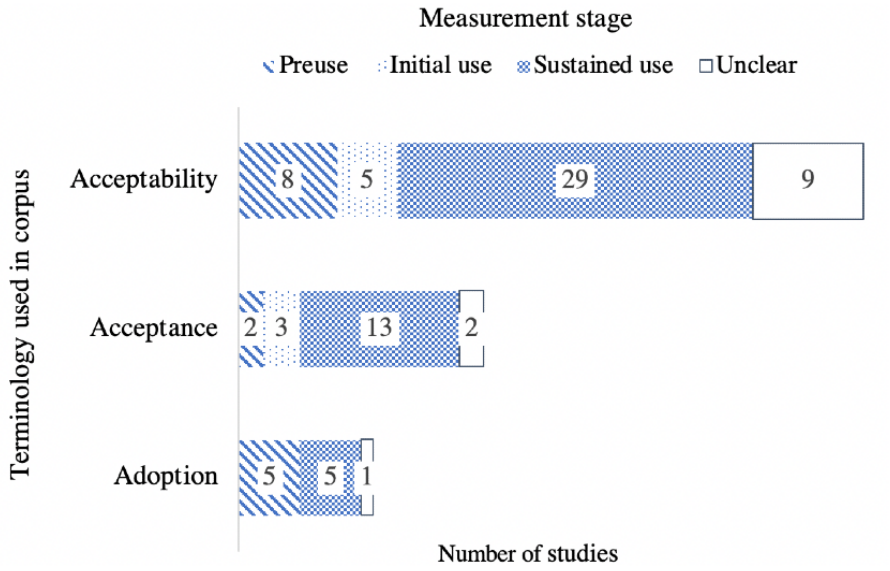
XSL•FO
RenderX

distinction between the 3 terms, but also their use in relation to the measurement timeline.

The majority of publications (n=51) referred to technology acceptability. Among these, a small number (n=8) addressed the preuse stage, sometimes equating acceptability to users' interest or willingness to use a system [46-50]. The other 34 papers referred to acceptability in the context of initial or sustained use.

Papers in the second group (n=20) referred to technology acceptance. Among these, 2 papers refer to it at preuse, and 13 in the context of sustained use.

Figure 2. Comparison of the terminology used with a classification following the measurement stage (some papers employed multiple terms).



The last part of the corpus (n=11) refers to technology adoption. Half of these papers focused on the sustained use stage and the other half on the preuse stage. Papers reported as unclear were those for which classification was not possible as the context of the study was ambiguous. Although the terminology used showed a focus on acceptability, looking at the corpus through the lens of the measurement timeline highlighted that more than two-third of the papers (n=47) explored the sustained use stage. This distribution suggested a trend in the corpus to employ the term *acceptability*, regardless of the context and stage of the user journey. Table 2 presents the range of interpretations for acceptability, acceptance, and adoption extracted from the corpus.

XSL•FO
RenderX

Table 2. Meanings associated with terminology (some studies referred to several concepts).

Terms and associated meanings	Occurrences [references], n
Acceptability	
Perceived usefulness	14 [51-64]
User satisfaction	11 [54,55,57,60,65-71]
System usability	8 [48,57,67,72-76]
User feedback	8 [52,53,61,73,75,77-79]
Rate of recommendation	8 [51,53,54,61,62,73,79,80]
Actual usage	8 [64,79,81-86]
Perceived efficiency	6 [51,59,66,73,79,86]
Perceived ease of use	5 [52,53,63,64,73]
Intention to use	5 [46,47,50,64,87]
User engagement	4 [58,79,81,88]
User enjoyment	4 [51,58,64,79]
Attitude toward using	2 [49,64]
Quality of the system	1 [64]
Acceptance	
Perceived usefulness	5 [64,89-92]
Intention to use	5 [64,87,90,91,93]
Actual usage	5 [64,85,90,91,94]
User satisfaction	4 [71,89,93,95]
Perceived ease of use	3 [64,90,91]
Attitude toward using	3 [64,90,91]
Perceived efficiency	2 [89,93]
System usability	2 [76,96]
Quality of the system	2 [64,93]
User feedback	1 [97]
User enjoyment	1 [64]
Adoption	
Actual usage	7 [84-86,91,98-100]
Intention to use	3 [91,101,102]
Perceived usefulness	2 [63,91]
Perceived ease of use	2 [63,91]
Perceived efficiency	1 [86]
Attitude toward using	1 [91]

The distribution shown in Table 2 reveals that acceptability was mainly understood as perceived usefulness (a concept from TAM) or user satisfaction. System usability, user feedback, and the other TAM constructs (perceived ease of use, attitude toward using, intention to use, and actual usage) were often mentioned. It also emerged that researchers may consider a perceived reduced stigma [49] and high similarity between the behavior of the technology and traditional health care [61] as markers for acceptability. Furthermore, one study explored sociocultural aspects of technology acceptability in developing regions, explaining that a system needs to take into account "the preferences and aspirations of individual service users and the cultures of their communities" [103]. A striking finding was that, out of 6 years of research, only 2 works explored mHealth for children [87,104], with Farooqui et al [104] alone studying children's acceptance.

Fewer studies employed the term acceptance, associating it with constructs from the TAM—perhaps because the model itself uses the term acceptance.

TAM constructs were also found in the interpretations of adoption. Some of this work seeks to adapt or extend the TAM,

XSL•FO
RenderX

such as Khatun et al [105], who developed the concept of *readiness to adopt* mHealth in developing countries. The authors argued that the TAM "does not consider the influence of human factors, the internal resources of the user or the external environmental and ecological factors" and proposed a model integrating TAM constructs and others more specific to rural and developing areas (access, sociodemographics, awareness of mHealth services, and trust).

Among the literature reviews present in the corpus, some associated acceptability or acceptance with feasibility [97,106]. Feasibility, as defined by the National Institute for Health Research glossary, relates to whether a study can be carried out. A feasibility study explores not only technical parameters but also human factors likely to be important for the conduct of the main study. Although a feasibility study may include the assessment of participants' acceptance, it is not the case that all feasibility studies will do so.

A small number of studies (n=13) cited conceptual definitions from the literature. The TAM was cited in reference to the 3 concepts: acceptability [64,65], acceptance [64,90,91], and adoption [91,102,105].

However, some studies have highlighted the limitations of existing models. Zhu et al [102] argue that conceptual models should not regard mobile services as a generic concept but specifically address particular use cases (eg, mobile services for health monitoring). They present their own model for technology acceptance, combining TAM constructs with health-related concepts from the health belief model [107].

In the same vein, Povey et al [108] stated that existing models were not suitable for their study. The authors attempted to build a model for the acceptability of e-mental health for an indigenous population and acknowledged that the resulting model is similar to the HITAM [31].

Ammenwerth et al [109] argued that existing models such as the TAM and task technology fit failed to address the interaction between user and task. To bridge this gap, the authors developed and validated the fit between individuals, task, and technology (FITT) framework for information technology adoption in clinical environments. Only one study [85] then employed FITT to measure acceptance of the use of mobile apps among physical therapists. Finally, another study [98] cited Agarwal and Prasad [110], supporting the idea of 2 stages: initial adoption and long-term engagement.

These references to models and definitions within the literature show researchers' willingness to reuse existing theories. However, the adaptations of these models and their inconsistent interpretation also reveal their limitations and the lack of a common framework to study technology acceptance.

Measurement

We extracted measurements employed in the corpus to assess technology acceptance and classified them by type and temporality (Table 3).

Table 3. Measurements performed (some studies performed several measurements).

Measurement	Occurrences per stage		
	Preuse (11 studies), n	Initial use (7 studies), n	Sustained use (40 studies), n
Customized survey	7	2	28
Standardized survey	2	2	8
Focus groups (qualitative)	2	1	1
Interviews (qualitative)	1	1	4
Completion of task	0	1	1
Usage	0	0	6

Table 3 illustrates a strong preference for the use of surveys (73.0%, 49/67), independently of the timeline, with a prevalence of custom questionnaires (55.0%, 37/67). Almost all studies assessing system usability (which is a frequent interpretation of acceptability, see Table 2), made use of standardized surveys, which reflects the range of reliable tools available to evaluate this concept [72,74].

Other studies used existing surveys incorporating TAM constructs [64,91] or developed their own with elements from different models to fit their research better [102]. Owing to the need to assess technology acceptance in developing countries and remote areas, Chen et al [64] suggested that models found in the literature may need to be validated on a larger variety of populations.

The rest of the corpus measured technology acceptance through participant feedback (via custom surveys, focus groups, and interviews), adherence, usage, and rate of completion of tasks. Although this wide set of metrics contributes to expanding the number of assessment tools, it also impedes consistency within technology acceptance research.

The low number of measures applied at the pre and initial use stages also reveals that technology acceptance was rarely investigated at the design stage (Table 4).

XSL•FO
RenderX

Table 4. Measurement timeline of constructs (some studies measured several constructs).

Constructs (occurrences)	Measurement stage [references]		
	Preuse (n=13)	Initial use (n=7)	Sustained use (n=87)
Perceived usefulness (n=18)	[61]	[57,58]	[52-56,58-60,62-64,89-92]
User satisfaction (n=15)	—[a]	[57,65]	[54,55,60,62,66-71,89,93,95]
Intention to use (n=11)	[46,47,50,101,102,105]	—	[64,87,90,91,93]
System usability (n=10)	[48]	[57]	[55,67,72-76,96]
Rate of recommendation (n=10)	[61]	—	[53-55,62,73,79,80,104,111]
Actual usage (n=9)	—	—	[64,79,81,83,84,86,90,91,94]
User feedback (n=8)	[61,77]	[78]	[52,53,73,75,79]
Perceived ease of use (n=7)	—	—	[52,53,63,64,73,90,91]
Perceived efficiency (n=6)	—	—	[59,66,73,79,86,89,96]
Attitude toward use (n=4)	[49]		[64,90,91]
User enjoyment (n=4)	[58]	[58]	[64,79]
User engagement (n=3)	—	—	[79,81,88]
Quality of the system (n=2)	—	—	[64,93]

[a]No study measured the construct at that specific measurement stage.

Interestingly, although existing acceptance models provide sets of measurable constructs, researchers' efforts have focused on perceived usefulness. Only the studies in Table 5 measured the full set of constructs contained in the technology acceptance models which they cite.

Table 5. Use of existing technology acceptance models.

Acceptance models used in the corpus	Measurement stage			Additional constructs
	Preuse	Initial use	Sustained use	
TAM[a]	[102]	—[b]	—	Perceived disease threat, perceived risk, initial trust, and technology anxiety
	—	—	[64]	System quality, social influence, perceived enjoyment, and smartphone experience
	—	—	[90]	Demographics (age, position at work, usage time of PDA[c], and skill level of using a PDA)
	—	—	[91]	None
Information system success model	—	—	[93]	None

[a]TAM: technology acceptance model (Davis [30]).

[b]Model was not used at that measurement stage.

[c]PDA: personal digital assistant.

Table 5 shows that almost all studies that relied on the TAM added constructs to capture the influence of context-related factors on acceptance. Indeed, Zhu et al [102] followed the added variables approach described by Holden and Karsh [112] to evaluate the acceptance of their technology within the specific Chinese health care context. In addition to the TAM constructs, they used 4 context-related constructs taken from other studies. Similarly, Chen et al [64] used a questionnaire based on the TAM with additional constructs from other studies to fit the context of their application. Finally, Wang et al [90] showed the influence of certain demographic factors (age, position at work, usage time of the personal digital assistant [PDA], and skill level of using a PDA) on the TAM constructs (perceived usefulness and ease of use).

On the other hand, Ammenwerth et al [93] used a survey based on the information system success model [113] to evaluate acceptance postuse at 2 different points in time. The first survey assessed the 7 constructs from the model, whereas the second only assessed intention to use and net benefit. This could suggest that these constructs are seen as more stable or important, or that the existing constructs are not suitable for all study designs.

As expected from the analysis of definitions, these studies measured a wide variety of constructs. New questionnaires were developed, and researchers stressed the limitations of existing measurement frameworks. Without the validation of these new surveys, the comparison and replication of studies may be difficult. Thus, we argue that standardized ways to measure

XSL•FO
RenderX

technology acceptance are needed to support objective and coherent assessments at the different stages of the user journey.

Discussion

This scoping review analyzed the terminology used to refer to technology acceptance and extracted the different interpretations given and measurement methods employed.

Terminology

The findings reveal that technology acceptance was mostly referred to as acceptability and, to a lesser extent, acceptance or adoption. A small part of the corpus converges toward the differentiation of these terms as distinct concepts, corresponding to stages (ie, preuse acceptability, initial and sustained use acceptance) or events (ie, point of adoption) in the user journey of technology acceptance. However, the rest of the corpus does not distinguish between these terms. One reason could be that the existing definitions arguing for that distinction were published separately; hence, providing a coherent and more precise terminology and set of definitions can aid researchers in communicating which concepts they are referring to.

Definitions

The corpus provided a variety of interpretations of the concept of technology acceptance. Among these, some were based on constructs present in literature models such as the TAM [30], indicating that existing models do influence acceptance studies. Other definitions should be associated with the appropriate model constructs; for example, user satisfaction and feedback should refer to the TAM's construct *attitude toward use*. Similarly, factors related to the sensitive nature of health care technologies (such as perceived stigma) should be linked to the appropriate constructs' *subjective norm* (in the HITAM), *social influence* (in the pervasive technology acceptance model), or *sharing* (in the model by Cheung et al [34]). Thus, it is likely that the extent to which an individual is concerned by these risks will impact their acceptance of the technology. Following previous research [102,108], we argue that acceptance models should take into account the cultural and health context of end users.

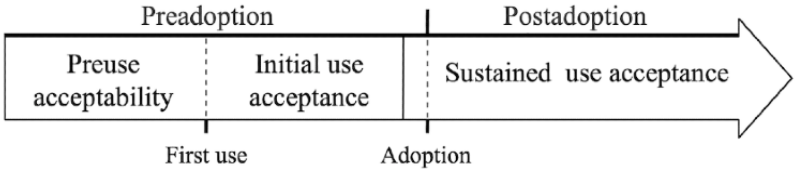
Models and Measurement

Although part of the corpus employed standardized tools, the majority of the studies used custom surveys. This aligns with the findings of Wozney et al [5] that the major part of their corpus used nonvalidated measures of acceptability. This could be because of a lack of validated tools to assess technology acceptance in the context of mHealth. Many researchers felt the need to design their own survey to have a measurement instrument adapted to the specific issues of their target population (eg, technology access and cost). Similarly, researchers who used existing acceptance models felt the need to add context-specific constructs. It would be interesting to see a community effort to validate new tools, create adapted tools for important contexts such as mHealth, and adapt existing models and questionnaires so that they embrace the changes in users' acceptance as they use the technology. Finally, technology acceptance was rarely investigated at the design stage. This is unexpected as exploring acceptance issues could greatly inform design work [114] and reduce risks that an implemented technology is rejected or abandoned. On the use of machine learning in clinical contexts, Thieme et al [115] argue that collaborating with health care users at the design stage may increase the chances of acceptance of the technology. Existing acceptance models were only used once to measure acceptance at the preuse stage (Table 5). A reason for this might be that existing models do not target the design stage and are not adapted to measurement at the preuse stage.

Technology Acceptance Lifecycle

In line with the existing literature definitions and informed by our analysis of the corpus, we argue for clearly distinguishing between the different stages of technology acceptance. We have seen in our analysis both a wide variety of interpretations of these concepts, and a range of measurement approaches applied across the lifecycle, mostly based on ad hoc tools. This variety creates ambiguities in the reporting and understanding of results and makes it difficult to draw conclusions on the acceptance of the systems studied. We believe that a better understanding of the process of technology acceptance would greatly benefit the community in terms of researchers articulating their findings with regard to the entire process. To contribute toward clarification of the measurement of technology acceptance, we propose the Technology Acceptance Lifecycle (TAL). The TAL consists of a timeline to anchor the definitions of technology acceptance within the overall process (Figure 3).

Figure 3. Proposed terminology for technology acceptance lifecycle.



Our motivation is to highlight the evolving nature of technology acceptance across the different stages of the user journey with a technology. The two main stages of the TAL follow Karahanna et al [28] and Roger's interpretations [29] in favor of a distinction between the pre- and postadoption stages. With regard to the definitions of Distler et al [21], Martin et al [20], Garces et al [23], and Terrade et al [22], we argue for a distinction between the stages of acceptability (preuse) and acceptance (initial use) and propose the continuum acceptability-acceptance-sustained use. The TAL proposes a more explicit terminology, embedding temporality in the name of the different stages. The continuum then becomes preuse acceptability—initial use acceptance—sustained use acceptance. Finally, sometime during the sustained use stage, the user would reach the point of

XSL•FO
RenderX

adoption of the technology. According to Rogers [29], a system can be considered adopted when users make full use of it. However, the literature does not specify the conditions for full use to be achieved. On the basis of the existing literature definitions, the TAL articulates the process of technology acceptance and its different stages across the user journey. We argue that acceptance at the initial and sustained use stages is likely to be impacted by factors related to the actual use of the technology and user engagement. Therefore, we suggest that the research community considers acceptance as a process (rather than a discrete measure) and adopts assessment approaches that take into account the temporal dimension and possible evolution of acceptance. We believe that the proposed TAL could help align the research field and provide researchers with a timeline that they can follow to assess technology acceptance and terminology to communicate their research clearly. Further research is needed to establish the influencing factors at the different stages of the process and to develop and validate measurement methods adapted to these stages.

Limitations

Our search focused on the PubMed database and relied on the MeSH classification. We did not apply any time restriction and obtained a reasonably sized and highly relevant sample. However, a more exhaustive sample could be obtained by expanding the search to other terms.

Conclusions

This review has identified the common interpretations and measurement approaches that have been used to assess the acceptance of mHealth technologies. To our knowledge, this is the first study to look at the terminology employed and examine the basis and consistency of the definitions employed in mHealth research. In addition, this review has described how researchers made use of existing models to measure mHealth acceptance and the lack of readily available assessment tools that are appropriate for specific study contexts and use at the design stage. This paper has uncovered the need for a common definition and measurement framework to address technology acceptance, particularly in the domains of health and mental health. A common set of definitions and more consistent approaches to measurement would support both developers of mHealth applications in addressing user acceptance of their systems and the communication and reporting of results in interdisciplinary studies of mHealth interventions. With the TAL, we propose a more explicit terminology and a representation of the process of technology acceptance throughout the user journey.

Our findings highlight the importance of better articulating the specific concepts highlighted in the TAL and developing appropriate measurement tools, ideally standardized, for each of these concepts. This perspective encourages developers to consider acceptance across the user journey and allows researchers to be more explicit about what they are investigating. Finally, efforts from the mHealth community are needed to adapt existing acceptance models for use in sensitive contexts such as mental health interventions.

Acknowledgments

This work was supported by AffecTech: Personal Technologies for Affective Health, Innovative Training Network funded by the H2020 People Programme under Marie Skłodowska-Curie grant agreement number 722022. This research was supported, in part, by Science Foundation Ireland grants 13/RC/2106 (Adapt) and 13/RC/2094 (Lero).

Conflicts of Interest

None declared.

References

1. Sanches P, Janson A, Karpashevich P, Nadal C, Qu C, Dauden RC, et al. HCI and Affective Health: Taking Stock of a Decade of Studies and Charting Future Research Directions. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300475]
2. Coman A, Sas C. A hybrid intervention for challenging the stigma of mental illness. Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law 2016;9(58):73 [FREE Full text]
3. Stowell E, Lyson M, Saksono H, Wurth R, Jimison H, Pavel M, et al. Designing and Evaluating mHealth Interventions for Vulnerable Populations: A Systematic Review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173589]
4. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
5. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
6. Matthews M, Doherty G. In the Mood: Engaging Teenagers in Psychotherapy Using Mobile Phones. In: Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; March 15-17, 2011; Maryland, United States. [doi: 10.1145/1978942.1979379]

XSL•FO
RenderX

7. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
8. Qu C, Sas C, Doherty G. Exploring and Designing for Memory Impairments in Depression. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300740]
9. Salovaara A, Hook K, Cheverst K, Twidale M, Chalmers M, Sas C. Appropriation and creative use: linking user studies and design. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems. USA: Association for Computing Machinery; 2011 Presented at: CHI '11; 2011; Vancouver, BC, Canada p. 37-40. [doi: 10.1145/1979742.1979585]
10. Al-Youssef I. Student acceptance and use of internet-based distance education in Saudi Electronic University (SEU): a mixed method study. Ohio University: OhioLINK Electronic Theses and Dissertations Center; 2015. URL: https://etd.ohiolink.edu/!etd.send_file?accession=ohiou1424175265&disposition=inline [accessed 2020-06-01]
11. Dillon A, Morris M. User acceptance of new information technology: theories and models. In: Annual Review of Information Science and Technology. Medford, NJ: Information Today; 1996:3-32.
12. Fischer AJ, Dart EH, Leblanc H, Hartman KL, Steeves RO, Gresham FM. An investigation of the acceptability of videoconferencing within a school-based behavioral consultation framework. Psychology in the Schools 2016;53(3):240-252. [doi: 10.1002/pits.21900]
13. Schade J, Schlag B. Acceptability of urban transport pricing strategies. Transp Res Part F Traffic Psychol Behav 2003 Mar;6(1):45-61. [doi: 10.1016/s1369-8478(02)00046-3]
14. Adell E. Acceptance of driver support systems. 2010 Presented at: Proceedings of the European conference on human centred design for intelligent transport systems; 2010; Berlin, Germany p. 475-486 URL: https://www.researchgate.net/publication/229049067_Acceptance_of_driver_support_systems
15. Acceptability. Cambridge Dictionary - Cambridge University Press. 2020. URL: https://dictionary.cambridge.org/dictionary/english/acceptability [accessed 2020-05-21]
16. Acceptance. Cambridge Dictionary - Cambridge University Press. 2020. URL: https://dictionary.cambridge.org/dictionary/english/acceptance [accessed 2020-05-21]
17. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Administration and Policy in Mental Health and Mental Health Services Research 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
18. Forlizzi J, Battarbee K. Understanding Experience in Interactive Systems. In: Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques. 2004 Presented at: DIS'04; May 22-23, 2004; Miami, FL. [doi: 10.1145/1013115.1013152]
19. Hassenzahl M, Tractinsky N. User experience - a research agenda. Behaviour & Information Technology 2006;25(2):91-97. [doi: 10.1080/01449290500330331]
20. Martin N, Erhel S, Jamet E, Rouxel G. What Links Between User Experience and Acceptability? In: Proceedings of the 27th Conference on l'Interaction Homme-Machine. 2015 Presented at: IHM'15; October 27-30, 2015; Toulouse, France. [doi: 10.1145/2820619.2825015]
21. Distler V, Lallemand C, Bellet T. Acceptability and Acceptance of Autonomous Mobility on Demand: The Impact of an Immersive Experience. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 21-26, 2018; Montreal, QC,. [doi: 10.1145/3173574.3174186]
22. Terrade F, Pasquier H, Reerinck-Boulanger J, Guingouain G, Somat A. [L'acceptabilité sociale: la prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques]. Le Travail Humain 2009;72(4):395. [doi: 10.3917/th.724.0383]
23. Arbelaez Garces G, Rakotondranaivo A, Bonjour E. An acceptability estimation and analysis methodology based on Bayesian networks. Int J Ind Ergon 2016;53:245-256. [doi: 10.1016/j.ergon.2016.02.005]
24. Reerink-Boulanger J. [Services technologiques intégrés dans l'habitat des personnes âgées: examen des déterminants individuels, sociaux et organisationnels de leur acceptabilité] (Doctoral dissertation). HAL Archives. 2012. URL: https://tel.archives-ouvertes.fr/tel-00666879/document [accessed 2020-06-02]
25. Overdijk M, Van Diggelen W. Technology appropriation in face-to-face collaborative learning. In: Proceedings of the European Conference on Technology Enhanced Learning. 2006 Presented at: EC-TEL'06; October 1-4, 2006; Crete, Greece p. 89-96. [doi: 10.1007/11876663_56]
26. Bar F, Weber MS, Pisani F. Mobile technology appropriation in a distant mirror: baroquization, creolization, and cannibalism. New Media Soc 2016 Feb 05;18(4):617-636. [doi: 10.1177/1461444816629474]
27. Carroll J, Howard S, Peck J, Murphy J. From adoption to use: the process of appropriating a mobile phone. Australasian J Inf Syst 2003;10(2):38-48. [doi: 10.3127/ajis.v10i2.151]
28. Karahanna E, Straub DW, Chervany NL. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. MIS Q 1999;23(2):183. [doi: 10.2307/249751]
29. Rogers E. Diffusion of Innovations. Toronto, ON: Simon and Schuster; 2010.

XSL•FO
RenderX

30. Davis F. A technology acceptance model for empirically testing new end-user information systems: Theory and results. USA: Massachusetts Institute of Technology; 1986.
31. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
32. Connelly K. On Developing a Technology Acceptance Model for Pervasive Computing. 9th International Conference on Ubiquitous Computing -Workshop of Ubiquitous System Evaluation (USE) New York, NY. In: In 9th International Conference on Ubiquitous Computing - Workshop of Ubiquitous System Evaluation.: Springer; 2007 Presented at: UBICOMP; September 2007; Innsbruck, Austria p. 520 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.587.422&rep=rep1&type=pdf
33. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177. [doi: 10.2196/mhealth.7886] [Medline: 29212629]
34. Cheung ML, Chau KY, Lam MH, Tse G, Ho KY, Flint SW, et al. Examining consumers' adoption of wearable healthcare technology: the role of health attributes. Int J Environ Res Public Health 2019 Jun 26;16(13):2257. [doi: 10.3390/ijerph16132257] [Medline: 31247962]
35. Doherty K, Doherty G. Engagement in HCI. ACM Comput Surv 2019 Jan 23;51(5):1-39. [doi: 10.1145/3234149]
36. Karapanos E, Hassenzahl M, Martens J. User Experience over Time. In: In CHI '08 Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2008 Presented at: CHI'08; May 2008; Florence, Italy p. 3561-3566. [doi: 10.1145/1358628.1358891]
37. Karapanos E, Zimmerman J, Forlizzi J. User Experience Over Time: an Initial Framework. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. NY, USA: ACM; 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA, US p. 729-738. [doi: 10.1145/1518701.1518814]
38. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
39. Dworkin M, Panchal P, Jimenez A, Garofalo R, Haberer J, Wiebel W. How acceptable is a wireless pill bottle that monitors and texts in response to missed doses: focus groups with young African American MSM living with HIV. Open Forum Infect Dis 2017;4:S422. [doi: 10.1093/ofid/ofx163.1060]
40. Lawson-Body A, Willoughby L, Lawson-Body L, Tamandja EM. Students' acceptance of E-books: an application of UTAUT. J Comp Inf Syst 2018 Apr 19;60(3):256-267. [doi: 10.1080/08874417.2018.1463577]
41. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
42. Allen I, Seaman J. Staying the Course: Online Education in the United States, 2008. Sloan Consortium. Newburyport, USA; 2008. URL: https://onlinelearningsurvey.com/reports/staying-the-course.pdf [accessed 2020-05-21]
43. Otto L, Harst L, Schlieter H, Wollschlaeger B, Richter P, Timpel P. Towards a unified understanding of ehealth and related terms-proposal of a consolidated terminological basis. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: HEALTHINF. 2018 Presented at: International Joint Conference on Biomedical Engineering Systems and Technologies; 2018; Funchal, Madeira, Portugal p. 533-539. [doi: 10.5220/0006651005330539]
44. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
45. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
46. Sharpe JD, Zhou Z, Escobar-Viera CG, Morano JP, Lucero RJ, Ibañez GE, et al. Interest in using mobile technology to help self-manage alcohol use among persons living with the human immunodeficiency virus: A Florida Cohort cross-sectional study. Subst Abus 2018 Jan 02;39(1):77-82 [FREE Full text] [doi: 10.1080/08897077.2017.1356793] [Medline: 28723300]
47. Abelson JS, Symer M, Peters A, Charlson M, Yeo H. Mobile health apps and recovery after surgery: what are patients willing to do? Am J Surg 2017 Oct;214(4):616-622. [doi: 10.1016/j.amjsurg.2017.06.009] [Medline: 28666581]
48. Gardner CL, Liu F, Fontelo P, Flanagan MC, Hoang A, Burke HB. Assessing the usability by clinicians of VISION: a hierarchical display of patient-collected physiological information to clinicians. BMC Med Inform Decis Mak 2017 Apr 14;17(1):41 [FREE Full text] [doi: 10.1186/s12911-017-0435-3] [Medline: 28410579]
49. Liu X, Wang R, Zhou D, Hong Z. Feasibility and acceptability of smartphone applications for seizure self-management in China: questionnaire study among people with epilepsy. Epilepsy Behav 2016 Feb;55:57-61. [doi: 10.1016/j.yebeh.2015.11.024] [Medline: 26745631]
50. Peterson CM, Apolzan JW, Wright C, Martin CK. Video chat technology to remotely quantify dietary, supplement and medication adherence in clinical trials. Br J Nutr 2016 Nov;116(9):1646-1655 [FREE Full text] [doi: 10.1017/S0007114516003524] [Medline: 27753427]

XSL•FO
RenderX

51. Helsel BC, Williams JE, Lawson K, Liang J, Markowitz J. Telemedicine and mobile health technology are effective in the management of digestive diseases: a systematic review. Dig Dis Sci 2018 Jun;63(6):1392-1408. [doi: 10.1007/s10620-018-5054-z] [Medline: 29663265]
52. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GY. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF app trial. Am J Med 2017 Dec;130(12):1388-1396 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
53. Westergaard RP, Genz A, Panico K, Surkan PJ, Keruly J, Hutton HE, et al. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addict Sci Clin Pract 2017 Apr 26;12(1):11 [FREE Full text] [doi: 10.1186/s13722-017-0076-y] [Medline: 28441962]
54. Schlosser DA, Campellone TR, Truong B, Anguera JA, Vergani S, Vinogradov S, et al. The feasibility, acceptability, and outcomes of PRIME-D: a novel mobile intervention treatment for depression. Depress Anxiety 2017 Jun;34(6):546-554 [FREE Full text] [doi: 10.1002/da.22624] [Medline: 28419621]
55. Gordon JS, Armin J, Hingle MD, Giacobbi P, Cunningham JK, Johnson T, et al. Development and evaluation of the See Me Smoke-Free multi-behavioral mhealth app for women smokers. Transl Behav Med 2017 Jun;7(2):172-184 [FREE Full text] [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
56. Hicks TA, Thomas SP, Wilson SM, Calhoun PS, Kuhn ER, Beckham JC. A preliminary investigation of a relapse prevention mobile application to maintain smoking abstinence among individuals with posttraumatic stress disorder. J Dual Diagn 2017;13(1):15-20 [FREE Full text] [doi: 10.1080/15504263.2016.1267828] [Medline: 27918881]
57. O'Brien KH, LeCloux M, Ross A, Gironda C, Wharff EA. A pilot study of the acceptability and usability of a smartphone application intervention for suicidal adolescents and their parents. Arch Suicide Res 2017;21(2):254-264. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
58. Rizvi SL, Hughes CD, Thomas MC. The DBT coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
59. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
60. Sureshkumar K, Murthy G, Natarajan S, Naveen C, Goenka S, Kuper H. Evaluation of the feasibility and acceptability of the 'care for stroke' intervention in India, a smartphone-enabled, carer-supported, educational intervention for management of disability following stroke. BMJ Open 2016 Feb 2;6(2):e009243 [FREE Full text] [doi: 10.1136/bmjopen-2015-009243] [Medline: 26839011]
61. Juarascio AS, Goldstein SP, Manasse SM, Forman EM, Butryn ML. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
62. Goldstein CM, Gathright EC, Dolansky MA, Gunstad J, Sterns A, Redle JD, et al. Randomized controlled feasibility trial of two telemedicine medication reminder systems for older adults with heart failure. J Telemed Telecare 2014 Sep;20(6):293-309 [FREE Full text] [doi: 10.1177/1357633X14541039] [Medline: 24958355]
63. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-125 [FREE Full text] [doi: 10.1177/1932296814556506] [Medline: 25361643]
64. Chen YS, Wong JE, Ayob AF, Othman NE, Poh BK. Can Malaysian young adults report dietary intake using a food diary mobile application? A pilot study on acceptability and compliance. Nutrients 2017 Jan 13;9(1):65 [FREE Full text] [doi: 10.3390/nu9010062] [Medline: 28098770]
65. Brittain K, Kamp K, Cassandras C, Salaysay Z, Gómez-Márquez J. A mobile app to increase informed decisions about colorectal cancer screening among African American and Caucasian women: a pilot study. Gastroenterol Nurs 2018;41(4):297-303. [doi: 10.1097/SGA.0000000000000319] [Medline: 30063684]
66. Niendam TA, Tully LM, Iosif A, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: a longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan;96:239-246. [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
67. Jacobson AE, Vesely SK, Haamid F, Christian-Rancy M, O'Brien SH. Mobile application vs paper pictorial blood assessment chart to track menses in young women: a randomized cross-over design. J Pediatr Adolesc Gynecol 2018 Apr;31(2):84-98. [doi: 10.1016/j.jpag.2017.09.009] [Medline: 29030160]
68. Eisenhauer CM, Hageman PA, Rowland S, Becker BJ, Barnason SA, Pullen CH. Acceptability of mhealth technology for self-monitoring eating and activity among rural men. Public Health Nurs 2017 Mar;34(2):138-146. [doi: 10.1111/phn.12297] [Medline: 27757986]
69. Isetta V, Torres M, González K, Ruiz C, Dalmases M, Embid C, et al. A new mhealth application to support treatment of sleep apnoea patients. J Telemed Telecare 2017 Jan;23(1):14-28. [doi: 10.1177/1357633X15621848] [Medline: 26672606]

XSL•FO
RenderX

70. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
71. Patel S, Jacobus-Kantor L, Marshall L, Ritchie C, Kaplinski M, Khurana PS, et al. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Technol 2013 May 1;7(3):630-649 [FREE Full text] [doi: 10.1177/193229681300700307] [Medline: 23759395]
72. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383 [FREE Full text] [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
73. Santo K, Chow CK, Thiagalingam A, Rogers K, Chalmers J, Redfern J. MEDication reminder APPs to improve medication adherence in coronary heart disease (MedApp-CHD) study: a randomised controlled trial protocol. BMJ Open 2017 Oct 8;7(10):e017540 [FREE Full text] [doi: 10.1136/bmjopen-2017-017540] [Medline: 28993388]
74. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
75. Hartzler AL, Venkatakrishnan A, Mohan S, Silva M, Lozano P, Ralston JD, et al. Acceptability of a team-based mobile health (mhealth) application for lifestyle self-management in individuals with chronic illnesses. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:3277-3281. [doi: 10.1109/EMBC.2016.7591428] [Medline: 28269007]
76. Sarzynski E, Decker B, Thul A, Weismantel D, Melaragni R, Cholakis E, et al. Beta testing a novel smartphone application to improve medication adherence. Telemed J E Health 2017 Apr;23(4):339-348. [doi: 10.1089/tmj.2016.0100] [Medline: 27564971]
77. Radhakrishnan K, Toprac P, O'Hair M, Bias R, Kim MT, Bradley P, et al. Interactive digital e-health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374 [FREE Full text] [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
78. Ahern DK, Parker D, Eaton C, Rafferty C, Wroblewski J, Goldman R. Patient-facing technology for identification of COPD in primary care. J Innov Health Inform 2016 Jul 15;23(2):541-546 [FREE Full text] [doi: 10.14236/jhi.v23i2.824] [Medline: 27869583]
79. Simmons ES, Paul R, Shic F. Brief report: a mobile application to treat prosodic deficits in autism spectrum disorder and other communication impairments: a pilot study. J Autism Dev Disord 2016 Jan;46(1):320-327. [doi: 10.1007/s10803-015-2573-8] [Medline: 26329637]
80. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
81. Muroff J, Robinson W, Chassler D, López LM, Gaitan E, Lundgren L, et al. Use of a smartphone recovery tool for Latinos with co-occurring alcohol and other drug disorders and mental disorders. J Dual Diagn 2017;13(4):280-290. [doi: 10.1080/15504263.2017.1348649] [Medline: 28692420]
82. Hickey AM, Freedson PS. Utility of consumer physical activity trackers as an intervention tool in cardiovascular disease prevention and treatment. Prog Cardiovasc Dis 2016;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
83. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
84. Ofili EO, Pemu PE, Quarshie A, Mensah EA, Rollins L, Ojutalayo F, et al. Democratizing discovery health with N=ME. Trans Am Clin Climatol Assoc 2018;129:215-234 [FREE Full text] [Medline: 30166716]
85. Noblin A, Shettian M, Cortelyou-Ward K, Schack Dugre J. Exploring physical therapists' perceptions of mobile application usage utilizing the FITT framework. Inform Health Soc Care 2017 Mar;42(2):180-193. [doi: 10.1080/17538157.2016.1178118] [Medline: 27259680]
86. Yasini M, Marchand G. Adoption and use of a mobile health application in older adults for cognitive stimulation. Stud Health Technol Inform 2016;221:13-17. [Medline: 27071867]
87. Ginsburg AS, Tawiah Agyemang C, Ambler G, Delarosa J, Brunette W, Levari S, et al. mPneumonia, an Innovation for Diagnosing and Treating Childhood Pneumonia in Low-Resource Settings: A Feasibility, Usability and Acceptability Study in Ghana. PLoS One 2016;11(10):e0165201 [FREE Full text] [doi: 10.1371/journal.pone.0165201] [Medline: 27788179]
88. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and acceptability of ecological momentary assessment of daily functioning among older adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840 [FREE Full text] [doi: 10.1016/j.jagp.2016.11.019] [Medline: 28065496]
89. Rico TM, dos Santos MK, Fernandes VP, Madruga SW, Noguez PT, Barcelos CR, et al. Text messaging (SMS) helping cancer care in patients undergoing chemotherapy treatment: a pilot study. J Med Syst 2017 Oct 9;41(11):181. [doi: 10.1007/s10916-017-0831-3] [Medline: 28990135]

XSL•FO
RenderX

90. Wang Y, Xiao Q, Sun L, Wu Y. Chinese nurses' acceptance of PDA: a cross-sectional survey using a technology acceptance model. Stud Health Technol Inform 2016;225:889-890. [Medline: 27332394]
91. DeVito DA, Song M, Myers B, Hawkins RP, Aubrecht J, Begey A, et al. Clinical trials of health information technology interventions intended for patient use: unique issues and considerations. Clin Trials 2013;10(6):896-906 [FREE Full text] [doi: 10.1177/1740774513493149] [Medline: 23867222]
92. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
93. Ammenwerth E, Woess S, Baumgartner C, Fetz B, van der Heidt A, Kastner P, et al. Evaluation of an integrated telemonitoring surveillance system in patients with coronary heart disease. Methods Inf Med 2015;54(5):388-397. [doi: 10.3414/ME15-02-0002] [Medline: 26395147]
94. Becker S, Brandl C, Meister S, Nagel E, Miron-Shatz T, Mitchell A, et al. Demographic and health related data of users of a mobile application to support drug adherence is associated with usage duration and intensity. PLoS One 2015;10(1):e0116980 [FREE Full text] [doi: 10.1371/journal.pone.0116980] [Medline: 25629939]
95. Wu X, Oliveria SA, Yagerman S, Chen L, DeFazio J, Braun R, et al. Feasibility and efficacy of patient-initiated mobile teledermoscopy for short-term monitoring of clinically atypical nevi. JAMA Dermatol 2015 May;151(5):489-496. [doi: 10.1001/jamadermatol.2014.3837] [Medline: 25629626]
96. Fallah M, Yasini M. A medication reminder mobile app: does it work for different age ranges. Stud Health Technol Inform 2017;235:68-72. [Medline: 28423757]
97. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
98. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
99. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mhealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
100. Schulte M, Liang D, Wu F, Lan Y, Tsay W, Du J, et al. A smartphone application supporting recovery from heroin addiction: perspectives of patients and providers in China, Taiwan, and the USA. J Neuroimmune Pharmacol 2016 Sep;11(3):511-522 [FREE Full text] [doi: 10.1007/s11481-016-9653-1] [Medline: 26846506]
101. Giménez-Pérez G, Recasens A, Simó O, Aguas T, Suárez A, Vila M, et al. Use of communication technologies by people with type 1 diabetes in the social networking era. A chance for improvement. Prim Care Diabetes 2016 Apr;10(2):121-128. [doi: 10.1016/j.pcd.2015.09.002] [Medline: 26428527]
102. Zhu Z, Liu Y, Che X, Chen X. Moderating factors influencing adoption of a mobile chronic disease management system in China. Inform Health Soc Care 2018 Jan;43(1):22-41. [doi: 10.1080/17538157.2016.1255631] [Medline: 28068149]
103. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mhealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
104. Farooqui N, Phillips G, Barrett C, Stukus D. Acceptability of an interactive asthma management mobile health application for children and adolescents. Ann Allergy Asthma Immunol 2015 Jun;114(6):527-529. [doi: 10.1016/j.anai.2015.03.006] [Medline: 26021895]
105. Khatun F, Heywood AE, Ray PK, Hanifi SM, Bhuiya A, Liaw S. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
106. Jongbloed K, Parmar S, van der Kop M, Spittal PM, Lester RT. Recent evidence for emerging digital technologies to support global HIV engagement in care. Curr HIV/AIDS Rep 2015 Dec;12(4):451-461 [FREE Full text] [doi: 10.1007/s11904-015-0291-7] [Medline: 26454756]
107. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974 Dec 1;2(4):354-386. [doi: 10.1177/109019817400200405]
108. Povey J, Mills PP, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for aboriginal and Torres strait islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
109. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 9;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]

XSL•FO
RenderX

110. Agarwal R, Prasad J. The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. Decis Sci 1997 Jul;28(3):557-582. [doi: 10.1111/j.1540-5915.1997.tb01322.x]
111. Hochstenbach LM, Zwakhalen SM, Courtens AM, van Kleef M, de Witte LP. Feasibility of a mobile and web-based intervention to support self-management in outpatients with cancer pain. Eur J Oncol Nurs 2016 Aug;23:97-105 [FREE Full text] [doi: 10.1016/j.ejon.2016.03.009] [Medline: 27456381]
112. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
113. Delone W, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inform Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
114. Matthews M, Voida S, Abdullah S, Doherty G, Choudhury T, Im S, et al. In Situ Design for Mental Illness: Considering the Pathology of Bipolar Disorder in mHealth Design. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. NY, USA: ACM; 2015 Presented at: CHI'17; May 2017; Denver, USA p. 86-97. [doi: 10.1145/2785830.2785866]
115. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support effective ML system design. ACM Trans Comput Hum Interact 2020:-. [doi: 10.1145/3398069]

Abbreviations

FITT: fit between individuals, task, and technology
HCI: human-computer interaction
HITAM: health information technology acceptance model
MeSH: Medical Subject Headings
mHealth: mobile health
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability framework
PDA: personal digital assistant
TAL: Technology Acceptance Lifecycle
TAM: technology acceptance model

Edited by G Eysenbach; submitted 29.11.19; peer-reviewed by M Bellio, A Graham; comments to author 31.01.20; revised version received 14.04.20; accepted 30.04.20; published 06.07.20.

Please cite as:
Nadal C, Sas C, Doherty G
Technology Acceptance in Mobile Health: Scoping Review of Definitions, Models, and Measurement
J Med Internet Res 2020;22(7):e17256
URL: https://www.jmir.org/2020/7/e17256
doi:10.2196/17256
PMID:32628122

©Camille Nadal, Corina Sas, Gavin Doherty. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Acceptability and Usability of Digital Health Interventions for Adults With Depression, Anxiety, and Somatoform Disorders: Qualitative Systematic Review and Meta-Synthesis

Shireen Patel[1], BSc, MSc; Athfah Akhtar[2], PhD; Sam Malins[1], PhD, ClinPsy; Nicola Wright[3], PhD; Emma Rowley[4], PhD; Emma Young[5], MSc; Stephanie Sampson[6], LLM, MA; Richard Morriss[1,5,6], ChB, MD

[1]Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, United Kingdom
[2]School of Social Sciences, Birmingham City University, Birmingham, United Kingdom
[3]Faculty of Medicine, School of Health Sciences, University of Nottingham, Nottingham, United Kingdom
[4]School of Medicine, University of Nottingham, Nottingham, United Kingdom
[5]Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, United Kingdom
[6]NIHR MindTech MedTech Co-operative, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Shireen Patel, BSc, MSc
Division of Psychiatry and Applied Psychology, School of Medicine
University of Nottingham
Institute of Mental Health Building, University of Nottingham Innovation Park
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 115 8231434
Email: shireen.patel@nottingham.ac.uk

Abstract

Background: The prevalence of mental health disorders continues to rise, with almost 4% of the world population having an anxiety disorder and almost 3.5% having depression in 2017. Despite the high prevalence, only one-third of people with depression or anxiety receive treatment. Over the last decade, the use of digital health interventions (DHIs) has risen rapidly as a means of accessing mental health care and continues to increase. Although there is evidence supporting the effectiveness of DHIs for the treatment of mental health conditions, little is known about what aspects are valued by users and how they might be improved.

Objective: This systematic review aimed to identify, appraise, and synthesize the qualitative literature available on service users' views and experiences regarding the acceptability and usability of DHIs for depression, anxiety, and somatoform disorders.

Methods: A systematic search strategy was developed, and searches were run in 7 electronic databases. Qualitative and mixed methods studies published in English were included. A meta-synthesis was used to interpret and synthesize the findings from the included studies.

Results: A total of 24 studies were included in the meta-synthesis, and 3 key themes emerged with descriptive subthemes. The 3 key themes were initial motivations and approaches to DHIs, personalization of treatment, and the value of receiving personal support in DHIs. The meta-synthesis suggests that participants' initial beliefs about DHIs can have an important effect on their engagement with these types of interventions. Personal support was valued very highly as a major component of the success of DHIs. The main reason for this was the way it enabled individual personalization of care.

Conclusions: Findings from the systematic review have implications for the design of future DHIs to improve uptake, retention, and outcomes in DHIs for depression, anxiety, and somatoform disorders. DHIs need to be personalized to the specific needs of the individual. Future research should explore whether the findings could be generalized to other health conditions.

(J Med Internet Res 2020;22(7):e16228) doi:10.2196/16228

KEYWORDS

digital health interventions; depression; anxiety; somatoform disorders; smartphone; mobile phone

XSL•FO
RenderX

Introduction

Background

The prevalence of mental health disorders continues to rise [1]. It is estimated that globally in 2017, 264 million people (3.4%) experienced depression, and 284 million (3.7%) of the population experienced an anxiety disorder. This included phobias, obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), or generalized anxiety disorder (GAD) [2]. The worldwide prevalence of somatoform disorders or Medically Unexplained Symptoms in Primary Care is 25%-50% [3]. From an economic and social perspective, the costs of mental health disorders are high. The total costs of mental ill health are estimated at more than 4% of the gross domestic product, or over €600 (US $650.31 million), across the 28 EU countries [4]. More than two-third of individuals with depressive disorders also have anxiety symptoms, and 40%-70% simultaneously meet the criteria for at least one type of anxiety disorder [5,6]. Depression, anxiety, and somatoform disorders are often comorbid, and treatment can reduce symptomology in all 3 areas [7], implying that treatments could be effective across comorbid mental health conditions. Furthermore, where large-scale mental health interventions have been implemented, there is evidence of coaction—where a change in one disorder is achieved, it is more likely that change will be achieved with other health disorders [8].

Despite the high prevalence, only 33% of adults with depression and anxiety receive treatment in England [9], indicating that few people who need treatment receive it. The barriers to accessing psychological therapy include a shortage of therapists, long waiting times, and the stigma of accessing psychological treatment [10]. A recent authoritative review identified the use of digital health interventions (DHIs) as central to addressing population-level mental health. The absence of widely disseminated, accessible mental health interventions has also been identified as a key reason for the lack of improvement in mental health treatment [8]. DHIs have the potential to improve accessibility and meet the requirements expected from mental health services [10].

DHIs can be defined as interventions that provide information, therapy, and support (emotional, decisional, or behavioral) for physical or mental health disorders via a technological or digital platform (website, computer, mobile phone app, SMS text message, email, videoconferencing, wearable device, patient portals, or personal health records or virtual reality [VR]) [11,12]. The variety and functionality of digital health technologies continue to evolve. The forms of DHIs include computerized cognitive behavioral therapy (cCBT) but can also include telehealth or telemedicine utilizing telecommunications processes such as SMS text messaging, email, and videoconferencing for remote delivery of therapy. Recently, there has also been the emergence of *mHealth* or mobile health, which incorporates smartphone apps, remote monitoring, and wearables. DHIs can be standalone interventions, used as a stepped care model, provided in addition to face-to-face treatment or mediated by a health care professional such as a therapist [13,14].

Numerous systematic reviews have found that for people with depression and anxiety disorders, DHIs can be just as effective as face-to-face therapy [15,16]. However, dropout rates for completion of computerized psychological therapy are high [17,18]. Waller and Gilbody [17] conducted a systematic review of the quantitative and qualitative evidence identifying the barriers to the uptake of cCBT. They concluded that, a median of, 38% (range 4%-84%) of participants who were recruited started cCBT, and a median of only 56% (range 12%-100%) of starters completed it. They concluded that these low figures were influenced more by personal circumstances than technological aspects. Waller and Gilbody [17] acknowledged that more data was needed to explore why so few started or discontinued therapy. The inclusion of therapist support and guidance has been found to improve treatment completion rates and increase treatment effect sizes [19-21], implying that therapist-supported DHIs rather than self-guided DHIs may lead to more positive experiences for participants. The concept of blending face-to-face therapy with computer-delivered therapy or via videoconferencing has recently begun to emerge, which attempts to address some of the barriers to engaging with DHIs [22-24]. Furthermore, a recent article [25] identified that the top 10 priorities for DHIs in mental health include considering how DHIs could be combined with human support to improve its effectiveness.

Existing qualitative systematic reviews provide insight into the potential barriers and facilitators of recruiting and retaining participants in DHI research trials. O'Connor et al [26] investigated the factors affecting patient and public engagement and recruitment to DHIs. They identified that there was a need for greater awareness and understanding about DHIs to improve public understanding of DHIs [26]. They also recommended that DHIs need to incorporate some kind of social interaction to improve recruitment. DHIs also need to be tailored to individual needs. Knowles et al [27] explored user experience of digital technologies as a low-intensity intervention delivered with minimal or no professional support for depression and anxiety, concluding that personalization and tailoring content toward users' needs improved the experience of computerized therapy. Furthermore, they recommended for future research to explore if these findings could be extrapolated to other delivery formats and treatment modalities.

Given the rise in prevalence of depression, anxiety, and somatoform disorders and advances in the development of DHIs, but low uptake and completion rates, it is important to conduct a systematic review exploring perceived barriers and facilitators to the acceptability and usability of a range of DHIs. This systematic review aimed to build on previous literature to explore the diverse nature of DHIs and whether potential facilitators and barriers are consistent across different types and formats of DHIs. We wanted to investigate whether there were specific issues relevant to the integration of support compared with self-guided DHIs and whether varying levels of support, intensity, and delivery formats influenced patient experience.

This is the first meta-synthesis to our knowledge that will look at the acceptability and usability of DHIs across different types and formats and across a range of depression and anxiety disorders. This systematic review includes all DHIs regardless

XSL•FO
RenderX

of the delivery method (ie, text-based, automated, blended therapy). It compares very different experiences to consider whether therapies involving digital aspects share common issues and if consistent themes are found across different formats and modes of delivery. A systematic review of the qualitative literature may provide further explanation and synthesis of factors that may facilitate adherence and outcome of DHIs for comorbid mental health conditions. The key findings of this systematic review will inform the design of mental health digital interventions. This will support the development of digital mental health interventions tailored to service users' needs and thus improve uptake, adherence, and experience, which will have a positive impact on service users and providers.

Objectives of the Systematic Review

This systematic review aimed to understand the experiences of service users with regard to DHIs for depression, anxiety, and somatoform disorders. Specific objectives were to systematically identify, appraise, and meta-synthesize available qualitative literature that explored the following:

1. Service user's perceptions of DHIs regarding their acceptability and usability.
2. Aspects of DHIs that are valued and work well and those that could be improved or altered.
3. Why service users chose to use, continue with, or stop using DHIs.

Methods

The protocol for the systematic review was published on the International Prospective Register of Systematic Reviews (ID: CRD42018104016) before completion of the systematic review. The systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [28] statement checklist and Transparency in Reporting the Synthesis of Qualitative Research [29] guidelines (Multimedia Appendices 1 and 2).

Search Strategy

A comprehensive search strategy was developed with an expert health information specialist (EY). A scoping search was conducted to identify key papers and associated search terms to inform the design of the search strategy. A systematic search was conducted for published papers that contained qualitative information about service user experiences of participating in a DHI trial for depression, anxiety, or somatoform disorders. A combination of free terms and controlled vocabulary terms was used to ensure that all relevant studies were identified. Search terms were split into 2 categories: DHIs and mental health conditions (depression, anxiety, and or somatoform disorders). Qualitative search filters identified by the InterTASC Information Specialists Sub-Group were also applied [30]. The search strategy was developed for MEDLINE (Multimedia Appendix 3) and then adapted and applied to PsycINFO, Cumulative Index for Nursing and Allied Health Literature (CINAHL), EMBASE, ISI Web of Science, Scopus, and the Cochrane Library. We searched the electronic databases for eligible studies from database inception to July 17, 2018. The search was conducted by EY. From the search results, papers that were published in the same study but in different publications were not removed to determine more closely if they reported different data aspects of the research. Search results were uploaded and stored using EndNote version X8.

Inclusion and Exclusion Criteria

Textbox 1 describes the inclusion and exclusion criteria.

XSL•FO
RenderX

Textbox 1. Inclusion criteria and exclusion criteria.

Inclusion criteria

- English language
- Community, primary, and secondary care
- Original qualitative studies, studies involving secondary analysis of qualitative data, or qualitative studies that were part of a mixed methods study (eg, the study also had a quantitative component, but the major component was qualitative, and a qualitative methodology was described). Studies should include a substantial amount of qualitative methods, including interviews, observations, and open-ended evaluation forms. Free textboxes on evaluation forms were included if there was richness in the data provided (ie, sufficient quotes to support the analysis)
- Papers should include some form of qualitative data analysis such as thematic or inductive analysis
- Papers reporting on participants who had experienced the use of digital health interventions (DHIs, also called *internet interventions* or *electronic health interventions*), where the DHI was primarily used to treat depression, anxiety, and/or somatoform disorders. This included interventions that provided information and support (emotional, decisional, and/or behavioral) via a technological or digital platform (website, computer, mobile phone app, SMS text message, email, videoconferencing, wearable device, patient portals or personal health records, or virtual reality)

Exclusion criteria

- Gray literature or literature not published in a peer reviewed journal
- Dissertation or theses
- Published abstracts or conference proceedings
- Any type of literature review, systematic review, and meta-synthesis
- Experiences of health care professionals or parents or carers
- Studies where the primary digital health interventions (DHIs) were telephone-based with no additional technological function (eg, telephone counseling or triaging service), internet-based health tools that were not defined as interventions (eg, internet health searching), or an implantable device that was remotely monitored
- Interventions to improve adherence to medication, improve assessment or diagnosis, or where digital interventions were not the major constituent of the intervention
- Peer-to-peer networks and DHIs of social support via the internet, use of social media, online support groups, or DHIs consisting of group therapy
- Data collected during the testing of the usability and design of DHIs
- Males and females aged <18 years; studies were included if ≥50% of the sample were aged ≥18 years

Screening and Data Extraction

Data Screening

Duplicated studies were removed in EndNote, and the remaining articles were exported into a Microsoft Excel document for screening. Qualitative results from the same overall study that were split across different publications were not removed until the full-text screening. Papers were removed if they reported the same qualitative findings; otherwise, they were included. In addition, we carried out backward citation by hand searching the references of all included studies and key papers. Forward citation searching of the included studies was carried out in Scopus. Initial screening of titles and abstracts was conducted by 1 reviewer (SP), with a second reviewer (AA) screening a random 10% sample to confirm congruence. Interrater agreement for full-text screening was 99.4%, and disagreements were resolved through discussion.

Data Extraction

Data were extracted by SP from papers included in the systematic review using a data extraction form developed by SP and AA. The primary focus of data extraction was the identification of specific qualitative findings—reported themes and subthemes related to the phenomena of interest, which were subsequently synthesized as described below. All text from the papers labeled as results or findings were entered into a Microsoft Word document. Additionally, descriptive data, including details about DHIs, study aims, methods and analysis, country of research, and demographics of participants, were extracted. The form was initially piloted on 3 papers by both reviewers. AA completed data extraction for 30% of the articles to confirm congruence. Data extraction forms were compared to ensure data accuracy and comprehensiveness. Any disagreements were resolved through discussion until a consensus was reached.

Quality Appraisal

Quality assessment of papers included in the meta-synthesis was undertaken by SP and AA using the Critical Appraisal Skills Programme (CASP) criteria [31]. Studies were not excluded on the basis of quality because we were more concerned about including papers that contained depth in data collection and analysis, which might provide valuable information regarding participant experience of DHIs. We included papers that collected data through semistructured interviews. We included papers that collected data through open response text if there was richness in the data provided.

Meta-Synthesis

A meta-synthesis approach was used to organize and interpret the data. The findings of the included studies were synthesized

XSL•FO
RenderX

using methods proposed by Noblit and Hare [32]. Papers were read and re-read by SP and AA, and first- and second-order constructs were extracted using a Microsoft Word template form. First-order constructs were defined as direct participant quotes reported in the papers. Second-order constructs were defined as the authors' interpretations of participants' quotes expressed as themes extracted from both the results and findings sections of papers. SP and AA independently sifted the second-order constructs, compiling new third-order constructs that summarized and encompassed the various themes across studies. Third-order constructs refer to synthesized constructs that emerge from the analysis of first- and second-order constructs. Constructs were reviewed to see how the themes were similar when compared across papers to make sense of the variability in participant experience of DHIs. A draft summary of the analytical themes was written by SP and AA and shared with the review team consisting of service providers and service user representatives. Themes were refined until consensus was reached.

Results

Summary of Search Results

The initial search yielded 9030 records. After duplicate records were removed, there were a total of 8936 records. Following title and abstract screening, 8836 articles were excluded, resulting in a total of 100 papers being reviewed in full; 1 paper that was not identified in our search but included in a key systematic review paper [27] was included [33], and 1 paper was identified and included through hand searching [34]. A total of 26 papers met the eligibility criteria and were included in the meta-synthesis. These were amalgamated into 24 studies because 2 of the included studies reported data in 2 papers. One study that was included was reported in a qualitative paper as part of a Health Technology Assessment report, the themes highlighted in the 2 papers differed, and both findings were included [35,36]. For another study that was reported in 2 papers, 1 paper specifically focused on motivation to persist with the DHI, while another paper focused on patient experience and the implementation of digitally delivered cognitive behavioral therapy (CBT) [37,38].

The majority of the studies were nested within a randomized controlled trial (RCT; n=15). All the included studies were deemed to be of sufficient quality to contribute to the meta-synthesis. All papers reported a clear statement of the aims of the research and were deemed to be valuable in their contribution to the themes. Question 6 of the CASP referring to the relationship between the researcher and participants was acknowledged in only 13 papers. Kuhn et al [39] contributed the least to the meta-synthesis because the paper provided minimal information regarding the data analysis method and how themes were derived (Multimedia Appendix 4).

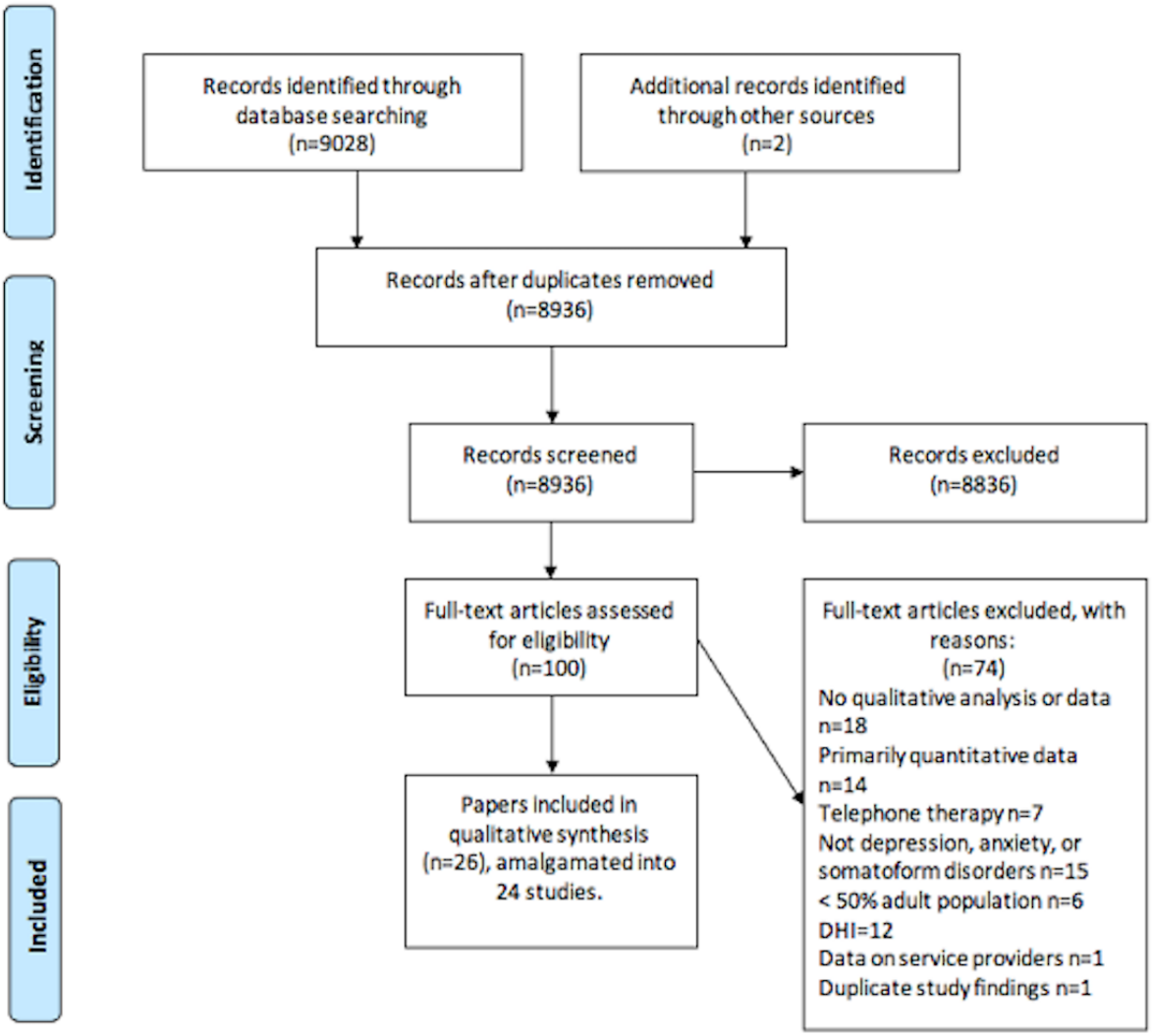
The PRISMA diagram (Figure 1) illustrates the flow of study identification and selection. The main reasons for exclusion included no qualitative analysis or data, primarily quantitative data, primary focus not being depression, anxiety, or somatoform disorders, and lack of data on the experience of using a DHI. Multimedia Appendix 5 summarizes the included studies. The sample characteristics of the included studies are described in Multimedia Appendix 6.

The included studies were published between 2007 and 2018. The majority of the studies were carried out in European countries, primarily England and Sweden (n=11). A total of 13 studies looked at all types of depressive disorders, including major depression and dysthymic disorder, postpartum depression, and studies where depression was comorbid with cardiovascular disease and multiple sclerosis; 6 studies examined anxiety disorders, including panic disorder, PTSD, OCD, GAD, and postpartum anxiety; and 4 studies examined depression and/or anxiety. No qualitative studies have been conducted on the use of a DHI for the treatment of somatoform disorders. The majority of the participants were recruited from the community, and 1 study recruited participants from a multiple sclerosis outpatient clinic. Only 11 papers reported the ethnicity of participants. The participants in the studies were primarily of a white ethnic background and younger or middle-aged. The majority of the studies collected data via interviews, and 2 studies collected data via open-ended questionnaires. Of the total, 18 studies were purely qualitative studies, and 6 studies were mixed methods studies.

The studies varied in types and formats of DHIs. Of the 24 studies, the majority provided additional support via email, telephone calls, or SMS text messages (n=13); 6 studies included some form of face-to-face support, whereas 5 studies were purely self-guided DHIs with no support provided. In relation to platforms, the majority of the studies were DHIs accessible on desk-based computers (n=19). Only 1 study used a blended approach consisting of video-based therapy. Of the eligible studies, 2 studies consisted of smartphone apps, and 1 study consisted of web- and smartphone-based monitoring. Only 1 study included the use of a computer telephony system designed to monitor and support self-care. In terms of treatment approach, the majority of studies were based on CBT principles (n=19).

XSL•FO
RenderX

Figure 1. Flow diagram of study identification and selection adapted from Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA).



Results of Meta-Synthesis

The meta-synthesis revealed 3 major themes and 9 subthemes related to the 3 objectives of the systematic review:

1. Service user perceptions of DHIs regarding their acceptability and usability.
2. Aspects of DHIs that are valued and work well and those that could be improved or altered.
3. Why service users chose to use, continue with, or stop using DHIs.
4. Selected supporting quotations from across the studies is provided in Multimedia Appendix 7 to illustrate this meta-synthesis. It shows first-, second-, and third-order constructs and synthesized subthemes with contrasting positive and negative participant experiences of DHIs.

The first theme, *initial motivations and approaches to DHIs*, had 2 subthemes:

1. Initial motivations: hope, accessibility, and cynicism.
2. Participant approaches to engaging with a DHI: active versus passive.

The second theme, *personalization of treatment*, had 3 subthemes:

1. Flexibility and autonomy.
2. Stigma and privacy.
3. Functionality, content, and interface.

The third theme, *the value of receiving personal support in DHIs*, had 4 subthemes:

1. Support for understanding DHIs.
2. Support to enhance commitment and motivation.
3. Suitability and desire for additional support.

XSL•FO
RenderX

4. Support to develop a virtual therapeutic relationship.

Theme 1: Initial Motivations and Approaches to Digital Health Interventions

Participants' initial motivations and approaches to engaging with DHIs had a significant impact on the perceived helpfulness of the intervention. Those who approached the DHI with a sense of hope that it might be helpful and had an active, committed approach to see the treatment through had more positive experiences of treatment and reported greater benefits than those who were initially more cynical and utilized a passive approach in their engagement with DHIs. Thus, the participant approach to DHIs impacted the recruitment and therapeutic process. This began when participants contemplated participating in the study and continued throughout their engagement with the DHI. This is explored further in the following subthemes.

Initial Motivations: Hope, Accessibility, and Cynicism

A total of 15 papers reported on participant expectations of participating in a DHI trial [34-48]. Participants initially decided to engage with a DHI for several reasons. These included hope for recovery and the desire to improve health and reduce symptoms through self-management. The prospect of using DHIs encouraged participants to feel empowered and manage their health by taking responsibility. Participants highlighted that participating in a DHI would enable them to develop coping strategies to manage their difficulties and increase their self-efficacy. DHIs provided a sense of agency to move from a passive to an active role in managing their condition.

DHIs were also viewed as novel approaches to treatment, which provided an alternative opportunity to receive help. DHIs were perceived to be more easily available because they increased accessibility, flexibility, and choice in accessing therapy:

> *I think it's an easy way to access help without having to like – you know it's easy, it fits into your lifestyle, it's convenient.* [48]

There were some negative expectations about DHIs, including skepticism about its helpfulness and concerns about whether a therapeutic relationship could be established remotely. However, in some cases, the ambivalence was overturned once participants commenced treatment, and there was surprise at how quickly a relationship could be established remotely.

Participant Approaches to Engaging With a Digital Health Intervention: Active Versus Passive

Participants' approach to DHIs and technology in general affected motivation to continue engaging with DHIs. This was reported in 17 papers [24,34,37-39,41,44-46,48-55]. Participants who took a more active approach could see the unique benefits of DHIs over medication or face-to-face therapy:

> *Rather than just saying well here's your pills or sit there and talk to somebody for 35 minutes…actually felt like I was doing something to help myself.* [54]

Participants with an active approach embraced independent work. This involved actively processing information received (eg, taking time to reflect on the sessions), educating themselves about their condition, and applying the learning to their daily living.

Engaging with the DHI gave participants a sense of empowerment, understanding, and awareness about their condition and its triggers, which encouraged them to utilize the tools for self-management. It gave participants a sense of accomplishment and provided greater understanding, and participants felt "inspired to take more control" [38].

This theme was strongly represented in Bendelin et al [49], who identified that an active, self-reliant approach to treatment was related to outcomes that were more positive. However, participants who had a passive style of working struggled to apply the treatment and were more likely to discontinue treatment.

With regard to completing therapy, participants with an active approach felt a sense of obligation or personal commitment to complete the therapy because they had agreed to participate in the DHI and owed it to the researcher or research team to complete the treatment. Others reported that they completed sessions because they valued the importance of research.

Participants with a more passive approach, however, struggled to maintain motivation. They found the nature of DHIs to be "quite difficult," "quite stressful" [52], and isolating. They preferred face-to-face sessions and felt that this was an essential component of personalized practical and emotional support.

Theme 2: Personalization of Treatment

The degree and ways in which DHIs were personalized to participants' situations and health status were deemed to impact the value of the treatment. The flexibility and convenience of DHIs had differential effects. For some participants, this made it more accessible and possible for them to engage in treatment in a way that traditional approaches could not. However, for others, the lack of structure, protected time, and accountability, present in more formal face-to-face therapy, meant that they forgot to complete sessions or disengaged from DHIs. Stigma and privacy were also a double-edged sword: for some, the anonymity of DHIs helped them to trust the process and engage; for others, the lack of a separate, private space to engage with difficult issues felt unsafe. There was broad agreement that DHIs with a simple interface and succinct content were preferred. There was also consensus that reminders, feedback on progress, and acknowledgment of achievements helped to support engagement.

Flexibility and Autonomy of the Digital Health Interventions

Flexibility and autonomy were emphasized in the majority of papers, with 19 papers reporting the flexible and autonomous nature of DHIs [24,35-42,45-49,52-56]. Some participants perceived DHIs to be more accessible and flexible. This led participants to feel that they had more autonomy and choice over their treatment. They used DHIs more responsively when they needed it, and this positively impacted treatment completion:

> *I like the fact that I could do it in my own time at home...cause I have three children so it wasn't like I*

XSL•FO
RenderX

> *would have to try to have to make appointments and get child care so I could do it when they were in bed or you know whenever it sort of seemed to fit in with my lifestyle I suppose.* [56]

Conversely, for some participants, DHIs lacked the structure and protected the treatment time they wanted, which subsequently impacted the motivation to complete treatment. Where interventions were self-guided and did not include monitoring, participants felt less obligated to complete sessions, particularly if they had competing priorities.

Stigma and Privacy

DHIs appealed to some participants because they were perceived to reduce the stigma and anxiety associated with seeking psychological help for mental health conditions. This was reported in 13 papers [35,36,41,42,44,48,49,51-53,55-57]. For participants who had not accepted their condition or felt afraid to express their thoughts, DHIs provided a safe platform to access support from the comfort of their own home. Participants felt more comfortable expressing their feelings on a computer rather than face-to-face because the DHI felt more private. Participants felt less judged and were able to be more honest and open in expressing their feelings to a computer rather than a person. They liked that they "didn't have to tell anybody else face to face," and found the DHI to be "a way of coping privately but in a structured way"' [36]. However, some participants had concerns about the security and privacy of data [51]. Some believed that a trusting and therapeutic relationship could only be formed face-to-face.

Functionality, Content, and Interface

There was great variability in the DHIs used, consisting of different interventions and varying levels and forms of support. Themes related to DHI functionality, content and interface were highlighted in 21 papers [24,34-39,41-48,50-52,55-57]. Participants reported that content simplicity, reminders, and progress monitoring were very important aspects of functionality, the absence of which impacted treatment completion and satisfaction. This is because it influenced user identification with the material and provided feedback. Accessibility on a range of platforms, content relevance, and ease of navigation, readability, and inclusion of interactive elements impacted user acceptability and engagement with DHIs.

Theme 3: The Value of Responsive Personal Support

This theme was identified in 24 of 25 papers. Only Kuhn et al [39] did not report this. Participants were able to seek treatment to help them self-manage their symptoms via the use of DHIs, but they still valued some form of human, responsive, personal support even if this support was not provided face-to-face. The key elements of additional support valued by participants were support that was personal and human and support that was rapidly responsive to their emotional state, personal difficulties, and achievements. Participants identified that additional support in DHIs helped them better understand DHIs, increased commitment and motivation, and helped form a more therapeutic engagement with DHIs. The rapidly responsive contact with a supporting person or therapist seemed to be missing from those who had poorer experiences of DHI. The presence and value of the provision of some form of personal support were identified to be integral in the majority of studies and forms the most influential theme.

Support for Understanding Digital Health Interventions and Treatment

Incorporating some form of support in DHIs aided participants' understanding of the purpose of the intervention. This was particularly pertinent where participants were ambivalent about participating in a DHI or were unsure about the need or value of receiving psychological support. This was emphasized in 16 papers [24,35-38,40,44-48,51,52,54,55,57]. Where support was not provided, participants misunderstood the difference between a research trial and the DHI and would often assume that trial participation was part of therapy. Guided support provided participants with direction about the interpretation of the treatment session content and made therapy more personally relevant.

Support for Enhancing Commitment and Motivation

Incorporating some form of support to enhance commitment and motivation was highlighted in 17 papers [35,37,38,41,43-51,53-55,57]. Owing to the autonomous nature of DHIs, participants reported forgetting or feeling less obligated to engage in treatment compared with face-to-face therapy, as highlighted in the previous theme. Without additional support, they struggled to relate to and apply the therapy content to their own condition, leading to disengagement from the DHI. Receiving feedback from a therapist or others allowed participants to monitor their progress, prevented forgetfulness, and encouraged participants to continue with therapy. Some form of communication was helpful and was achieved via a number of mediums, including face-to-face and remotely via emails, phone calls, and SMS text messages. Thus, receiving support facilitated understanding of symptoms, prevented forgetfulness to complete modules, and provided encouragement to overcome challenges and reduced isolation. Disengagement from a DHI was more likely in the absence of support as it reduced commitment and motivation to complete therapy.

> *From my point of view, the contact with the therapist was an essential aspect of therapy. Therefore, I lost all my interest in the therapy and didn't want to continue.* [43]

Suitability and Desire for Additional Support

Questions regarding the suitability of DHIs for some problems were raised in 12 papers, alongside some patients' desire for additional responsive support when DHIs became challenging or unsuitable [24,34,43,45,46,50-56]. DHI therapy sessions could be physically and mentally exhausting, and for some participants, it exacerbated symptoms of low mood and anxiety. Therefore, some participants wanted additional support to manage these negative feelings. The absence of support made module completion overwhelming for some participants, leading them to prioritize other commitments or discontinue the treatment.

XSL•FO
RenderX

Hind et al [52], focusing on DHIs for people with depression and physical comorbidities, found that completion of DHI sessions placed physical demands on participants. Where support was not provided, some form of support was recommended to overcome feelings of isolation and enable emotional expression [34,51,52]. This subtheme was not highlighted by any of the papers that included face-to-face support [24,33,35-38,44,57].

Support for Developing a Virtual Therapeutic Relationship

The interpersonal and relational aspects remained an essential ingredient of therapy, even if it was delivered as a DHI. This subtheme was reported in 18 papers [24,33-35,37,38,41,43-48,50,51,53-55]. Participants who engaged in DHIs reported feeling surprised at how quickly a relationship could be formed remotely with a person. They valued expressing feelings in written form because it enabled self-reflection and communication of emotion without interruptions. Attributes associated with developing a therapeutic relationship face-to-face were also identified in the DHIs. This included building a positive relationship between trust and feeling understood. Participants referred to the therapist as "a real friend" and likened the DHI to a "face-to-face session" [24].

Participants who disengaged from DHIs found them to be impersonal and expressed a preference for face-to-face therapy. The absence of visual cues such as eye contact and gestures was perceived to reduce emotional closeness, and made participants question whether the therapist was giving them their undivided attention. The use of written communication methods and the associated time delay between responses were seen as barriers to developing a therapeutic relationship. The absence of face-to-face contact was perceived to lack empathy, be machine-like "cold, not very friendly" [35], and this negatively impacted the therapeutic relationship.

Summary of Meta-Synthesis

The themes emerging from the meta-synthesis acknowledge the variability within individuals' experiences of DHIs for depression, anxiety, and somatoform disorders. DHIs were perceived both positively and negatively by participants, and this was influenced by the participants' expectations and needs. Personalization was an overarching theme; participants who had a preference for flexibility and autonomy perceived these to be the benefits of DHIs. In addition, the functionality of the DHIs was also important to participants. The key means of personalizing treatment identified as helpful by participants was the addition of rapid, responsive personal or human support. Thus, DHIs need to be responsive to participant preferences and experiences.

Discussion

Principal Findings

This systematic review aimed to meta-synthesize qualitative studies exploring the views of people who had been invited to participate in DHIs for depression, anxiety, and somatoform disorders. The first aim of the review was to examine service user perceptions of DHIs concerning their acceptability and usability. Findings from the meta-synthesis indicate that the acceptability and usability of DHIs differ on the basis of initial personal perceptions of DHIs, associated motivations, and the degree of responsive support offered. Participants' perceptions of DHIs can be positive or negative depending on expectations, preferences, and approaches to DHIs. Some participants felt more comfortable expressing their feelings on a computer rather than sharing it with somebody face-to-face because the DHI felt more private. Other participants found the nature of DHIs to be impersonal and preferred face-to-face therapy. This did not differ by DHI type or format, suggesting that the acceptability and usability of DHIs is influenced by individual perceptions and preferences. This study highlights that therapeutic work in DHIs is a dynamic process and is perceived positively or negatively depending on how well the DHI is adapted to the participants' preferences. The personalization of DHIs was an overarching theme, implying that DHIs need to consider individual preferences, circumstances, and needs to improve therapy completion rates for DHIs.

The second objective of the review was to examine aspects of DHIs that are valued and work well and those that could be improved or altered. The third aim of the review was to explore why service users chose to use, continue with, or stop using DHIs. This meta-synthesis emphasizes the significance of receiving personal support in DHIs and is consistent with the findings of O'Connor et al [26] and Knowles et al [27], both highlighted the need for personalization and availability of support in DHIs. The meta-synthesis also informs some of the research priorities identified by Hollis et al [25]. This study highlights that the suitability of DHIs is based on differing needs and that DHIs could be optimized by incorporating additional support. This systematic review extends on previous findings of Knowles et al [27] by demonstrating that personalized support was valued across studies irrespective of DHI type, format, or disorder. Participants who received self-guided DHIs expressed dissatisfaction with the lack of human interaction and expressed that it was required to increase commitment as it enabled personal support and feedback. Participants who received face-to-face contact as part of the DHI expressed that incorporating interpersonal features such as the provision of support was central because it personalized therapy. Personalized support facilitated an understanding of therapy, increased commitment and motivation to continue treatment, and helped form a more therapeutic engagement with DHIs. Similarly, participants in DHIs consisting of email or phone support reiterated that the presence of a supporter personalized therapy. However, they also expressed a desire for more contact with a supporter. Participants in the 2 studies that provided no therapy and only technical support or reminders [50,54] expressed a need for personalization in the form of feedback and emotional support. The functionality of DHIs was perceived to facilitate or hinder engagement. DHIs that were easily accessible and interactive were viewed as more beneficial than DHIs that were harder to navigate and inaccessible on a range of platforms. DHIs consisting of smartphone apps were perceived to be easily accessible. A simple interface and succinct content with reminders, feedback on progress, and acknowledgment of achievements also helped to support the completion of therapy.

XSL•FO
RenderX

This review also highlights that the flexible and autonomous nature of DHIs and the desire to gain greater awareness to self-manage conditions can affect participants' choice of whether to engage with a DHI or not. These aspects are consistent with a previous systematic review on DHIs [26], which also emphasized the significance of personal agency and motivation to improve awareness and self-management of health when considering recruitment to DHIs. The current meta-synthesis enhanced these findings by highlighting how individual expectations and perceptions influenced participant engagement and approaches to DHIs. Participants with an active approach perceived DHIs to be more favorable compared with face-to-face therapy and were more likely to actively engage with DHIs by reflecting on and applying session content to their daily lives. Subsequently, these participants were more likely to complete treatment compared with participants who found DHIs isolating. DHIs attracted participation when they were perceived to be accessible and tailored to individual needs. Participants could access treatment from their own homes, which was more convenient and appropriate for those who would otherwise not access treatment. DHIs were also perceived to be more appealing because they reduced the stigma of accessing treatment. However, for some participants, the autonomous nature of DHIs made it easier to prioritize other tasks and disengage from treatment because the sense of commitment to treatment was reduced.

The meta-synthesis illustrates the different requirements of support that can potentially be provided in a number of ways. Thus, on the basis of participant preferences and needs, DHIs could be tailored to meet individual presentations. DHI functions and level of support likely to be required could be determined by initially assessing participant expectations and needs, as opposed to uniformly offering DHIs as an all-or-nothing option.

Strengths and Limitations

This is the first meta-synthesis to our knowledge that has looked at the acceptability and usability of DHIs across different types and formats and across a range of depression and anxiety disorders. This systematic review included all DHIs regardless of the delivery method (ie, text-based, automated, blended therapy), so it compared very different experiences. The themes emerging from the meta-synthesis highlight that personalized support was valued across studies irrespective of DHI type, format, or health condition. In particular, some form of human interaction was valued because it personalized therapy and increased motivation to complete therapy. This supports the generalizability of this finding, given that participants reported similar themes from a range of DHIs.

A limitation of including a range of different interventions is that some comparisons may have been incompatible or inappropriate across these rather different technologies. However, given the broad inclusion criteria, it is particularly important that personal support was still highlighted as a theme across studies and gives greater weight to its importance. We excluded peer-to-peer networks and DHIs of social support via the internet, use of social media, online support groups, or DHIs consisting of group therapy. Future research could explore whether participants' experiences of these delivery formats differ or are similar to our findings. It is worth highlighting that technological competence was only identified in 3 studies [24,41,51] as a potential barrier to engaging with a DHI. This could be because the views captured in the papers are mainly of participants who chose to engage in a DHI and do not capture the experiences of participants who decided not to participate in a DHI. The majority of the DHIs were CBT-based despite including a broad range of mental health conditions, treatment settings, and types of DHIs. Recommended therapies, such as interpersonal therapy for depression, were notably absent from studies included [58]. In terms of DHI variability, the use of videoconferencing to provide therapy was only included in 1 study [24] and could be investigated further in the future. Videoconferencing has the potential to provide real-time face-to-face therapy remotely, thus improving accessibility but also increasing the level of personalized support. A recent RCT investigating the clinical and cost-effectiveness of videoconferencing CBT delivered to repeat unscheduled care users with health anxiety found completion rates to be substantially higher than other DHI studies [7]. Recent years have also seen an increase in the use of smartphone apps for the treatment of depression and anxiety and the use of VR and interactive games for the treatment of anxiety and phobias. Therefore, it will be useful to update this systematic review to incorporate research regarding the acceptability and usability of smartphone apps to ascertain whether similar themes emerge.

This systematic review was not carried out using double screening, with only 30% of data being extracted by both reviewers. However, the high level of congruence found for the subset sample implied that the screening methods were rigorous. The systematic review only included papers published in English, which may reflect the fact that the majority of the studies were conducted in European and American countries. However, major sources of technology production are found in non–English speaking countries, for example, China, Japan, and India. The majority of studies were for the treatment of depression, and there were no studies on the use of DHIs for the treatment of somatoform disorders; therefore, the conclusions might not apply to somatoform disorders. We excluded participants with bipolar disorder despite depression being a characteristic feature of bipolar disorder; hence, we were unable to see whether the findings could translate to this condition.

Implications and Recommendations

On the basis of our findings from the meta-synthesis, we propose the following implications and recommendations to policy makers and commissioners:

1. Expectations and pre-existing beliefs about DHIs and their effectiveness can have an impact on patient experience and engagement. Therefore, addressing these expectations before beginning a DHI or building this into the initial stages of DHIs would help manage any misconceptions and address early barriers. This might also explain the high dropout rate from DHIs if people perceive them to be less effective than traditional approaches. An initial assessment should also include addressing patient preferences in terms of autonomy, level of support, and medium of

XSL•FO
RenderX

communication. This would help identify whether additional support is needed, and if so, the level required. Responsiveness to these kinds of unnecessary barriers could aid engagement and retention in DHIs.

2. Personalization of DHIs and personalized support was an overarching element that ran throughout all themes. Communication of some form was valued by the majority of the participants. It clarifies the intervention's purpose, personalizes therapy, and increases self-discipline and motivation to engage with DHIs. Thus, engagement with DHIs can be enhanced by including personal reminders for therapeutic activities and giving participants' individualized feedback on their progress with therapeutic tasks. Ensuring that the interface and content is succinct and easy to navigate around would also likely reduce dropout from DHIs. Thus, future DHIs need to consider how feedback and reminders can be incorporated and presented.
3. Both the previous recommendations and many other individual problems described by participants in this study would be addressed by making additional support available to those using DHIs. There is a clear and consistent theme of preferred qualities for this additional support, particularly where the person is passive or indifferent in their willingness to engage with a DHI. Additional support should be personalized, incorporate some form of human interaction, and be rapidly responsive. This kind of support was especially desirable when participants met significant barriers to DHIs. Therefore, treatment engagement, experience, and response should be monitored throughout DHIs, so those who do not respond or have a poor experience can receive an appropriate level of additional support quickly.

Conclusions

This review indicates that addressing patients' initial expectations of DHIs could help improve uptake, therapy completion, and effectiveness. Furthermore, the addition of rapid, responsive personal and human support, albeit offered remotely, could improve patient experience of DHIs, particularly when patients find DHIs challenging. The recommendations offered by this review have the potential to improve recruitment and retention rates to participation in DHIs and guide the design of DHIs so that they are personalized and improve overall patient experience.

Acknowledgments

The review was funded by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The funders had no role in the study design, data collection, analysis, interpretation of results, or write-up. This research was supported by the NIHR MindTech MedTech Co-operative and NIHR Nottingham Biomedical Research Centre. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care. RM is funded by NIHR Applied Research Collaboration East Midlands, NIHR MindTech Medical Technology, and Invitro Collaboration and Nottingham NIHR Biomedical Research Centre Mental Health and Technology theme. The authors would like to thank Dr David Smart, Dr Marcus Bicknell, Dr Karim Safiy, Dr Jennifer Martin, and Patient and Public Involvement or Engagement representatives, David Waldram and Fred Higton, for their valuable support and input in developing the systematic review protocol and analysis of findings. The authors would also like to thank Naomi Thorpe and Safiyah Osell for their support.

Authors' Contributions

SP, AA, and EY designed the search strings. EY conducted the search. AA and SP screened the papers for inclusion. AA and SP extracted data and appraised the quality of the papers. SP and AA coded the results section of each included paper and developed descriptive themes through inductive analysis. SP, AA, SM, SS, ER, NW, and RM reflected on and interpreted the descriptive themes to develop analytical themes. All authors contributed to the writing and refinement of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2009 checklist.
[DOCX File , 31 KB - jmir_v22i7e16228_app1.docx]

Multimedia Appendix 2
ENTREQ checklist.
[DOCX File , 32 KB - jmir_v22i7e16228_app2.docx]

Multimedia Appendix 3
Search strategy for MEDLINE.
[DOCX File , 16 KB - jmir_v22i7e16228_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
CASP Quality appraisal of included studies.
[DOCX File , 24 KB - jmir_v22i7e16228_app4.docx]

Multimedia Appendix 5
Summary of included studies.
[DOCX File , 25 KB - jmir_v22i7e16228_app5.docx]

Multimedia Appendix 6
Sample characteristics of included studies.
[DOCX File , 48 KB - jmir_v22i7e16228_app6.docx]

Multimedia Appendix 7
First, second, and third order constructs and sub-themes and contrasting positive and negative participant experiences.
[DOCX File , 32 KB - jmir_v22i7e16228_app7.docx]

References

1. Stansfeld S, Clark C, Bebbington P. Common mental disorders. In: McManus S, Bebbington P, Jenkins R, Brugha R, editors. Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey. Leeds, UK: NHS Digital; 2016:1-32.
2. Ritchie H, Roser M. Our World in Data. 2019. Mental Health URL: https://ourworldindata.org/mental-health [accessed 2019-08-15]
3. Edwards T, Stern A, Clarke D, Ivbijaro G, Kasney L. The treatment of patients with medically unexplained symptoms in primary care: a review of the literature. Ment Health Fam Med 2010 Dec;7(4):209-221 [FREE Full text] [Medline: 22477945]
4. OECD, European Union. Health at a Glance: Europe 2018 State of Health in the EU Cycle. Paris: OECD Publishing; 2018.
5. Kessler RC, Merikangas KR, Wang PS. Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. Annu Rev Clin Psychol 2007;3:137-158. [doi: 10.1146/annurev.clinpsy.3.022806.091444] [Medline: 17716051]
6. Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJ, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands study of depression and anxiety (NESDA). J Clin Psychiatry 2011 Mar;72(3):341-348. [doi: 10.4088/JCP.10m06176blu] [Medline: 21294994]
7. Morriss R, Patel S, Malins S, Guo B, Higton F, James M, et al. Clinical and economic outcomes of remotely delivered cognitive behaviour therapy versus treatment as usual for repeat unscheduled care users with severe health anxiety: a multicentre randomised controlled trial. BMC Med 2019 Jan 23;17(1):16 [FREE Full text] [doi: 10.1186/s12916-019-1253-5] [Medline: 30670044]
8. Prochaska JO, Norcross JC, Saul SF. Generating psychotherapy breakthroughs: transtheoretical strategies from population health psychology. Am Psychol 2019 Nov 25:- epub ahead of print. [doi: 10.1037/amp0000568] [Medline: 31763861]
9. McManus S, Bebbington P, Jenkins R, Brugha T. Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey. Leeds, UK: NHS Digital; 2014.
10. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
11. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems-a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
12. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: a systematic review protocol. JMIR Res Protoc 2015 Apr 28;4(2):e47 [FREE Full text] [doi: 10.2196/resprot.3990] [Medline: 25921274]
13. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
14. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How can ehealth technology address challenges related to multimorbidity? Perspectives from patients with multiple chronic conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
15. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry 2010 Mar;196(3):173-178. [doi: 10.1192/bjp.bp.109.066001] [Medline: 20194536]
16. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]

XSL•FO
RenderX

17. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
18. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450. [doi: 10.1111/bjc.12055] [Medline: 24831119]
19. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
20. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
21. Kelders SM. University of Twente Research Information. Understanding Adherence to Web-based Interventions URL: https://research.utwente.nl/en/publications/understanding-adherence-to-web-based-interventions [accessed 2020-05-18]
22. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008 Sep;193(3):181-184. [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]
23. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
24. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
25. Hollis C, Sampson S, Simons L, Davies E, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind alliance priority setting partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7]
26. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
27. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
29. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
30. Glanville J, Lefebvre C, Wright K. Google Sites. 2019. The InterTASC Information Specialists' Sub-Group Search Filter Resource URL: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home [accessed 2018-08-12]
31. Critical Appraisal Skills Programme: CASP. 2015. URL: http://www.casp-uk.net/ [accessed 2018-11-15]
32. Noblit G, Hare R. Meta-Ethnography: Synthesizing Qualitative Studies. Thousand Oaks, CA: Sage Publications; 1988.
33. Farzanfar R, Frishkopf S, Friedman R, Ludena K. Evaluating an automated mental health care system: making meaning of human–computer interaction. Comput Hum Behav 2007 May;23(3):1167-1182. [doi: 10.1016/j.chb.2004.11.015]
34. Purves DG, Dutton J. An exploration of the therapeutic process while using computerised cognitive behaviour therapy. Couns Psychother Res 2013 Dec;13(4):308-316. [doi: 10.1080/14733145.2012.761259]
35. Lovell K, Bower P, Gellatly J, Byford S, Bee P, McMillan D, et al. Clinical effectiveness, cost-effectiveness and acceptability of low-intensity interventions in the management of obsessive-compulsive disorder: the obsessive-compulsive treatment efficacy randomised controlled Trial (OCTET). Health Technol Assess 2017 Jun;21(37):1-132 [FREE Full text] [doi: 10.3310/hta21370] [Medline: 28681717]
36. Knopp-Hoffer J, Knowles S, Bower P, Lovell K, Bee PE. 'One man's medicine is another man's poison': a qualitative study of user perspectives on low intensity interventions for obsessive-compulsive disorder (OCD). BMC Health Serv Res 2016 May 18;16:188 [FREE Full text] [doi: 10.1186/s12913-016-1433-3] [Medline: 27194033]
37. Lillevoll KR, Wilhelmsen M, Kolstrup N, Høifødt RS, Waterloo K, Eisemann M, et al. Patients' experiences of helpfulness in guided internet-based treatment for depression: qualitative study of integrated therapeutic dimensions. J Med Internet Res 2013 Jun 20;15(6):e126 [FREE Full text] [doi: 10.2196/jmir.2531] [Medline: 23786763]
38. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 7;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
39. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]

XSL•FO
RenderX

40. Advocat J, Lindsay J. Internet-based trials and the creation of health consumers. Soc Sci Med 2010 Feb;70(3):485-492. [doi: 10.1016/j.socscimed.2009.10.051] [Medline: 19926185]
41. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
42. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
43. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. Internet Interv 2017 Dec;10:29-38 [FREE Full text] [doi: 10.1016/j.invent.2017.09.001] [Medline: 30135750]
44. Gega L, Smith J, Reynolds S. Cognitive behaviour therapy (CBT) for depression by computer vs therapist: patient experiences and therapeutic processes. Psychother Res 2013;23(2):218-231. [doi: 10.1080/10503307.2013.766941] [Medline: 23390994]
45. Johansson O, Michel T, Andersson G, Paxling B. Experiences of non-adherence to internet-delivered cognitive behavior therapy: a qualitative study. Internet Interv 2015 May;2(2):137-142. [doi: 10.1016/j.invent.2015.02.006]
46. Ly KH, Janni E, Wrede R, Sedem M, Donker T, Carlbring P, et al. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. Internet Interv 2015 Mar;2(1):60-68. [doi: 10.1016/j.invent.2014.12.002]
47. Richards D, Murphy T, Viganó N, Timulak L, Doherty G, Sharry J, et al. Acceptability, satisfaction and perceived efficacy of an internet-delivered treatment for depression. Internet Interv 2016 Sep;5:12-22 [FREE Full text] [doi: 10.1016/j.invent.2016.06.007] [Medline: 30135802]
48. Walsh A, Richards D. Experiences and engagement with the design features and strategies of an internet-delivered treatment programme for generalised anxiety disorder: a service-based evaluation. Br J Guid Counc 2016 Feb 28;45(1):16-31. [doi: 10.1080/03069885.2016.1153039]
49. Bendelin N, Hesser H, Dahl J, Carlbring P, Nelson KZ, Andersson G. Experiences of guided Internet-based cognitive-behavioural treatment for depression: a qualitative study. BMC Psychiatry 2011 Jun 30;11:107 [FREE Full text] [doi: 10.1186/1471-244X-11-107] [Medline: 21718523]
50. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
51. Gerhards SA, Abma TA, Arntz A, de Graaf LE, Evers SM, Huibers MJ, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. J Affect Disord 2011 Mar;129(1-3):117-125. [doi: 10.1016/j.jad.2010.09.012] [Medline: 20889214]
52. Hind D, O'Cathain A, Cooper CL, Parry GD, Isaac CL, Rose A, et al. The acceptability of computerised cognitive behavioural therapy for the treatment of depression in people with chronic physical disease: a qualitative study of people with multiple sclerosis. Psychol Health 2010 Jul;25(6):699-712. [doi: 10.1080/08870440902842739] [Medline: 20204955]
53. Holst A, Nejati S, Björkelund C, Eriksson MC, Hange D, Kivi M, et al. Patients' experiences of a computerised self-help program for treating depression - a qualitative study of internet mediated cognitive behavioural therapy in primary care. Scand J Prim Health Care 2017 Mar;35(1):46-53 [FREE Full text] [doi: 10.1080/02813432.2017.1288813] [Medline: 28277055]
54. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open 2015 Nov 30;5(11):e008581 [FREE Full text] [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]
55. Pugh N, Hadjistavropoulos H, Hampton A, Bowen A, Williams J. Client experiences of guided internet cognitive behavior therapy for postpartum depression: a qualitative study. Arch Womens Ment Health 2015 Apr;18(2):209-219. [doi: 10.1007/s00737-014-0449-0] [Medline: 25109484]
56. Ashford MT, Olander EK, Rowe H, Fisher JR, Ayers S. Feasibility and acceptability of a web-based treatment with telephone support for postpartum women with anxiety: randomized controlled trial. JMIR Ment Health 2018 Apr 20;5(2):e19 [FREE Full text] [doi: 10.2196/mental.9106] [Medline: 29678804]
57. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mhealth augmentation of collaborative care: a mixed methods pilot study. Gen Hosp Psychiatry 2018;51:22-29 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]
58. NICE: The National Institute for Health and Care Excellence. 2019. Depression in Adults: Recognition and Management URL: https://www.nice.org.uk/guidance/cg90 [accessed 2019-09-10]

Abbreviations

CASP: Critical Appraisal Skills Program
CBT: cognitive behavioral therapy
cCBT: computerized cognitive behavioral therapy

XSL•FO
RenderX

DHI: digital health intervention
GAD: generalized anxiety disorder
NIHR: National Institute for Health Research
OCD: obsessive-compulsive disorder
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
VR: virtual reality

Edited by G Eysenbach; submitted 12.09.19; peer-reviewed by A Parks, A Gulliver; comments to author 07.11.19; revised version received 20.02.20; accepted 23.03.20; published 06.07.20.

Please cite as:
Patel S, Akhtar A, Malins S, Wright N, Rowley E, Young E, Sampson S, Morriss R
The Acceptability and Usability of Digital Health Interventions for Adults With Depression, Anxiety, and Somatoform Disorders: Qualitative Systematic Review and Meta-Synthesis
J Med Internet Res 2020;22(7):e16228
URL: https://www.jmir.org/2020/7/e16228
doi:10.2196/16228
PMID:32628116

©Shireen Patel, Athfah Akhtar, Sam Malins, Nicola Wright, Emma Rowley, Emma Young, Stephanie Sampson, Richard Morriss. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review

Dillys Larbi[1], MPH, MSc; Pietro Randine[1,2], MCS; Eirik Årsand[1,2], PhD; Konstantinos Antypas[1,3], PhD; Meghan Bradway[1,4], MBA; Elia Gabarron[1], PhD

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

[2]Department of Computer Science, Faculty of Science and Technology, UiT The Arctic University of Norway, Tromsø, Norway

[3]SINTEF Digital, Oslo, Norway

[4]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Meghan Bradway, MBA
Norwegian Centre for E-health Research
University Hospital of North Norway
Forskningsparken i Breivika 3rd Fl
Sykehusveien 23
Tromsø, 9019
Norway
Phone: 47 91193393
Email: meghan.bradway@ehealthresearch.no

Abstract

Background: There is growing evidence that apps and digital interventions have a positive impact on diabetes self-management. Standard self-management for patients with diabetes could therefore be supplemented by apps and digital interventions to increase patients' skills. Several initiatives, models, and frameworks suggest how health apps and digital interventions could be evaluated, but there are few standards for this. And although there are many methods for evaluating apps and digital interventions, a more specific approach might be needed for assessing digital diabetes self-management interventions.

Objective: This review aims to identify which methods and criteria are used to evaluate apps and digital interventions for diabetes self-management, and to describe how patients were involved in these evaluations.

Methods: We searched CINAHL, EMBASE, MEDLINE, and Web of Science for articles published from 2015 that referred to the evaluation of apps and digital interventions for diabetes self-management and involved patients in the evaluation. We then conducted a narrative qualitative synthesis of the findings, structured around the included studies' quality, methods of evaluation, and evaluation criteria.

Results: Of 1681 articles identified, 31 fulfilled the inclusion criteria. A total of 7 articles were considered of high confidence in the evidence. Apps were the most commonly used platform for diabetes self-management (18/31, 58%), and type 2 diabetes (T2D) was the targeted health condition most studies focused on (12/31, 38%). Questionnaires, interviews, and user-group meetings were the most common methods of evaluation. Furthermore, the most evaluated criteria for apps and digital diabetes self-management interventions were cognitive impact, clinical impact, and usability. Feasibility and security and privacy were not evaluated by studies considered of high confidence in the evidence.

Conclusions: There were few studies with high confidence in the evidence that involved patients in the evaluation of apps and digital interventions for diabetes self-management. Additional evaluation criteria, such as sustainability and interoperability, should be focused on more in future studies to provide a better understanding of the effects and potential of apps and digital interventions for diabetes self-management.

(J Med Internet Res 2020;22(7):e18480) doi:10.2196/18480

KEYWORDS

self-management; diabetes mellitus; mobile applications; computer communication networks; mHealth; eHealth; health care evaluation mechanisms

XSL•FO
RenderX

Introduction

As the number of people with diabetes continues to rise worldwide [1], the need to increase patients' self-management skills is crucial to improve clinical outcomes and reduce health-related costs [2,3]. There is growing evidence that apps and digital interventions such as websites (web), social media, and other online services have a positive impact on diabetes self-management [4-12], suggesting that standard self-management could be supplemented by digital interventions to aid and improve patients' skills [4-12]. While some apps and digital interventions have benefited patients, not all of them seem to be based on research, and some of these digital interventions could even compromise the safety of patients with diabetes [13].

To improve diabetes self-management with apps and digital interventions, the World Health Organization and the European Commission [14,15] deem it necessary that the available apps and digital interventions are accurate and reliable. Several initiatives, models, and frameworks suggest how some of these apps and digital interventions could be evaluated [16-19]. These approaches commonly name background information, privacy and security, evidence on the provided information, ease of use, or interoperability as issues that need to be addressed [16-18]. Regarding how to evaluate these criteria, several methods of different complexity have been proposed. These include simple questions to be answered by health care professionals (HCPs) and patients, whereas more complex methodology approaches, such as laboratory-based testing, field testing, and N-of-1 design, are used by researchers [18,20]. Although the aforementioned issues are relevant for diabetes self-management apps and digital interventions, a more specific approach is needed for assessing the growing number and rapidly changing functionalities of these digital diabetes self-management interventions.

Another relevant issue is who should be involved in these evaluations. As patients are often required to make critical decisions based on their own generated health information [21], people with diabetes should be involved in these evaluations. However, a previous assessment of digital health interventions demonstrated limited consideration of user perceptions, and also that of health care personnel [22].

In this systematic review, we identify the specific methods and evaluation criteria that were used to assess apps and digital interventions for diabetes self-management. We also report how patients were involved in these assessments.

Methods

This review followed the PRISMA approach [23], and its systematic review protocol is registered in PROSPERO (Registration number: CRD42018115246).

Data Sources and Search Strategy

We performed a single data search in June 2018. The search strategy covered all studies that assessed diabetes self-management apps and digital interventions, involved patients, and were published in English after 2015. We chose a short search period to get a rapid overview of the most recent methods and evaluation criteria. The search strategy covered the following databases: CINAHL, EMBASE, MEDLINE, and Web of Science. The full search strategy is available in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We included articles for review if they were (1) primary studies referring to the evaluation of apps or digital interventions for diabetes self-management; and (2) involved patients in the evaluation.

Article were excluded if (1) the evaluation only measured medical values (ie, weight, glycated hemoglobin [HbA_{1c}], blood glucose); (2) it was not a primary study; (3) it did not focus on apps or digital interventions for diabetes self-management; (4) the full-text was not available; (5) it was not a peer-reviewed publication; (6) it was not in English; or (7) it was published before 2015.

Eligibility and Data Collection Procedure

We uploaded all references captured by the search strategy to Rayyan and EndNote and removed duplicates. The eligibility of the articles was assessed in two stages. In the first stage, 2 independent reviewers (PR and EG) examined all titles and abstracts. Eligibility doubts were discussed and agreed with a third and fourth reviewer (KA and EÅ). In the second stage, the full texts of the selected articles were carefully examined by 2 independent reviewers (PR and EG) to confirm their eligibility.

Two reviewers (PR and MB) independently extracted and recorded the data from these articles on an Excel spreadsheet (Microsoft). We extracted the following information from each article: type of platform, targeted health condition, study population, methods of evaluation, and evaluation criteria. Incongruences with the data extraction were discussed among the research group.

Confidence in the Evidence and Risk of Bias Assessment

Two reviewers (EG and KA) assessed the confidence in the evidence and risk of bias of the articles. We used an approach based on the CERQual guidelines [24] to assess the confidence in the evidence of the qualitative primary studies, by evaluating their methodological limitations, relevance, and adequacy. We followed the GRADE guidelines [25] to assess mixed-methods studies, quantitative studies, and randomized trials.

Strategy for Data Synthesis

We provide a narrative qualitative synthesis of the findings from the included articles, structured around confidence in the evidence and risk of bias; type of platform (apps, web, or multiplatform [ie, ≥2 types of platform delivering the same intervention in a study]); targeted health condition (type 1 diabetes [T1D], T2D, gestational diabetes mellitus, both T1D and T2D, and unspecified diabetes type); methods of evaluation (questionnaires, interviews, user-group meetings, health measures, system usage analysis, or other); and evaluation criteria (usability, clinical impact, cognitive impact, behavioral impact, feasibility, engagement, acceptability and acceptance, or security and privacy).

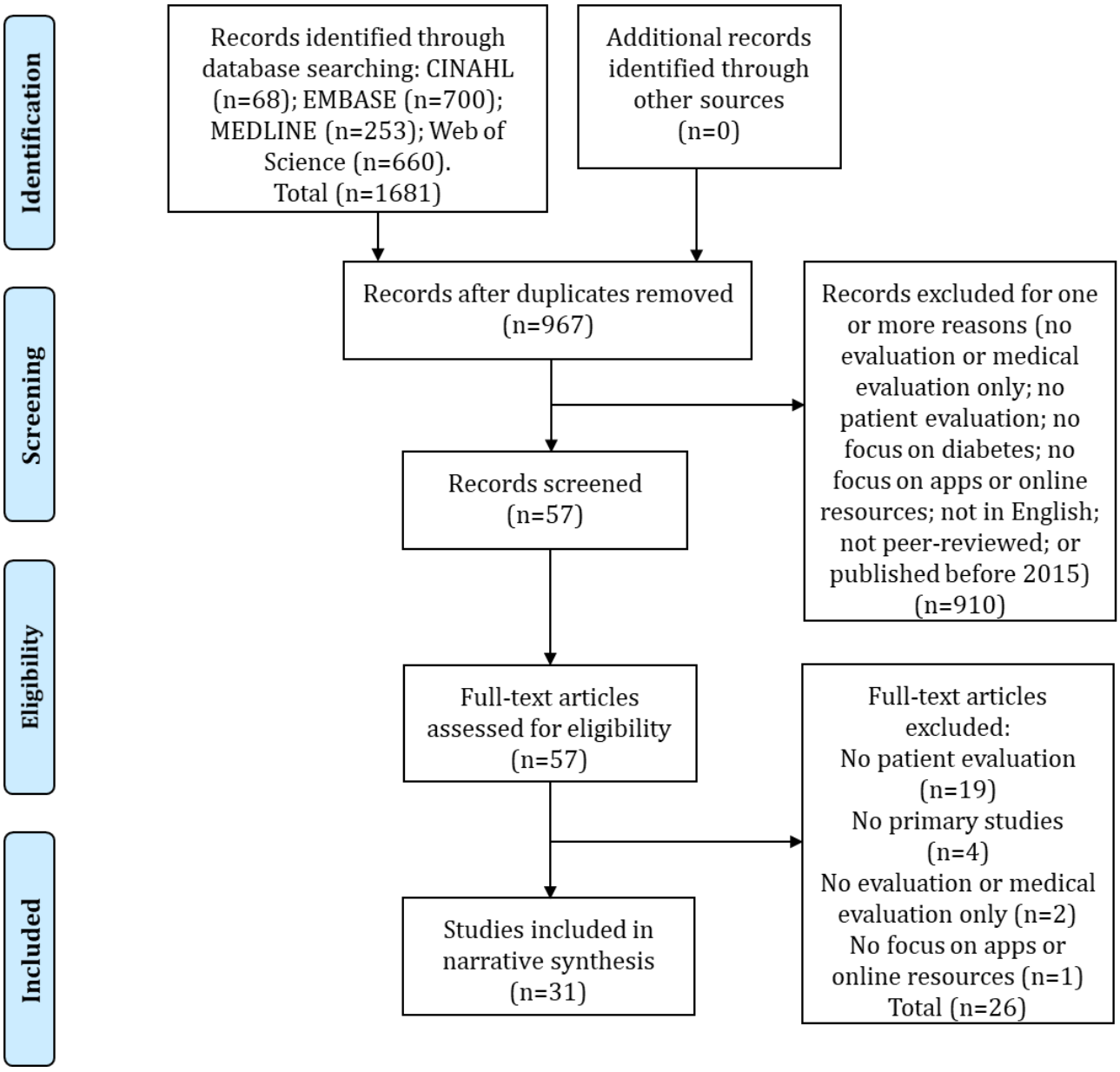
XSL•FO
RenderX

Results

Identified and Included Studies

The search strategy resulted in 1681 articles. After removing duplicates, 967 articles remained. In the abstract screening, we excluded 910 articles in accordance with one or more of the exclusion criteria. A total of 57 articles were eligible for full-text screening, 26 of which were excluded (see Multimedia Appendix 2). A total of 31 articles were eventually included in the review [26-56] (see Multimedia Appendix 3). The PRISMA diagram in Figure 1 summarizes the selection process. The confidence in the evidence was considered high in 7 articles [27,33,36,43,51,52,54]; moderate to high in 1 [56]; moderate in 17 [26,28-32,35,37,39,41,42,45,46,48,49,53,55], and low in 6 [34,38,40,44,47,50].

Figure 1. PRISMA flowchart of the selection procedure.



Study Population

The 31 articles in this review included evaluations from 3689 participants. The number of participants in each study ranged from 7 [41,50] to 1041 [43]. In addition to including patients with diabetes in their evaluations, some of the studies expanded the participant group to include HCPs (8/31, 26%) [26,27,29,37,42,48,53,56], developers (4/31, 13%) [26,44,51,56], researchers (3/31, 10%) [29,38,44], informal caregivers (eg, parents, family members) (4/31, 13%) [29,44,53,56], and others (including experts and other unspecified individuals) (8/31, 26%) [26,29,32,37,38,48,55,56].

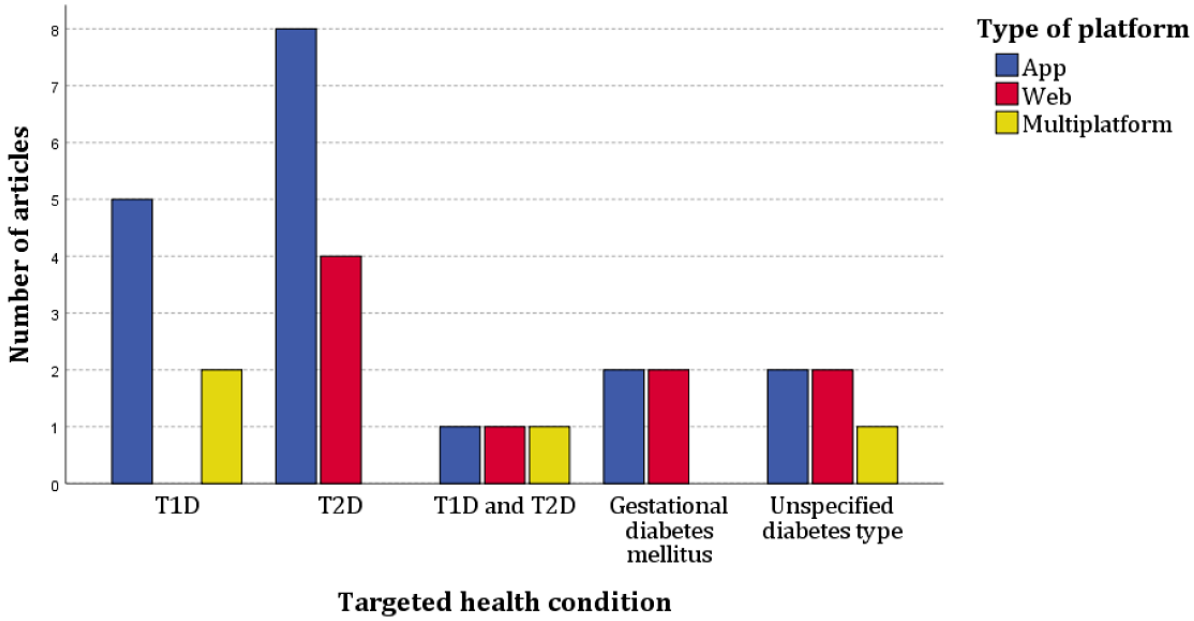
Type of Platform and Targeted Health Condition

Most of the 31 included studies evaluated interventions delivered via apps (18/31, 58%) [29-34,36-39,41,46-49,52,54,56], followed by web (9/31, 29%) [27,28,35,43,45,50,51,53,55] and multiplatform (4/31, 13%) [26,40,42,44]. In the studies that conducted a randomized controlled trial, the self-management platform was the main mode of intervention compared with a standard paper diary [33], the intervention plus counseling via

XSL•FO
RenderX

telephone call [54], and a plain text version of the web intervention [43] as opposed to an interactive version. In addition, the same intervention was referred to by some of the studies: Young with Diabetes app [29,36], My Diabetes My Way [30,55], and WellDoc [31,50]. The evaluated digital self-management interventions targeted mostly T2D (12/31, 38%) [27,31,32,34,43,46-51,54], followed by T1D (7/31, 23%) [29,33,36,40,41,44,56], unspecified diabetes type (5/31, 16%) [26,37,39,53,55], gestational diabetes mellitus (4/31, 13%) [35,38,45,52], and T1D and T2D (3/31, 10%) [28,30,42] (Figure 2).

Figure 2. Distribution of types of platform and targeted health conditions among included articles (n=31).



Identified Evaluation Methods

The methods of evaluation were grouped into 6 categories: questionnaires, interviews, user-group meetings, health-related measures, system usage analysis, and other measurements. We also identified 20 specific methods that were either used once or multiple times by the studies during the evaluation process.

The interrater agreement for the methods of evaluation was found to be κ=0.550, which represents a moderate agreement [57]. A summary of the specific methods of evaluation and studies that used them is presented in Table 1.

Questionnaires were the most common method used to evaluate diabetes self-management apps and digital interventions (21/31 studies, 68%) [29-40,42,43,45,46,50,51,54-56]. Standardized questionnaires were the most frequently used: 16 in total, each used one or multiple times among 13 studies [31,33-36,38,40,42,43,45,50,51,54]. The second most common method of evaluation was interviews (13/31 studies, 42%) [28,29,31,32,35,36,38,45,48,49,52,53,56], mainly semistructured interviews, which were used 14 times in 11 studies [28,31,32,35,36,45,48,49,52,53,56]. Other methods of evaluation that were identified in the included studies were user-group meetings (11/31, 35%) [26,27,29,37,40-42,45,47,48,56], health-related measures (9/31, 29%) [31,33,36,42,45,49,51,54,55], system usage analysis (8/31, 26%) [29,34,35,42,43,45,51,54], and other measurements (7/31, 23%) [26,37,38,44,48,51,55]. Table 1 summarizes the specific methods of evaluation, the number of times these methods were used, and the number of studies that employed these methods.

Among the 7 studies considered of high confidence in the evidence, the evaluations of the apps and digital diabetes self-management interventions were based mostly on standardized questionnaires [33,36,43,51,54], medical tests [33,36,51,54], and usage log analysis [43,51,54], followed by author-created questionnaires [43,51], semistructured interviews [36,52], focus groups [27], self-reported health measures [33], self-reported usage [43], alpha testing [51], and other oral and written feedback [51].

XSL•FO
RenderX

Table 1. Specific methods of evaluation and studies that used them.

Method of evaluation, specific type (n=times used), and details	Reference(s)
Questionnaires	
Standardized questionnaires (n=26)	
Block Food Frequency Assessment	[45]
Dietary Knowledge, Attitude, and Behavior Questionnaire	[51]
Health Care Climate Questionnaire	[36]
Paffenbarger Questionnaire	[45]
Patient Enablement Instrument	[43]
Patient Health Questionnaire-9	[50]
Patient Reported Diabetes Symptoms Scale	[50]
Perceived Competence in Diabetes	[36]
Problem Areas in Diabetes-5	[31]
Problem Areas in Diabetes	[33,36,40,42]
RAND 36-Item Health Survey 1.0	[33]
Self-Efficacy for Diabetes Scale	[50]
System Usability Scale	[33-35,38,40,42]
The Health Education Impact Questionnaire	[54]
The Service User Technology Acceptability Questionnaire	[54]
36-Item Short Form Survey	[50]
Author-created questionnaires (n=20)	
N/A[a]	[29,30,32,37-39,42,43,45,46,51,55,56]
User-group meetings	
Focus groups (n=9)	
N/A	[27,37,40-42,45,47]
Workshops (n=7)	
N/A	[26,29,48,56]
Interviews	
Semistructured interviews (n=14)	
N/A	[28,31,32,35,36,45,48,49,52,53,56]
Unspecified interview format (n=3)	
N/A	[29,38]
System usage analysis	
Usage log analysis (n=8)	
N/A	[34,35,43,45,51,54]
Self-reported usage (n=2)	
N/A	[42,43]
Think-aloud protocol (n=1)	
N/A	[29]
Health-related measures	
Medical tests (n=8)	

XSL•FO
RenderX

Method of evaluation, specific type (n=times used), and details	Reference(s)
HbA_{1c}	[31,33,36,45,49,51,54,55]
Fasting blood glucose	[51]
Blood pressure and cholesterol	[55]
Gestational weight gain	[45]
Self-reported health measures (n=5)	
Self-reported blood glucose	[33,42]
Self-reported physical activity and nutritional habits	[42]
Other measurements	
Security assessment (n=1)	
N/A	[44]
Scenarios (n=2)	
N/A	[37,38]
Cost-effectiveness (n=1)	
N/A	[55]
Alpha testing (n=1)	
N/A	[51]
Observation (n=2)	
N/A	[48]
Rating system (n=1)	
Star rating	[26]
Heuristics method (n=1)	
Bertini's mobile tool	[37]
Anecdotal feedback (n=1)	
Open text review	[26]
Other oral and written feedback (n=2)	
N/A	[44,51]

[a]N/A: not applicable.

Identified Evaluation Criteria

The evaluated criteria were grouped into 8 categories: usability, clinical impact, cognitive impact, behavioral impact, feasibility, engagement, acceptability and acceptance, and security and privacy. The included studies evaluated one or several of these identified criteria. The interrater agreement (κ) for the evaluation criteria was found to be 0.563, which represents a moderate agreement [57].

Among the 7 studies considered of high confidence in the evidence, the most commonly evaluated criteria were clinical impact [33,36,51,54], cognitive impact [33,36,43,54], and engagement [43,51,54], followed by usability [33,51], behavioral impact [33,51], and acceptability and acceptance [51,54]. None of these studies considered of high confidence evaluated feasibility or security and privacy.

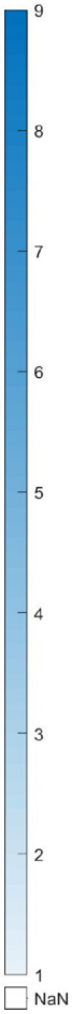
Qualitative and mixed-method studies that used thematic analysis in their evaluation focused mostly on usability as an evaluation criterion. Three of the studies considered of high confidence in evidence were qualitative and mixed-method studies. Of these, 2 evaluated cognitive impact [36,52] and usability [27,52], and 1 evaluated engagement [36].

Figure 3 shows the number of studies that used each of the specific methods to evaluate the identified criteria. It illustrates that several methods were used to evaluate one criterion in a single study. Likewise, some studies evaluated several criteria using one or more of the identified methods of evaluation. For example, of the 31 included studies, 9 [31,33,36,40,42,43,45,50,54] evaluated cognitive impact using standardized questionnaires.

XSL•FO
RenderX

Figure 3. Number of studies using the various methods of evaluation and evaluation criteria. Blank boxes (NaN): No studies within this category.

Method of Evaluation	Specific evaluation method \ Evaluation criteria	Usability	Feasibility	Acceptability and acceptance	Behavioral impact	Cognitive impact	Engagement	Clinical impact	Security and privacy
Questionnaires	Standardized	6		1	2	9		2	
	Author-created	8	5	5	2	2			
User-group meetings	Focus groups	5	2	2	2	1	1		
	Workshops	4	2						1
Interviews	Semistructured	7	3	1	5	5	3	1	
	Unspecified format	2				1			
System usage analysis	Usage log analysis	1	2				3		
	Patient-reported usage						2		
	Think-aloud test	1							
Health-related measures	Medical tests							8	
	Self-reported health measures				2			1	
Other measurements	Security assessment								1
	Scenarios		2						
	Cost-effectiveness		1						
	Alpha testing	1							
	Observation	1							
	Rating system	1							
	Heuristics method	1							
	Anecdotal feedback	1		1					
	Other oral and written feedback	2							



Discussion

Summary of the Findings

This review aimed to identify the existing methods and criteria used to assess apps and digital diabetes self-management interventions that involved patients in their evaluations. A total of 31 articles were included in the review, 7 of which were considered of high confidence in the evidence [27,33,36,43,51,52,54]. More than half of the studies (18/31, 58%) focused on the evaluation of apps for diabetes self-management, and 12 of the 31 studies addressed T2D. The most commonly used methods of evaluation were questionnaires, interviews, and user-group meetings. The most used evaluation criteria to assess apps and digital interventions for diabetes self-management were cognitive impact, clinical impact, and usability.

Specific Evaluation Criteria and Diabetes Patients' Assessment

In our review, we have found that studies dealing specifically with apps and digital interventions for diabetes self-management focus on the evaluation of more technology-related and users' interaction aspects (ie, acceptability and acceptance, and engagement). In addition, these studies focus on the impact that these digital self-management interventions have on the individual. Behavioral impact, cognitive impact, and clinical impact were used as relevant criteria for assessing all types of digital interventions for diabetes self-management. It is vital to measure the interventions' impact on their users because those that have shown benefits related to behavioral, cognitive, and clinical impact could reduce health-related costs [2,3].

Evidence shows that involving individuals in the assessment of different health interventions has a positive impact on health

XSL•FO
RenderX

[58]. We found few articles (n=31) in this review that involved patients in the evaluation of apps and digital interventions for diabetes self-management. The evaluations in which patients were involved in mostly focused on usability and cognitive impact. Evaluation criteria that could measure patients' continuous use of these apps and digital interventions for self-management could supplement both their qualitative responses and the more static traditional and clinical criteria. This is an opportunity for improvement, as none of the studies in this review evaluated the same criterion using both qualitative results from patients and quantitative measures.

Involving patients with diabetes in assessing apps and digital self-management interventions, and obtaining their feedback regarding additional evaluation criteria could also increase our knowledge about the features that support engagement with these technologies. This could also help create better digital health interventions that encourage more continuous and effective use [59]. The most common methods of evaluation with the patients were questionnaires, interviews, and user-group meetings. Simple methods such as these elicit the opinion and perceptions of users, as well as encourage them to critically analyze self-management apps and digital interventions. Therefore, such methods should be used in conjunction with complex methods used by researchers and developers [18,20], especially to measure the same criterion.

Improving Reported Evaluations of Digital Interventions for Diabetes Self-Management

Apps and digital health interventions have evolved quickly. Yet, compared with other sectors, the health industry seems to be behind with regard to digitalization [60]. Currently, most apps and digital interventions for self-management are not recommended as part of the treatment plan, maybe because their design and development do not take into consideration sustainability [61]. In fact, digital health interventions rarely advance beyond a pilot phase [62,63], or the duration of an intervention study.

In 2016, the mobile health (mHealth) evidence reporting and assessment checklist was developed by the World Health Organization to help with reporting evidence of the effectiveness of mHealth interventions [64]. The checklist recommended reporting on items that touch on sustainability, scalability, and transparency, such as infrastructure, interoperability, contextual adaptability, and replicability, which we still see are not much focused on in today's studies. Future studies should also consider these evaluation criteria in addition to gender and equity issues associated with the use of apps and digital interventions for diabetes self-management.

Evaluation reports for apps and digital interventions for diabetes self-management must be standardized, as recommended by the CONSORT-EHEALTH guidelines for reporting digital health interventions [65]. The lack of standardization made it challenging to compare studies as different authors used different terminologies to describe the same evaluation criterion. For example, one study [37] used the term *heuristics evaluation*, which was grouped under usability because it evaluated measures such as the visibility of app status, ease of input, and readability. Likewise, another study [32] evaluated *satisfaction*, which falls under usability because it evaluated among others, visual attractiveness and ease of use.

As electronic health (eHealth) research is a multidisciplinary field, we assume that the authors chose these terms based on the various educational or professional backgrounds and the corresponding target audiences. By following the World Health Organization classification of digital health interventions [66], terminologies related to the evaluation of apps and digital interventions for diabetes self-management could be standardized to facilitate straightforward interpretation and aggregation of research evidence.

Association Between Methods Used and Criteria Evaluated

In our review we have found that there was an almost even split of studies that used standardized questionnaires, author-created questionnaires, and semistructured interviews to evaluate usability. Our results are to some extent in line with the findings of a previous review that found that usability was mainly assessed though polls and questionnaires [67]. The usability of a digital self-management intervention is crucial to its successful adoption, its acceptance, and the individual's engagement with it. In addition, we found that cognitive impact was often assessed not only through standardized questionnaires, but also through semistructured interviews.

Comparing the methods for the evaluation of usability with those for the evaluation of cognitive impact, we identified that it was more common to use author-created questionnaire for usability. A possible explanation might be the wide variety of intervention delivery platforms (eg, different types of apps and online resources) that might create different evaluation needs not captured in existing standardized usability questionnaires. Another explanation might be the different research traditions in different disciplines. Usability might be more often a concern of computer science researchers, whereas cognitive impact a concern of health researchers and professionals.

Finally, health outcomes were almost exclusively evaluated by medical tests, showing the preference of health researchers and professionals in using standardized tests to determine the impact of digital interventions. Several other methods can be used to evaluate multiple criteria; however, depending on the aim and the type of study, researchers must endeavor to exhaust all available methods to ensure consistency of results.

Feasibility of Using Digital Self-Management Interventions in Clinical Workflow

Although most apps and digital health interventions are intended for self-management, some of them also provide access to the health care system, such as communication with HCPs and electronic health journals. The reviewed studies consistently reported that this is in response to patients' interest in being able to contact their HCPs or share results (eg, their blood glucose results with their health care team). This was the case not only within our review [35,36,42,44,47,48,56] but also by industry research groups [68,69]. This implies the potential and expectation for further involvement of HCPs in patients' use of apps and digital interventions for diabetes self-management.

XSL•FO
RenderX

Several studies, including many in this review, have shown that involving HCPs in digital interventions is associated with improved self-management of diabetes and the success of these interventions [31,48,49,52,70-72]. Therefore, studies focusing on apps and digital interventions for diabetes self-management should evaluate the possibilities of effortlessly integrating these interventions in the workflow of HCPs—the connection and interaction with electronic health journals and other existing health information systems. Such an integration can be achieved by evaluating the infrastructure needed for digital self-management interventions [64].

Limitations and Strengths

The search for articles covered a short period (2015-2018) and focused on articles published in the English language. Therefore, we may have missed relevant studies that reported additional evaluation methods or evaluation criteria. Our interrater agreement of the data extraction was only moderate; however, all incongruences were discussed among the research group. Our findings have provided a useful overview of the recent evaluation methods and criteria that researchers are using to assess current apps and digital interventions for diabetes self-management. Furthermore, our review included both quantitative and qualitative studies which provided a better characterization of different evaluation methods and criteria that are being used to assess digital diabetes self-management interventions.

Conclusions

There are only few studies that involved patients in the evaluation of apps and digital interventions for diabetes self-management, and even fewer still considered of high confidence in the evidence. The most common evaluation methods were questionnaires, interviews, and user-group meetings, whereas evaluation criteria were cognitive impact, clinical impact, and usability. Studies with high confidence in the evidence did not evaluate feasibility or security and privacy, neither were patients involved in evaluating the latter criterion which was evaluated in only 2 [29,44] of the included studies.

It is important to the successful implementation and continuous use of apps and digital interventions for diabetes self-management that patients are involved in evaluating every criteria. In that way, they can contribute to the development and modification of these digital interventions to better meet their specific self-management needs. Furthermore, the methods and criteria evaluated in digital diabetes self-management interventions should be expanded to assess and ensure sustainability and interoperability. In addition, studies should evaluate the association between cognitive, clinical, and behavioral impact of these apps and digital interventions, and health-related costs for individuals with diabetes. This could help improve health care associated with the management of diabetes and promote the incorporation of apps and digital interventions for self-management in the services provided at health care facilities.

Acknowledgments

This project is funded by Helse Nord (HNF1425-18). The coauthors acknowledge the advice of the project's Advisory board: Professor Gunnar Hartvigsen, Anne Grethe Olsen MD, and Dr. Med. Anne Helen Hansen, and also the support and contributions of Per Erlend Hasvold in his role as an "internal reviewer." Furthermore, we thank Dr. Steven Bradway for his assistance with the coarse data extraction and organization at the start of this review. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Authors' Contributions

KA was responsible for database searching; EG, EÅ, KA, and PR were responsible for title, abstract, and full-text screening; MB and PR performed independent data extraction; EG and KA evaluated risk of bias; and DL, EG, EÅ, MB, KA, and PR performed data analysis and interpretation. All the coauthors contributed to drafting and revising the review. All coauthors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy (search date: June 21, 2018).
[DOC File , 28 KB - jmir_v22i7e18480_app1.doc]

Multimedia Appendix 2
List of rejected articles after full-text review (n=26).
[DOC File , 51 KB - jmir_v22i7e18480_app2.doc]

Multimedia Appendix 3
Articles included in qualitative synthesis (n=31).
[DOC File , 267 KB - jmir_v22i7e18480_app3.doc]

XSL•FO
RenderX

Multimedia Appendix 4
PRISMA checklist.
[DOC File , 58 KB - jmir_v22i7e18480_app4.doc]

References

1. World Health Organization. Diabetes. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2019-12-01]
2. Barker I, Steventon A, Williamson R, Deeny SR. Self-management capability in patients with long-term conditions is associated with reduced healthcare utilisation across a whole health economy: cross-sectional analysis of electronic health records. BMJ Qual Saf 2018 Dec;27(12):989-999 [FREE Full text] [doi: 10.1136/bmjqs-2017-007635] [Medline: 30139822]
3. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
4. Wilhide ICC, Peeples MM, Kouyaté RCA. Evidence-Based mHealth Chronic Disease Mobile App Intervention Design: Development of a Framework. JMIR Res Protoc 2016;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
5. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
7. Gabarron E, Årsand E, Wynn R. Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
8. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
9. Ramadas A, Quek KF, Chan CKY, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
10. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
11. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
12. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: A global overview. Diabetes Metab Res Rev 2018 May;34(4):e2990. [doi: 10.1002/dmrr.2990] [Medline: 29431916]
13. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
14. European Commission. Report of the Working Group on mHealth Assessment Guidelines. 2017 Jun. URL: https://ec.europa.eu/digital-single-market/en/news/report-working-group-mhealth-assessment-guidelines [accessed 2019-12-01]
15. World Health Organization. mHealth. New Horizons for Health Through Mobile Technologies. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2019-12-01]
16. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digital Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
17. Health on the net. HON Code. 2019. URL: https://www.hon.ch/en/ [accessed 2020-01-01]
18. American Psychiatric Association. App Evaluation Model. 2020. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2020-01-01]
19. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies. London, UK: National Institute for Health and Care Excellence; 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2020-02-26]
20. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
21. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
22. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth Assessment: Conceptualization of a Global Framework. JMIR Mhealth Uhealth 2017 May 02;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]

XSL•FO
RenderX

23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
24. Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 2: how to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]
25. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
26. Ashurst EJ, Jones RB. Is the Health App Challenge approach of patient-led application conception, development, and review worthwhile? Health Policy Technol 2017 Mar;6(1):83-92. [doi: 10.1016/j.hlpt.2016.12.001]
27. Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, et al. Developing a Shared Patient-Centered, Web-Based Medication Platform for Type 2 Diabetes Patients and Their Health Care Providers: Qualitative Study on User Requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]
28. Brady E, Segar J, Sanders C. Accessing support and empowerment online: The experiences of individuals with diabetes. Health Expect 2017 Oct;20(5):1088-1095 [FREE Full text] [doi: 10.1111/hex.12552] [Medline: 28718928]
29. Castensøe-Seidenfaden P, Husted GR, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a Self-Management App for Young People With Type 1 Diabetes: Methodological Challenges, Experiences, and Recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
30. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: User preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
31. Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, et al. A Mobile App to Improve Self-Management of Individuals With Type 2 Diabetes: Qualitative Realist Evaluation. J Med Internet Res 2018 Mar 16;20(3):e81 [FREE Full text] [doi: 10.2196/jmir.8712] [Medline: 29549070]
32. Dewi DS, Irfoni AR, Rahman A. Kansei Engineering Approach for Designing a Self-monitoring Blood Glucose Application. IJTech 2017 Apr 29;8(2):272-282. [doi: 10.14716/ijtech.v8i2.6144]
33. Drion I, Pameijer LR, van Dijk PR, Groenier KH, Kleefstra N, Bilo HJG. The Effects of a Mobile Phone Application on Quality of Life in Patients With Type 1 Diabetes Mellitus: A Randomized Controlled Trial. J Diabetes Sci Technol 2015 May 11;9(5):1086-1091 [FREE Full text] [doi: 10.1177/1932296815585871] [Medline: 25963412]
34. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
35. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the Feasibility of Use of An Online Dietary Assessment Tool (myfood24) in Women with Gestational Diabetes. Nutrients 2018 Aug 23;10(9):1147 [FREE Full text] [doi: 10.3390/nu10091147] [Medline: 30142898]
36. Husted GR, Weis J, Teilmann G, Castensøe-Seidenfaden P. Exploring the Influence of a Smartphone App (Young with Diabetes) on Young People's Self-Management: Qualitative Study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e43 [FREE Full text] [doi: 10.2196/mhealth.8876] [Medline: 29490897]
37. Jeon E, Park H. Development of the IMB Model and an Evidence-Based Diabetes Self-management Mobile Application. Healthc Inform Res 2018 Apr;24(2):125-138 [FREE Full text] [doi: 10.4258/hir.2018.24.2.125] [Medline: 29770246]
38. Jo S, Park H. Development and Evaluation of a Smartphone Application for Managing Gestational Diabetes Mellitus. Healthc Inform Res 2016 Jan;22(1):11-21 [FREE Full text] [doi: 10.4258/hir.2016.22.1.11] [Medline: 26893946]
39. Kim YJ, Rhee SY, Byun JK, Park SY, Hong SM, Chin SO, et al. A Smartphone Application Significantly Improved Diabetes Self-Care Activities with High User Satisfaction. Diabetes Metab J 2015 Jun;39(3):207-217 [FREE Full text] [doi: 10.4093/dmj.2015.39.3.207] [Medline: 26124991]
40. Klaassen R, Bul KCM, Op den Akker R, van der Burg GJ, Kato PM, Di Bitonto P. Design and Evaluation of a Pervasive Coaching and Gamification Platform for Young Diabetes Patients. Sensors (Basel) 2018 Jan 30;18(2):402 [FREE Full text] [doi: 10.3390/s18020402] [Medline: 29385750]
41. Knight BA, McIntyre HD, Hickman IJ, Noud M. Qualitative assessment of user experiences of a novel smart phone application designed to support flexible intensive insulin therapy in type 1 diabetes. BMC Med Inform Decis Mak 2016 Dec 15;16:119 [FREE Full text] [doi: 10.1186/s12911-016-0356-6] [Medline: 27629774]
42. Lamprinos I, Demski H, Mantwill S, Kabak Y, Hildebrand C, Ploessnig M. Modular ICT-based patient empowerment framework for self-management of diabetes: Design perspectives and validation results. Int J Med Inform 2016 Jul;91:31-43. [doi: 10.1016/j.ijmedinf.2016.04.006] [Medline: 27185507]
43. Muller I, Rowsell A, Stuart B, Hayter V, Little P, Ganahl K, et al. Effects on Engagement and Health Literacy Outcomes of Web-Based Materials Promoting Physical Activity in People With Diabetes: An International Randomized Trial. J Med Internet Res 2017 Jan 23;19(1):e21 [FREE Full text] [doi: 10.2196/jmir.6601] [Medline: 28115299]

XSL•FO
RenderX

44. Neinstein A, Wong J, Look H, Arbiter B, Quirk K, McCanne S, et al. A case study in open source innovation: developing the Tidepool Platform for interoperability in type 1 diabetes management. J Am Med Inform Assoc 2016 Mar;23(2):324-332 [FREE Full text] [doi: 10.1093/jamia/ocv104] [Medline: 26338218]
45. Nicholson WK, Beckham AJ, Hatley K, Diamond M, Johnson L, Green SL, et al. The Gestational Diabetes Management System (GooDMomS): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. BMC Pregnancy Childbirth 2016 Sep 21;16(1):277 [FREE Full text] [doi: 10.1186/s12884-016-1064-z] [Medline: 27654119]
46. Park S, Burford S, Nolan C, Hanlen L. The Role of Digital Engagement in the Self-Management of Type 2 Diabetes. Health Commun 2016 Dec;31(12):1557-1565. [doi: 10.1080/10410236.2015.1089468] [Medline: 27124817]
47. Peng W, Yuan S, Holtz BE. Exploring the Challenges and Opportunities of Health Mobile Apps for Individuals with Type 2 Diabetes Living in Rural Communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
48. Petersen M, Hempler NF. Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):91 [FREE Full text] [doi: 10.1186/s12911-017-0493-6] [Medline: 28651639]
49. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2015 Jul 21:172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
50. Quinn CC, Khokhar B, Weed K, Barr E, Gruber-Baldini AL. Older Adult Self-Efficacy Study of Mobile Phone Diabetes Management. Diabetes Technol Ther 2015 Jul;17(7):455-461 [FREE Full text] [doi: 10.1089/dia.2014.0341] [Medline: 25692373]
51. Ramadas A, Chan CKY, Oldenburg B, Hussien Z, Quek KF. A web-based dietary intervention for people with type 2 diabetes: development, implementation, and evaluation. Int J Behav Med 2015 Jun;22(3):365-373 [FREE Full text] [doi: 10.1007/s12529-014-9445-z] [Medline: 25274015]
52. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
53. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
54. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
55. Wake DJ, He J, Czesak AM, Mughal F, Cunningham SG. MyDiabetesMyWay: An Evolving National Data Driven Diabetes Self-Management Platform. J Diabetes Sci Technol 2016 Dec;10(5):1050-1058 [FREE Full text] [doi: 10.1177/1932296816648168] [Medline: 27162192]
56. Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z. Exploration of Users' Perspectives and Needs and Design of a Type 1 Diabetes Management Mobile App: Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11400 [FREE Full text] [doi: 10.2196/11400] [Medline: 30249580]
57. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
58. Campbell M, Escobar O, Fenton C, Craig P. The impact of participatory budgeting on health and wellbeing: a scoping review of evaluations. BMC Public Health 2018 Jul 03;18(1):822 [FREE Full text] [doi: 10.1186/s12889-018-5735-8] [Medline: 29970044]
59. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the Development of Mobile Phone Apps to Support Diabetes Self-Management: Systematic Review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
60. Azzopardi-Muscat N, Ricciardi W, Odone A, Buttigieg S, Zeegers Paget D. Digitalization: potentials and pitfalls from a public health perspective. Eur J Public Health 2019 Oct 01;29(Supplement_3):1-2 [FREE Full text] [doi: 10.1093/eurpub/ckz169] [Medline: 31738438]
61. Wali S, Keshavjee K, Demers C. Moving towards sustainable electronic health applications. In: Heston TF, editor. eHealth - Making Health Care Smarter. London, UK: IntechOpen; 2018.
62. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
63. Wilson K, Gertz B, Arenth B, Salisbury N. The Journey to Scale: Moving Together Past Digital Health Pilots. Seattle, WA: PATH; 2014. URL: https://path.azureedge.net/media/documents/TS_dhs_journey_to_scale.pdf [accessed 2020-02-24]
64. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]

XSL•FO
RenderX

65. Eysenbach G, CONSORT-EHEALTH Group E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
66. World Health Organization. Classification of Digital Health Interventions (WHO/RHR/18.06). Geneva, Switzerland: World Health Organization; 2018.
67. Vera F, Noël R, Taramasco C. Standards, Processes and Instruments for Assessing Usability of Health Mobile Apps: A Systematic Literature Review. Stud Health Technol Inform 2019 Aug 21;264:1797-1798. [doi: 10.3233/SHTI190653] [Medline: 31438349]
68. Research2Guidance. mHealth Developer Economics: Connectivity in Digital Health. Berlin, Germany: Research2Guidance; 2018. URL: https://research2guidance.com/product/connectivity-in-digital-health/ [accessed 2019-11-26]
69. Research2Guidance. mHealth App Economics 2017: Current Status and Future Trends in Mobile Health. Berlin, Germany: Research2Guidance; 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-11-26]
70. Muralidharan S, Ranjani H, Anjana RM, Allender S, Mohan V. Mobile Health Technology in the Prevention and Management of Type 2 Diabetes. Indian J Endocrinol Metab 2017;21(2):334-340 [FREE Full text] [doi: 10.4103/ijem.IJEM_407_16] [Medline: 28459035]
71. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
72. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]

Abbreviations

HbA_{1c}: glycated hemoglobin
HCPs: health care professionals
T1D: type 1 diabetes
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by E Ding; comments to author 20.03.20; revised version received 24.04.20; accepted 13.05.20; published 06.07.20.

Please cite as:
Larbi D, Randine P, Årsand E, Antypas K, Bradway M, Gabarron E
Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review
J Med Internet Res 2020;22(7):e18480
URL: https://www.jmir.org/2020/7/e18480
doi:10.2196/18480
PMID:32628125

©Dillys Larbi, Pietro Randine, Eirik Årsand, Konstantinos Antypas, Meghan Bradway, Elia Gabarron. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials

Hengfen Dai[1*], MS; Caiyun Zheng[2*], MS; Chun Lin[3*], PhD; Yan Zhang[1], PhD; Hong Zhang[1], BS; Fan Chen[1], BS; Yunchun Liu[1], BS; Jingwen Xiao[1], MS; Chaoxin Chen[4], MS

[1]Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China

[2]Fuqing City Hospital, Fuzhou, China

[3]School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China

[4]School of Pharmacy, Fujian Medical University, Fuzhou, China

[*]these authors contributed equally

Corresponding Author:
Caiyun Zheng, MS
Fuqing City Hospital
267 Qingrong Avenue
Fuqing
Fuzhou, 350300
China
Phone: 86 059185153395
Email: saiun1992@163.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/8/e22761/

Abstract

Background: An increasing number of patients have received prophylactic or therapeutic oral anticoagulants (OACs) for thromboembolic complications of diseases. The use of OACs is associated with both clinical benefits and risks. Considering the challenges imposed by this class of drugs, as well as the enormous progress made in portable device technology, it is possible that technology-based interventions may improve clinical benefits for patients and optimize anticoagulation management.

Objective: This study was designed to comprehensively evaluate the role of technology-based interventions in the management of OACs.

Methods: We searched 6 databases—PubMed, EMBASE, Cochrane, Cumulative Index to Nursing and Allied Health Literature, Scopus, and PsycINFO—to retrieve relevant studies published as of November 1, 2019, to evaluate the effect of technology-based interventions on oral anticoagulation management. RevMan (version 5.3; Cochrane) software was used to evaluate and analyze clinical outcomes. The methodological quality of studies was assessed by the Cochrane risk of bias tool.

Results: A total of 15 randomized controlled trials (RCTs) were selected for analysis. They reported data for 2218 patients (1110 patients in the intervention groups and 1108 patients in the control groups). A meta-analysis was performed on the effectiveness and safety data reported in the RCTs. Technology-based interventions significantly improved the effectiveness of oral anticoagulation management (mean difference [MD]=6.07; 95% CI 0.84-11.30; I^2=72%; *P*=.02). The safety of oral anticoagulation management was also improved, but the results were not statistically significant. Bleeding events were reduced (major bleeding events MD=1.02; 95% CI 0.78-1.32; I^2=0%; *P*=.90; minor bleeding events MD=1.06, 95% CI 0.77-1.44; I^2=41%; *P*=.73) and thromboembolism events were reduced (MD=0.71; 95% CI 0.49-1.01; I^2=0%; *P*=.06). In general, patients were more satisfied with technology-based interventions, which could also improve their knowledge of anticoagulation management, improve their quality of life, and reduce mortality and hospitalization events.

Conclusions: Using technology to manage OACs can improve the effectiveness and safety of oral anticoagulation management, result in higher patient satisfaction, and allow greater understanding of anticoagulation.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e18386) doi:10.2196/18386

KEYWORDS

technology-based; oral anticoagulation management; meta-analysis; randomized controlled trials; telehealth; warfarin

Introduction

Oral Anticoagulation Management

Oral anticoagulants (OACs) have been used for decades, especially warfarin, which has been in use since the 1950s. Until recently warfarin has been a fundamental drug in clinical anticoagulant therapy, and although the results of two previously published meta-analyses of atrial fibrillation suggested that novel OACs are not inferior to warfarin in terms of safety and efficacy [1,2], warfarin has the advantage of being inexpensive and having a wide range of indications. Warfarin is widely used in conditions that are prone to thrombosis, such as atrial fibrillation, irregular heartbeat, myocardial infraction, artificial heart valve replacement, recurrent stroke, deep vein thrombosis, and pulmonary embolism [3]. However, it remains challenging to balance effectiveness and safety in treatment [4]. Statistical data from relevant literature estimated that over 6 million patients in the United States have received anticoagulant therapy [5], leading to increased risks of bleeding, thromboembolism, hospitalization, and mortality. Currently, rough statistics indicate that about 1 in 10 surgical patients in the United States receive OACs [6], and approximately 2 million people start warfarin therapy in the United States every year [7]. Therefore, effective anticoagulation management measures are urgently needed.

Technology-Based Interventions for Oral Anticoagulation Management

With the continuous development of telemedicine health service technology, more and more technological devices have been applied to help in the management of patients with chronic diseases, including those with diabetes, hypertension, heart failure, and others [8]. Telemedicine is defined as a long-distance medical practice that can be characterized by health services through a wide range of technical applications and services [9]. It is a new way of providing high-quality resources to primary medical institutions through remote consultation on the basis of internet convenience. Through this approach, medical services and medical activities are mainly carried out through computers, various remote communications, medical technology, and medical equipment. These remote forms of communication enable contact between patients and medical personnel, medical institutions, and medical equipment. This process, in turn, can assist in diagnoses, treatment, monitoring, and follow-up [10]. As a result, telemedicine has become an increasingly popular model for providing accurate international normalized ratio (INR) monitoring for patients taking warfarin. Other terms used to describe telemedicine include connected health [11], mobile health (mHealth) [12], and electronic health (eHealth) [13], which are collectively called technology-based interventions [14]. There is interest in learning whether these interventions can assist medication management and enhance patient compliance, with results suggesting some benefit [15].

Aim of the Study

Telemedicine technology has the potential to enhance multiple aspects of anticoagulant therapy management [16], including patient education, symptom monitoring, follow-up, and encouragement and tracking of medication adherence, given its accessibility by telephone, internet, voicemail messaging, and apps. However, no systematic review or meta-analysis has been published to summarize what is currently known on this topic. Thus, we aimed to evaluate the effectiveness and safety of technology-based interventions in OAC management by performing a meta-analysis.

Methods

Literature Search

Eligible studies were identified by searching PubMed, EMBASE, Cochrane, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, Scopus, and other relevant databases, and the results were combined using literature traceability methods. Searches were conducted on November 1, 2019. Search keywords included *warfarin therapy+*, *oral anticoagulation management+*, *telephone*, *eHealth*, *apps*, and *telemedicine+*. Search strategies are detailed in Multimedia Appendix 1. Only papers written in English were considered. No restriction regarding publication date was applied.

Inclusion and Exclusion Criteria

Inclusion Criteria

The studies included in the analysis met the following criteria: (1) the study was a randomized controlled trial (RCT); (2) the subjects were taking warfarin; (3) technology-based interventions were used to manage OACs; (4) results included the time in therapeutic range (TTR), bleeding, and thromboembolism events; and (5) the results were reported in English.

Exclusion Criteria

Exclusion criteria included: (1) studies that were retrospective, observational, reviews, model research, literature reviews, or conference summaries; (2) the results of the study did not involve the TTR, bleeding, and thromboembolism events; (3) the study was a duplicate report; and (4) technical intervention was only used as a means to collect data.

Document Screening and Data Extraction

All references were initially screened by title and abstract by 2 reviewers for relevance. Finally, full-text analysis for eligibility was performed independently by 2 authors, HD and CL. Disagreements were discussed and resolved by consensus or third-party arbitration.

The required data were extracted by a researcher using a literature data extraction table, and another researcher confirmed the accuracy and authenticity of the data. The extracted content

XSL•FO
RenderX

included (1) basic information of study, such as research topic, author, and date; (2) baseline characteristics of the study subjects, such as sample size, median age, and sex; (3) follow-up time for interventions; (4) efficacy and safety information after interventions, such as TTR, bleeding, and thromboembolism events; and (5) other outcome indicators, including time within expanded target INR range, mortality, and hospitalization events.

Literature Quality Evaluation

Risk of bias assessment of the included RCTs was performed using the Cochrane risk of bias tool based on the *Cochrane Handbook for Systematic Reviews of Interventions*' literature evaluation criteria [17].

Statistical Analysis

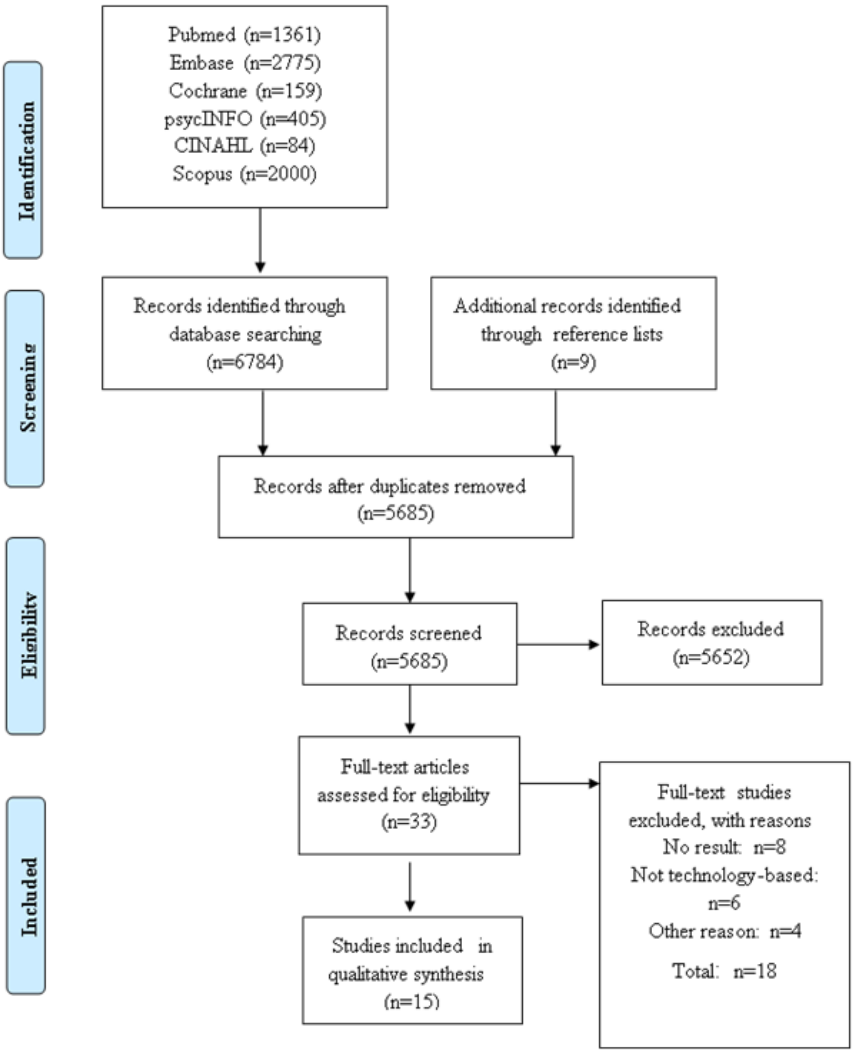
The meta-analysis of RCTs was performed using RevMan (version 5.3; Cochrane) software. Heterogeneity was assessed using a chi-square test, and quantitative analysis was performed using I^2. Values of $P \geq .05$ and $I^2 \leq 50\%$ were considered to represent no heterogeneity, in which case a fixed-effect model was used. If $P < .05$ and $I^2 > 50\%$, a random-effects model was used [18].

Results

Search Results

A total of 6784 papers were retrieved from the systematic literature search, and 9 papers were retrieved by other means, totaling 6793 papers. After removing duplicate studies, 2 authors independently reviewed and excluded another 5652 studies that did not meet predetermined selection criteria, based on the title and abstract of each paper. After reading 33 eligible full-text studies, 18 were excluded and 15 were selected for inclusion in the analysis [12,19-31]. The systematic search results are shown in Figure 1.

Figure 1. Flow chart of the systematic literature review process. CINAHL: Cumulative Index to Nursing and Allied Health Literature.



Basic Characteristics and Quality Evaluation of the Literature

The 15 RCTs included a total of 2218 patients: 1110 patients in the technology-based intervention group and 1108 patients in the traditional intervention group. General information (such as sample size, median age, sex, and specific intervention measures) from the studies is shown in Table 1. The Cochrane systematic evaluation method was used for quality evaluation, and overall, the included studies had low risk of bias and were high quality, as shown in Figure 2.

XSL•FO
RenderX

Table 1. Baseline characteristics of the included studies.

Study	Primary indication for therapy (n)	N[a]	Age in years, mean (SD)	Men, n (%)	Follow-up duration, months	Intervention	Description of technology-based interventions
Beyth 2000 [19]	VTE[b] (124) AF[c] (54) Cerebrovascular disease (49) HVR[d] (36) Others (41) PVD[e] (14) Myocardial infarction (7)	325	74.7 (6.75)	141 (43.38)	6	telephone	Recommendations for dose and subsequent INR[f] testing
Sidhu 2001 [20]	HVR (83)	83	60.9 (—)	46 (55.42)	24	telephone	Medical advice if patient's INR was too high (>4.0) or too low (< 1.5) Medical advice for any bleeding or thromboembolic events
Fitzmaurice 2002 [21]	—	49	66 (—)	37 (75.51)	6	software	Medical advice to override the dosing algorithm
Khan 2004 [22]	AF (79)	79	74 (—)	45 (56.96)	6	telephone	Recommendations for dose
Staresinic 2006 [23]	AF (79) VTE (23) Cerebrovascular accident (19) Coronary artery disease (12) PVR[g] (36) Others (23)	192	69.3 (9.1)	187 (97.40)	36	interim telephone	Telephone follow-up
Chan 2006 [24]	AF (72) HVR (24) DVT (17) PE[h] (9) Cerebrovascular accident (4) Valvular heart diseases (5) Cardiomyopathy (2) Miscellaneous (4)	137	59 (14)	62 (45.26)	24	telephone	Consultation for difficult INR control or adherence issues
Lalonde 2008 [25]	AF (149) DVT[i] (68) PE (26) Stroke (11) Cardiomyopathy (9) Myocardial infarction (8) Others (12)	250	65.45 (11.75)	128 (51.20)	6	telephone	Contact with pharmacist
Soliman 2008 [26]	Elective mechanical aortic valve replacement (—)	58	56 (8.95)	—	12	internet	Verify the anticoagulant dose on the website Pass the anticoagulant dosage exam on the website

XSL•FO
RenderX

Study	Primary indication for therapy (n)	N[a]	Age in years, mean (SD)	Men, n (%)	Follow-up duration, months	Intervention	Description of technology-based interventions
Schillig 2011 [27]	VTE (100) AF (302) Others (98)	500	66.05 (15.25)	276 (55.20)	1	telephone	Contact responsible physician and anticoagulation clinic that provided dosing regimen
Verret 2012 [28]	AF or flutter (58) PVR (—)	114	57.7 (10.5)	78 (68.42)	4	voicemail message	Communicate Provide INR results and perform necessary adjustments
Bungard 2012 [29]	AF (49) VTE (8) Others (5)	62	73 (—)	38 (61.29)	6	telephone	Discuss any potential factors that may influence the INR result Warfarin dosing instructions Schedule a follow-up phone call
Lakshmi 2013 [30]	Mitral valve replacement (16) AF (45) DVT (2) PE (2) Valvotomy (1) Bioprosthetic valve (2) Other cardiac risk (12)	80	55.97 (12.85)	52 (60.00)	6	telephonic contact	Call the clinical pharmacist for clarification on any anticoagulation-related issues
Brasen 2018 [13]	AF (56) DVT/PE (14) Valvular heart disease (2) Various diagnoses (cardiomyopathy, aneurism, thrombophilia, and stroke) (15)	87	69.4 (—)	69 (79.30)	10	telemedicine software	Physician could inform patient of result, new dosage, and date for next INR measurement
Ayutthaya 2018 [31]	Valvular heart disease (14) Mechanical prosthetic valves (3) AF (31) DVT (14) PE (1)	50	57.65 (10.95)	30 (60.00)	3	telephone	Pharmacists perform medicine use review by asking patients about problems/obstacles with managing warfarin, including adverse events and complications Assess medication adherence Provide reminders for the next scheduled visits

XSL•FO
RenderX

Study	Primary indication for therapy (n)	N[a]	Age in years, mean (SD)	Men, n (%)	Follow-up duration, months	Intervention	Description of technology-based interventions
Liang 2019 [32]	Non-valvular AF (80) Valvular AF (8) DVT (30) PE (12) Others (13) Multiple indications (9)	152	61.3 (15.4)	85 (55.92)	6	telephone	Pharmacists mainly assessed and reinforced adherence to warfarin and INR monitoring Education and recommendations according to participants' recent INR assessment, self-reported medication or dietary changes, and anticoagulation-related complications

[a]N: total number of participants in the study.

[b]VTE: venous thromboembolism.

[c]AF: atrial fibrillation.

[d]HVR: heart valve replacement.

[e]PVD: peripheral vascular disease.

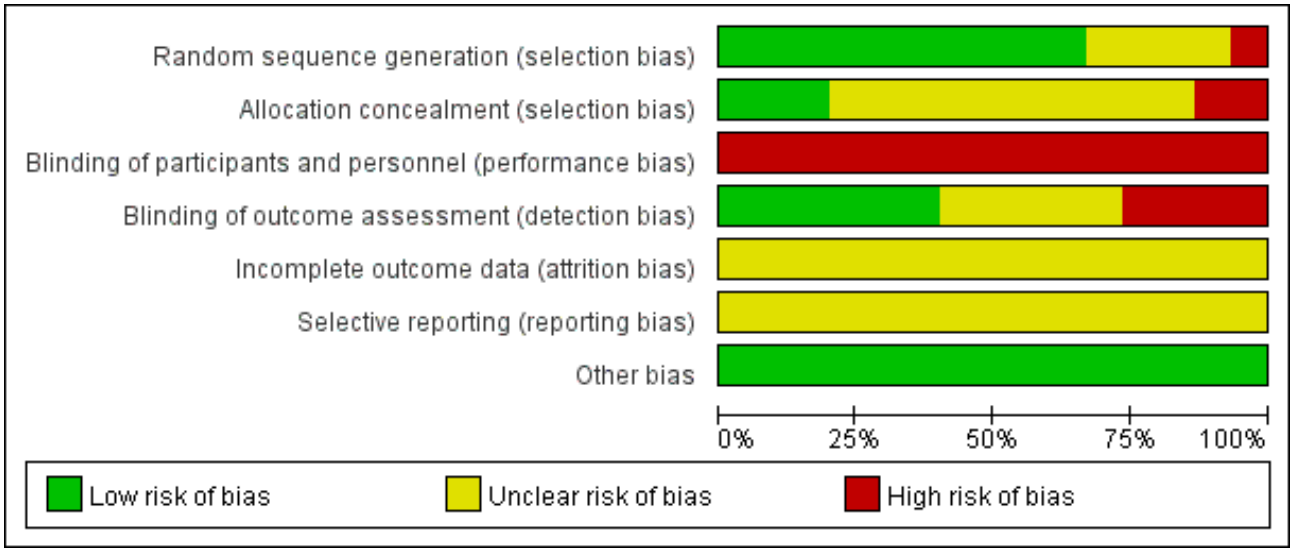
[f]INR: international normalized ratio.

[g]PVR: prosthetic valve replacement.

[h]PE: pulmonary embolism.

[i]DVT: deep vein thrombosis.

Figure 2. Bias risk map from the Cochrane systematic evaluation method to evaluate the quality of the included randomized controlled trials.



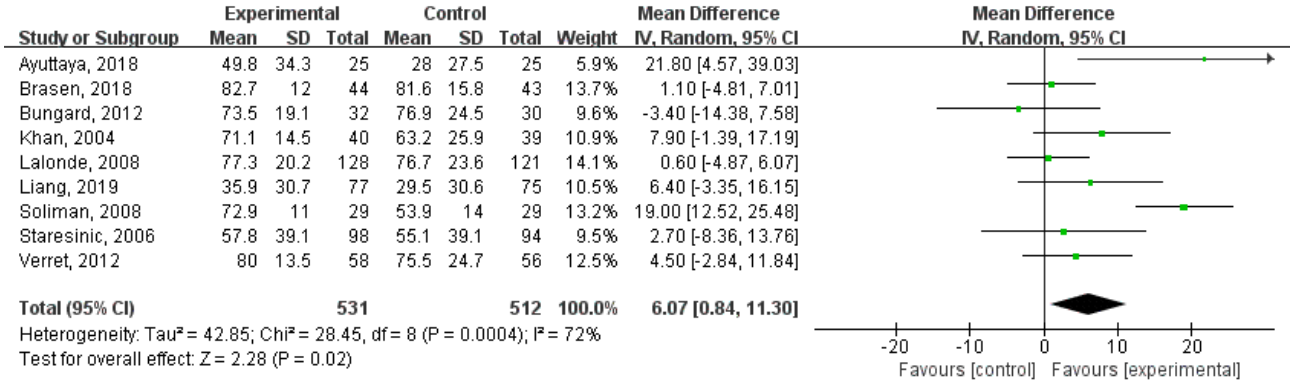
Meta-Analysis

Time Within Target INR Range

Of the 15 included studies, 9 included TTR values, for a total of 1043 patients. A random-effects analysis was used for the meta-analysis of these 9 RCTs. As shown in Figure 3, the TTR of the technology-based intervention group was significantly higher than that of the control group (mean difference [MD]=6.07; 95% CI 0.84-11.30; I^2=72%; *P*=.02).

XSL•FO
RenderX

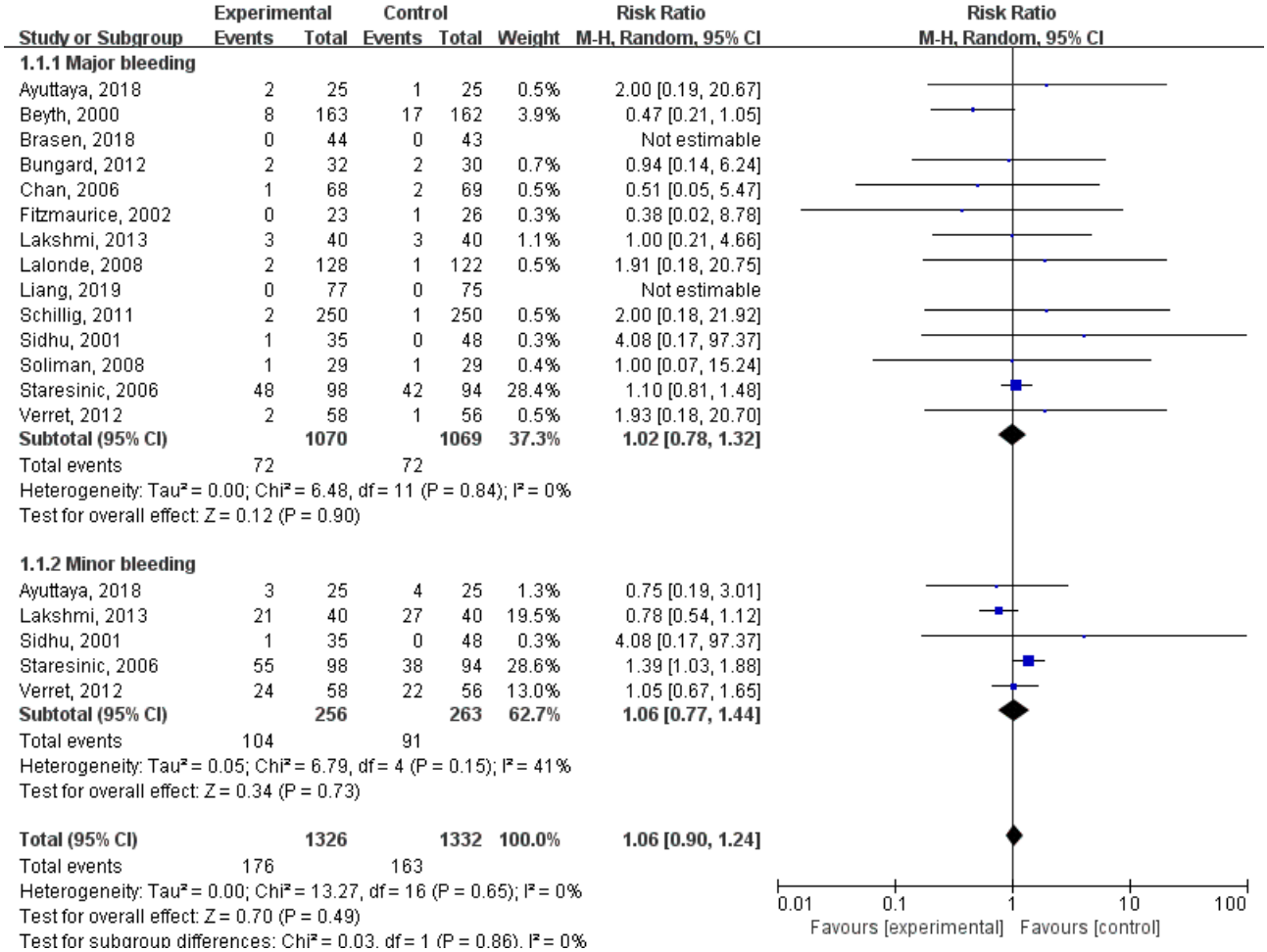
Figure 3. Results of a meta-analysis of the effects of technology-based interventions on time within target international normalized ratio range in oral anticoagulation management. IV: independent variable. Random: random effect model.



Bleeding Events

Of the 15 included studies, 14 included major bleeding values, for a total of 2139 patients. A random-effects model was used for the meta-analysis of these 14 RCTs. There were fewer major bleeding events in the technology-based intervention group than in the control group, but the difference was not statistically significant. (MD=1.02; 95% CI 0.78-1.32; I^2=0%; *P*=.90). There were 5 papers that included minor bleeding values, for a total of 519 patients. A random-effects analysis model was used for the meta-analysis of these 5 RCTs. There were fewer minor bleeding events in the technology-based intervention group than in the control group, but the difference was not statistically significant (MD=1.06; 95% CI 0.77-1.44; I^2=41%; *P*=.73). Major and minor bleeding event analyses are shown in Figure 4.

Figure 4. Results of a meta-analysis of the effects of technology-based interventions on major bleeding and minor bleeding events in oral anticoagulation management. M-H: Mantel-Haenszel method. Random: random effect model.

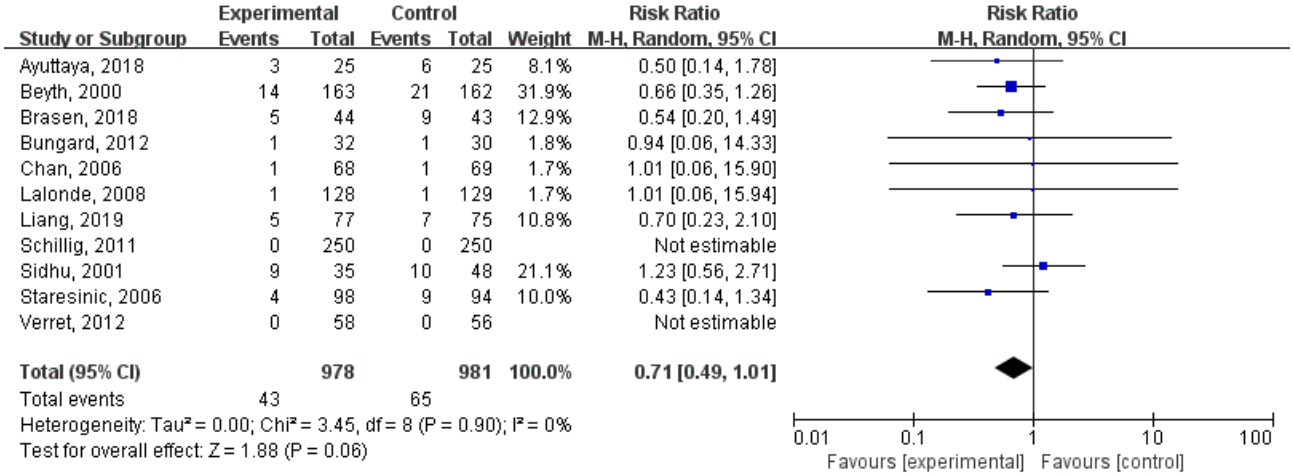


XSL•FO
RenderX

Thromboembolism Events

Of the 15 included papers, 11 had thromboembolism values, for a total of 1959 patients. A random-effects model was used for the meta-analysis of these 11 RCTs. As shown in Figure 5, there were fewer thromboembolism events in the technology-based intervention group than in the control group, but the difference was not statistically significant (MD=0.71; 95% CI 0.49-1.01; I^2=0%; *P*=.06).

Figure 5. Results of a meta-analysis of the effects of technology-based interventions on thromboembolism events in oral anticoagulation management. M-H: Mantel-Haenszel method. Random: random effect model.



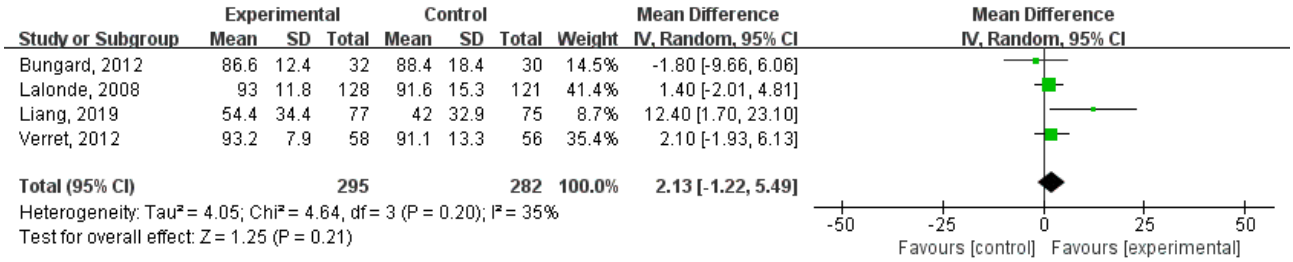
Sensitivity Analysis Results

The data were analyzed with fixed- and random-effects models, and the consistency of these results reflects the reliability of the combined results to some extent. The two effect models were used to analyze the combined effect of each risk factor and calculate 95% CIs. The results were similar, indicating that the results of this study are stable.

Other Results

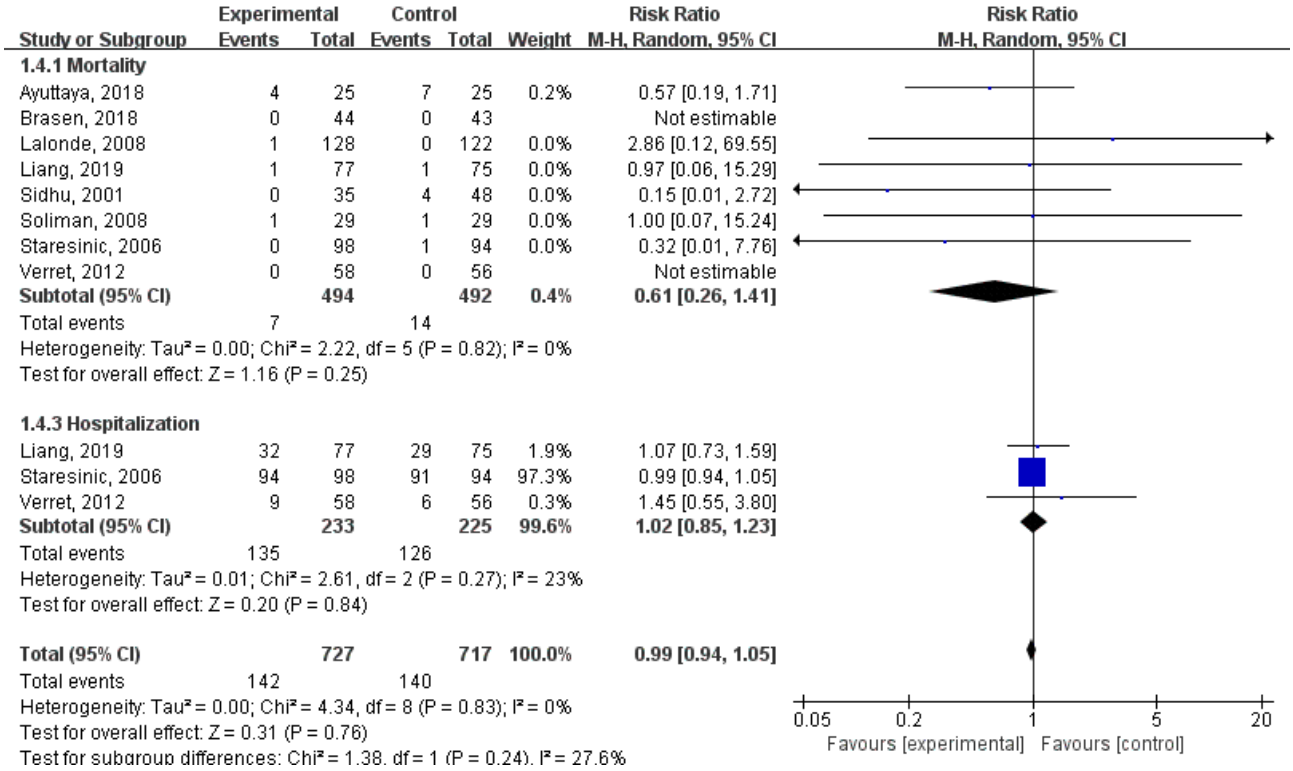
The time within expanded target INR range of the technology-based intervention group was higher than that of the control group, but the difference was not statistically significant (MD=2.13; 95% CI –1.22 to 5.49; I^2=35%; *P*=.21; Figure 6). There were fewer mortalities and hospitalization events in the technology-based intervention group than in the control group, but the difference was not statistically significant (mortality MD=0.61; 95% CI 0.26 to 1.41; I^2=0%; *P*=.25; hospitalization MD=1.02; 95% CI 0.85 to 1.23; I^2=23%; *P*=.84; Figure 7). There were 3 papers [28,30,32] that mentioned that patients' knowledge of anticoagulation was significantly improved through telemedicine intervention. There were 4 studies [22,25,26,28] that involved quality-of-life assessments; quality of life was higher in the intervention group than in the control group and it improved from baseline. There were 2 papers [21,25] that mentioned the higher costs in the technical intervention group than in the control group, and 1 paper [24] found that the costs for the nontechnical intervention group were higher. Patient satisfaction surveys were conducted in 5 studies [24,25,28-30], all of which indicated that the technical intervention group had higher satisfaction. There were 4 papers [25,26,28,32] that included data on INR tests, and the technical intervention group tended to have larger numbers of INR tests performed.

Figure 6. Results of a meta-analysis of the effects of technology-based interventions on time within extended target international normalized ratio range events in oral anticoagulation management. IV: independent variable. Random: random effect model.



XSL•FO
RenderX

Figure 7. Results of a meta-analysis of the effects of technology-based interventions on mortality and hospitalization events in oral anticoagulation management. M-H: Mantel-Haenszel method. Random: random effect model.



Discussion

Principal Findings

We performed a meta-analysis of 15 RCTs to determine whether technology-based interventions were beneficial for patients compared with traditional interventions in oral anticoagulation management. TTR reflects the quality and effectiveness of warfarin and is an important determinant of bleeding and thromboembolism in anticoagulation management. A total of 9 studies were involved in the TTR analysis, and almost all studies reported that technology-based interventions were effective. The number of bleeding (including major and minor) events in the technology-based intervention groups was lower than that in the control groups, but the difference was not statistically significant. Outcomes were similar for the thromboembolism events. Time within expanded target INR range, mortality, and hospitalization were not significantly different between the technology-based management model and traditional management model. However, patients' anticoagulation knowledge, patient satisfaction, quality of life, and number of INR tests performed were positively affected by technology-based management.

Oral anticoagulation management can be challenging when ensuring safety of the patient; as long as warfarin is used, regular monitoring is necessary [33]. The technology-based interventions used to manage oral anticoagulation typically include multiple methods, including telephone, internet, voicemail messaging, and apps. These tools are readily available in this era of technological advancement, and they reduce requirements for capacity and waiting time in outpatient clinics. Through this process, mHealth technologies have systematically demonstrated that they can help improve early diagnosis and treatment of atrial fibrillation and significantly reduce the cost of illness associated with atrial fibrillation [34]. One such example of mHealth technology is SintromacWeb, a management system of internet-based telecontrol tools that has been recognized to have better effectiveness and safety at oral anticoagulation therapy management than a conventional approach does [35]. Technology-based interventions can effectively and professionally teach patients about anticoagulation, provide interpretation of indicators, and provide a timely dose adjustment program.

Comment on Results

It is worth mentioning that the mean age of the patients included in each study was over 60 years, so these patients may have had a relatively poor grasp on learning how to use novel websites and apps. The telephone is an ultra-portable electronic device, and it was chosen as the intervention tool in most of the 15 RCT studies included. Telephone intervention is especially suitable for older patients who are not accustomed to using computers or online services, or who may have difficulties in vision, finger dexterity, and mental state.

Our overall results showed that technology-based interventions significantly improved the effectiveness of oral anticoagulation management but did not significantly increase the safety or other results of oral anticoagulation management. These results may be because of the different follow-up times, which ranged from 1 month to 36 months. The greatest difference between the intervention and control groups in the frequency of major bleeding events occurred in the first week (when anticoagulation therapy was started in the hospital) and after the first month of therapy (when outpatient therapy was presumably stabilized).

XSL•FO
RenderX

The 2 groups differed little during the second through fourth weeks of therapy, when patients were generally discharged and warfarin was first monitored in the outpatient setting [19]. That is to say, most patients might attach great importance to any type of intervention early on and then slowly relax later.

Limitations

We acknowledge that our research has some limitations. The included studies had relatively small sample sizes and varying follow-up times, with some as short as 1 month. Warfarin is the oldest and best known OAC, and as long as warfarin is used, monitoring is necessary. Therefore, in order to more comprehensively assess the safety of OACs, longer follow-up times are needed. Furthermore, patients' adherence to technology-based interventions may be lacking. It would be ideal if intervention compliance studies were added during follow-up. Despite these limitations, this study is the first to comprehensively analyze the effect of technology-based interventions on anticoagulation management. Therefore, larger sample sizes and longer clinical RCTs are needed to further evaluate the impact of technology-based interventions on anticoagulation management.

Comparison With Prior Work

To the best of our knowledge, this is the first meta-analysis of RCTs that evaluates the effectiveness and safety of technology-based interventions for oral anticoagulation management. In published systematic meta-analyses and reviews, most of the literature evaluating the effects of technology-based interventions has focused on atrial fibrillation [34] and cancer [36]. Most studies regarding the administration of OACs have focused on the effects of managers on the effectiveness and safety of anticoagulation management, such as the differences between pharmacist management and physician management or between patient self-management and anticoagulation clinic management [37-41]. The intervention method we chose is a technology-based approach that might have also been included in other studies, such as pharmacist-managed or self-managed interventions [37,41].

Although the study of telemedicine is not specific to OACs and is primarily qualitative, previous studies have highlighted similar challenges [42-44]. Clinical pharmaceutical care provided by Niznik et al [43] through telemedicine (mainly by telephone) in inpatient or outpatient settings was found to have an overall positive impact on outcomes related to clinical disease management, patient self-management, and adherence to the management of various chronic diseases. The common ground in the studies that had a positive impact on outcomes was the use of continuous or regular patterns of care, including telephone interaction and frequent monitoring and intervention [43]. Xia et al's [44] study was conducted to explore the effects of online and offline management of anticoagulants on therapeutic efficacy and adverse reactions. Considering the convenience and economy of technology-based interventions, online anticoagulation management is more suitable for patients with stable conditions and for whom transportation may be difficult, such as those with disabilities or who live far from the hospital. Lee et al's study [42] suggested that most outcomes for telemedicine were similar to those for conventional medical care, but the incidence of major thromboembolism events was significantly lower in the telemedicine group. However, the papers they included had a higher risk of bias and were of a lower quality study design, and the level of evidence supporting this conclusion was very low.

Conclusions

This meta-analysis explored the effects of technology-based interventions on oral anticoagulation management. The results demonstrate that the technology-based intervention group had significantly improved TTR compared with the traditional intervention group and that there were no significant differences between the 2 intervention models in terms of time within expanded target INR range and the incidence of major bleeding events, minor bleeding events, thromboembolic events, mortality, and hospitalization. Of the different management options, telephone intervention was found to be the most widely used and most convenient means of technology-based interventions; it enables patients to get a professional reply quickly and is not restricted by the patient's ability to use the internet. OAC management through technology-based intervention appears to be superior to OAC management through traditional intervention and may provide more convenient and higher quality anticoagulation services for patients. Further research is needed to explore more optimal technology-based interventions in oral anticoagulation management in a wider array of health care settings.

Acknowledgments

Funds were provided by the Fuzhou Science and Technology Project (Grant number: 2018-S-105-2) and the Startup Fund for Scientific Research, Fujian Medical University (Grant number: 2017XQ1206).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy to find literature on the use of technology-based interventions for oral anticoagulation management.
[DOC File , 36 KB - jmir_v22i7e18386_app1.doc]

References

XSL•FO
RenderX

1. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. Lancet 2014 Mar 15;383(9921):955-962. [doi: 10.1016/S0140-6736(13)62343-0] [Medline: 24315724]
2. Chatterjee S, Sardar P, Biondi-Zoccai G, Kumbhani DJ. New oral anticoagulants and the risk of intracranial hemorrhage: traditional and Bayesian meta-analysis and mixed treatment comparison of randomized trials of new oral anticoagulants in atrial fibrillation. JAMA Neurol 2013 Dec;70(12):1486-1490. [doi: 10.1001/jamaneurol.2013.4021] [Medline: 24166666]
3. Wardrop D, Keeling D. The story of the discovery of heparin and warfarin. Br J Haematol 2008 Jun;141(6):757-763. [doi: 10.1111/j.1365-2141.2008.07119.x] [Medline: 18355382]
4. Lee MTM, Klein TE. Pharmacogenetics of warfarin: challenges and opportunities. J Hum Genet 2013 Jun;58(6):334-338 [FREE Full text] [doi: 10.1038/jhg.2013.40] [Medline: 23657428]
5. Shah Z, Masoomi R, Tadros P. Managing Antiplatelet Therapy and Anticoagulants in Patients with Coronary Artery Disease and Atrial Fibrillation. J Atr Fibrillation 2015 Dec;8(4):1318 [FREE Full text] [doi: 10.4022/jafib.1318] [Medline: 27957230]
6. Spyropoulos AC, Douketis JD. How I treat anticoagulated patients undergoing an elective procedure or surgery. Blood 2012 Oct 11;120(15):2954-2962. [doi: 10.1182/blood-2012-06-415943] [Medline: 22932800]
7. Douketis JD, Spyropoulos AC, Spencer FA, Mayr M, Jaffer AK, Eckman MH, et al. Perioperative management of antithrombotic therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012 Feb;141(2 Suppl):e326S-e350S [FREE Full text] [doi: 10.1378/chest.11-2298] [Medline: 22315266]
8. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 04;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
9. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical and Judicial Affairs American Medical Association. Ethical practice in Telehealth and Telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
10. Nelson R. Telemedicine and Telehealth: The Potential to Improve Rural Access to Care. Am J Nurs 2017 Jun;117(6):17-18. [doi: 10.1097/01.NAJ.0000520244.60138.1c] [Medline: 28541980]
11. Krumm H, Reiss N, Burkert M, Schmidt T, Biehs S, Bohr C, et al. Development of a Computer-Aided Dosage and Telemonitoring System for Patients Under Oral Anticoagulation Therapy. Stud Health Technol Inform 2018;248:188-195. [Medline: 29726436]
12. Spyropoulos AC, Myrka A, Triller DM, Ragan S, York C, King JM, et al. Uptake and Utilization of the Management of Anticoagulation in the Periprocedural Period App: Longitudinal Analysis. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11090 [FREE Full text] [doi: 10.2196/11090] [Medline: 30578235]
13. Brasen CL, Madsen JS, Parkner T, Brandslund I. Home Management of Warfarin Treatment Through a Real-Time Supervised Telemedicine Solution: A Randomized Controlled Trial. Telemed J E Health 2019 Feb;25(2):109-115. [doi: 10.1089/tmj.2017.0260] [Medline: 29877778]
14. Kane JM. Technology-based interventions in health care. Epidemiol Psychiatr Sci 2014 Dec;23(4):323-326 [FREE Full text] [doi: 10.1017/S2045796014000444] [Medline: 25154596]
15. Deeken F, Rezo A, Hinz M, Discher R, Rapp MA. Evaluation of Technology-Based Interventions for Informal Caregivers of Patients With Dementia-A Meta-Analysis of Randomized Controlled Trials. Am J Geriatr Psychiatry 2019 Apr;27(4):426-445. [doi: 10.1016/j.jagp.2018.12.003] [Medline: 30642650]
16. Seaburg L, Hess EP, Coylewright M, Ting HH, McLeod CJ, Montori VM. Shared decision making in atrial fibrillation: where we are and where we should be going. Circulation 2014 Feb 11;129(6):704-710. [doi: 10.1161/CIRCULATIONAHA.113.004498] [Medline: 24515956]
17. Higgins JPT, Altman DG, GØtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
18. Jin YZ, Yan S, Yuan WX. Effect of isometric handgrip training on resting blood pressure in adults: a meta-analysis of randomized controlled trials. J Sports Med Phys Fitness 2017;57(1-2):154-160. [doi: 10.23736/S0022-4707.16.05887-4] [Medline: 26558836]
19. Beyth RJ, Quinn L, Landefeld CS. A multicomponent intervention to prevent major bleeding complications in older patients receiving warfarin. A randomized, controlled trial. Ann Intern Med 2000 Nov 07;133(9):687-695. [doi: 10.7326/0003-4819-133-9-200011070-00010] [Medline: 11074901]
20. Sidhu P, O'Kane HO. Self-managed anticoagulation: results from a two-year prospective randomized trial with heart valve patients. Ann Thorac Surg 2001 Nov;72(5):1523-1527. [doi: 10.1016/s0003-4975(01)03049-1] [Medline: 11722037]
21. Fitzmaurice DA, Murray ET, Gee KM, Allan TF, Hobbs FDR. A randomised controlled trial of patient self management of oral anticoagulation treatment compared with primary care management. J Clin Pathol 2002 Nov;55(11):845-849 [FREE Full text] [doi: 10.1136/jcp.55.11.845] [Medline: 12401823]

XSL•FO
RenderX

22. Khan TI, Kamali F, Kesteven P, Avery P, Wynne H. The value of education and self-monitoring in the management of warfarin therapy in older patients with unstable control of anticoagulation. Br J Haematol 2004 Aug;126(4):557-564. [doi: 10.1111/j.1365-2141.2004.05074.x] [Medline: 15287950]
23. Staresinic AG, Sorkness CA, Goodman BM, Pigarelli DW. Comparison of outcomes using 2 delivery models of anticoagulation care. Arch Intern Med 2006 May 08;166(9):997-1002. [doi: 10.1001/archinte.166.9.997] [Medline: 16682573]
24. Chan FWH, Wong RSM, Lau WH, Chan TYK, Cheng G, You JHS. Management of Chinese patients on warfarin therapy in two models of anticoagulation service - a prospective randomized trial. Br J Clin Pharmacol 2006 Nov;62(5):601-609 [FREE Full text] [doi: 10.1111/j.1365-2125.2006.02693.x] [Medline: 17061966]
25. Lalonde L, Martineau J, Blais N, Montigny M, Ginsberg J, Fournier M, et al. Is long-term pharmacist-managed anticoagulation service efficient? A pragmatic randomized controlled trial. Am Heart J 2008 Jul;156(1):148-154. [doi: 10.1016/j.ahj.2008.02.009] [Medline: 18585510]
26. Soliman Hamad MA, van Eekelen E, van Agt T, van Straten AHM. Self-management program improves anticoagulation control and quality of life: a prospective randomized study. Eur J Cardiothorac Surg 2009 Feb;35(2):265-269. [doi: 10.1016/j.ejcts.2008.10.020] [Medline: 19041254]
27. Schillig J, Kaatz S, Hudson M, Krol GD, Szandzik EG, Kalus JS. Clinical and safety impact of an inpatient pharmacist-directed anticoagulation service. J Hosp Med 2011;6(6):322-328. [doi: 10.1002/jhm.910] [Medline: 21834113]
28. Verret L, Couturier J, Rozon A, Saudrais-Janecek S, St-Onge A, Nguyen A, et al. Impact of a pharmacist-led warfarin self-management program on quality of life and anticoagulation control: a randomized trial. Pharmacotherapy 2012 Oct;32(10):871-879. [doi: 10.1002/j.1875-9114.2012.01116] [Medline: 23033226]
29. Bungard TJ, Ritchie B, Garg S, Tsuyuki RT. Sustained impact of anticoagulant control achieved in an anticoagulation management service after transfer of management to the primary care physician. Pharmacotherapy 2012 Feb;32(2):112-119. [doi: 10.1002/PHAR.1011] [Medline: 22392419]
30. Lakshmi R, James E, Kirthivasan R. Study on Impact of Clinical Pharmacist's Interventions in the Optimal Use of Oral Anticoagulants in Stroke Patients. Indian J Pharm Sci 2013 Jan;75(1):53-59 [FREE Full text] [doi: 10.4103/0250-474X.113550] [Medline: 23901161]
31. Sudas Na Ayutthaya N, Sakunrak I, Dhippayom T. Clinical Outcomes of Telemonitoring for Patients on Warfarin after Discharge from Hospital. Int J Telemed Appl 2018;2018:7503421 [FREE Full text] [doi: 10.1155/2018/7503421] [Medline: 30158972]
32. Liang JB, Lao CK, Tian L, Yang YY, Wu HM, Tong HHY, et al. Impact of a pharmacist-led education and follow-up service on anticoagulation control and safety outcomes at a tertiary hospital in China: a randomised controlled trial. Int J Pharm Pract 2020 Feb;28(1):97-106. [doi: 10.1111/ijpp.12584] [Medline: 31576625]
33. Tideman PA, Tirimacco R, St John A, Roberts GW. How to manage warfarin therapy. Aust Prescr 2015 Apr;38(2):44-48 [FREE Full text] [doi: 10.18773/austprescr.2015.016] [Medline: 26648615]
34. Giebel GD, Gissel C. Accuracy of mHealth Devices for Atrial Fibrillation Screening: Systematic Review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
35. Ferrando F, Mira Y. Effective and Safe Management of Oral Anticoagulation Therapy in Patients Who Use the Internet-Accessed Telecontrol Tool SintromacWeb. Interact J Med Res 2015 Apr 21;4(2):e10 [FREE Full text] [doi: 10.2196/ijmr.3610] [Medline: 25900775]
36. Agboola SO, Ju W, Elfiky A, Kvedar JC, Jethwani K. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. J Med Internet Res 2015 Mar 13;17(3):e65 [FREE Full text] [doi: 10.2196/jmir.4009] [Medline: 25793945]
37. Hou K, Yang H, Ye Z, Wang Y, Liu L, Cui X. Effectiveness of Pharmacist-led Anticoagulation Management on Clinical Outcomes: A Systematic Review and Meta-Analysis. J Pharm Pharm Sci 2017;20(1):378-396 [FREE Full text] [doi: 10.18433/J3SQ0B] [Medline: 29145935]
38. Manzoor BS, Cheng WH, Lee JC, Uppuluri EM, Nutescu EA. Quality of Pharmacist-Managed Anticoagulation Therapy in Long-Term Ambulatory Settings: A Systematic Review. Ann Pharmacother 2017 Dec;51(12):1122-1137. [doi: 10.1177/1060028017721241] [Medline: 28735551]
39. Wofford JL, Wells MD, Singh S. Best strategies for patient education about anticoagulation with warfarin: a systematic review. BMC Health Serv Res 2008 Feb 14;8:40 [FREE Full text] [doi: 10.1186/1472-6963-8-40] [Medline: 18275605]
40. Xu Z, Wang Z, Ou J, Xu Y, Yang S, Zhang X. Two monitoring methods of oral anticoagulant therapy in patients with mechanical heart valve prothesis: a meta-analysis. J Thromb Thrombolysis 2012 Jan;33(1):38-47. [doi: 10.1007/s11239-011-0626-1] [Medline: 21877235]
41. Zhou S, Sheng XY, Xiang Q, Wang ZN, Zhou Y, Cui YM. Comparing the effectiveness of pharmacist-managed warfarin anticoagulation with other models: a systematic review and meta-analysis. J Clin Pharm Ther 2016 Dec;41(6):602-611. [doi: 10.1111/jcpt.12438] [Medline: 27677651]
42. Lee M, Wang M, Liu J, Holbrook A. Do telehealth interventions improve oral anticoagulation management? A systematic review and meta-analysis. J Thromb Thrombolysis 2018 Apr;45(3):325-336. [doi: 10.1007/s11239-018-1609-2] [Medline: 29350322]

XSL•FO
RenderX

43. Niznik JD, He H, Kane-Gill SL. Impact of clinical pharmacist services delivered via telemedicine in the outpatient or ambulatory care setting: A systematic review. Res Social Adm Pharm 2018 Aug;14(8):707-717. [doi: 10.1016/j.sapharm.2017.10.011] [Medline: 29100941]
44. Xia X, Wu J, Zhang J. The effect of online versus hospital warfarin management on patient outcomes: a systematic review and meta-analysis. Int J Clin Pharm 2018 Dec;40(6):1420-1429. [doi: 10.1007/s11096-018-0731-8] [Medline: 30367374]

Abbreviations

eHealth: electronic health
INR: international normalized ratio
MD: mean difference
mHealth: mobile health
OAC: oral anticoagulant
RCT: randomized controlled trial
TTR: time in therapeutic range

Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by W Lizhu, D Triller; comments to author 11.03.20; revised version received 06.04.20; accepted 14.05.20; published 15.07.20.

Please cite as:
Dai H, Zheng C, Lin C, Zhang Y, Zhang H, Chen F, Liu Y, Xiao J, Chen C
Technology-Based Interventions in Oral Anticoagulation Management: Meta-Analysis of Randomized Controlled Trials
J Med Internet Res 2020;22(7):e18386
URL: https://www.jmir.org/2020/7/e18386
doi:10.2196/18386
PMID:32673227

©Hengfen Dai, Caiyun Zheng, Chun Lin, Yan Zhang, Hong Zhang, Fan Chen, Yunchun Liu, Jingwen Xiao, Chaoxin Chen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Self-Management Interventions for People With Osteoarthritis: Systematic Review With Meta-Analysis

Reza Safari[1], BSc, MSc, PhD; Jessica Jackson[1], BA, MSc, SCPHN, RMN; David Sheffield[2], BA, PhD

[1]Health and Social Care Research Centre, University of Derby, Derby, United Kingdom

[2]University of Derby Online Learning, Derby, United Kingdom

Corresponding Author:
Reza Safari, BSc, MSc, PhD
Health and Social Care Research Centre
University of Derby
Kedleston Road
Derby, DE22 1GB
United Kingdom
Phone: 44 1332 592418
Email: m.safari@derby.ac.uk

Abstract

Background: Osteoarthritis (OA) is not curable, but the symptoms can be managed through self-management programs (SMPs). Owing to the growing burden of OA on the health system and the need to ensure high-quality integrated services, delivering SMPs through digital technologies could be an economic and effective community-based approach.

Objective: This study aims to analyze the effectiveness of digital-based structured SMPs on patient outcomes in people with OA.

Methods: A total of 7 web-based and 3 gray literature databases were searched for randomized controlled trials assessing digital-based structured SMPs on self-reported outcomes including pain, physical function, disability, and health-related quality of life (QoL) in people with OA. Two reviewers independently screened the search results and reference lists of the identified papers and related reviews. Data on the intervention components and delivery and behavioral change techniques used were extracted. A meta-analysis, risk of bias sensitivity analysis, and subgroup analysis were performed where appropriate. The Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach was used to assess the quality of evidence.

Results: A total of 8 studies were included in this review involving 2687 patients with knee (n=2); knee, hip, or both (n=5); and unspecified joint (n=1) OA. SMPs were delivered via telephone plus audio and video, internet, or mobile apps. Studies reported that digital-based structured SMPs compared with the treatment as usual control group (n=7) resulted in a significant, homogeneous, medium reduction in pain and improvement in physical function (standardized mean difference [SMD] –0.28, 95% CI –0.38 to –0.18 and SMD –0.26, 95% CI –0.35 to –0.16, respectively) at posttreatment. The digital-based structured SMP effect on pain and function reduced slightly at the 12-month follow-up but remained to be medium and significant. The posttreatment effect of digital-based structured SMPs was small and significant for disability, but nonsignificant for QoL (SMD –0.10, 95% CI –0.17 to 0.03 and SMD –0.17, 95% CI –0.47 to 0.14, respectively; each reported in 1 study only). The 12-month follow-up effect of the intervention was very small for disability and QoL. The quality of evidence was rated as *moderate* for pain and physical function and *low* and *very low* for disability and QoL, respectively, using the GRADE approach.

Conclusions: Digital-based structured SMPs may result in improvement in pain and physical function that is largely sustained at the 12-month follow-up in people with knee and hip OA. The effects on disability and QoL are smaller and less clear. The quality of evidence is moderate to low, and further research is required to confirm the findings of the review and assess the effects of digital-based structured SMPs on other health-related outcomes.

(J Med Internet Res 2020;22(7):e15365) doi:10.2196/15365

KEYWORDS

osteoarthritis; self-management; internet-based intervention; mobile phone; eHealth; mHealth; systematic review; meta-analysis

XSL•FO
RenderX

Introduction

Osteoarthritis (OA) is a major burden for individuals, the health system, and the economy. The common symptoms of OA (ie, pain and subsequent physical inactivity) are risk factors in other chronic conditions, including cardiovascular problems, depression, stroke, cancers, and, consequently, premature mortality [1-6]. People with OA also experience fatigue, emotional distress, poor sleep, decreased productivity, social isolation, and poor quality of life (QoL) [7-9]. OA is not curable, but the pace of progression and symptoms can be managed. In a systematic review of guidelines for the management of OA [10], 12 of the 15 included guidelines made a strong recommendation for self-management and education. Evidence-based practice guidelines suggest that patients should be provided with tailored information to enhance understanding of the conditions and their progressive nature, tailored self-management programs (SMPs), information sharing, and regular contact with a multidisciplinary team to promote self-care and joint protection strategies [10-12].

SMPs are the structured and coordinated delivery of education and health behavior change interventions to empower people with OA to take care of their own condition [13,14]. Previous meta-analytic reviews show that SMPs and/or exercise interventions have small to medium benefits on health outcomes such as pain, function, and aspects of QoL [15-18]. A recent Cochrane review [19] revealed that, compared with usual care control groups, SMPs resulted in a significant but clinically unimportant reduction in pain up to 1 month postintervention (standardized mean difference [SMD] –0.26; 95% CI –0.41 to –0.10) and at 3- and 12-month follow-ups (SMD –0.17; 95% CI –0.26 to –0.08) [19]. No significant difference in patient-reported function was reported up to 1-month postintervention (SMD –0.01; 95% CI –019 to 018), but, at the 3- and 12-month follow-ups, function improved significantly in the SMP group (SMD –0.16; 95% CI –0.25 to –0.01). No effect of SMPs on QoL was reported in the review [19].

Due to the growing burden of arthritis on the health system, the increasing need to ensure high-quality integrated services; and the rapid advances in communication technology, health information, and services, delivering SMPs through digital technologies (eg, telephone, internet, mobile apps, and virtual reality equipment) could be an economical and effective community-based model of care [13,14]. Many people with OA continue to be in the workforce and have limited time to attend treatment sessions. In addition, people may live at a distance from physical health services [20,21]. Digital-based structured programs have the potential to be an important component in models of care for enhancing delivery to support self-management, offering a sustainable opportunity to improve patient outcomes, monitoring patients' symptoms for intervention adaptation, and increasing access to best practice [21]. It could also address the lack of continuity of care, lack of self-management support, and difficulty in accessing allied health professionals and pain management specialists [22]. A World Health Organization (WHO) guideline recommends the use of digital interventions in an interlinked manner, among others, to improve individuals' access to health services and information, and health workers can provide appropriate and high-quality care and can follow-up to ensure that individuals receive appropriate services [23].

There is a growing number of trials relating to digital-based structured interventions in OA conditions. A recent systematic review indicated that electronic health (eHealth; ie, internet, mobile, and telephone) exercise interventions, compared with no or other interventions, resulted in small effects in pain reduction, improved physical function, and improved health-related QoL [24]. Furthermore, moderate quality evidence indicates that telephone-based interventions (with educational material) reduce pain intensity and disability in people with OA of the knee or hip and spinal pain (back or neck pain) [25]. However, no review has focused explicitly on the effect of digital-based structured SMPs in people with OA. Therefore, this review aimed to determine the effectiveness of digital-based structured SMPs with controlled comparators on patient outcomes (pain, physical function, disability, and health-related QoL) in people with OA.

Methods

Systematic Literature Review

Guidance published in Preferred Reporting Items for Systematic reviews and Meta-Analyses [26] and the Cochrane Handbook of Systematic Reviews [27] was adhered to. The a priori protocol for the review is published in the International Prospective Register of Systematic Reviews (PROSPERO): CRD42018089322.

Search Strategy

Studies were identified by searching web-based databases with support and consultation provided by an institutional librarian. The search included the Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica Database (EMBASE), Cochrane Central Register of Controlled Trials (CENTRAL), Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Physiotherapy Evidence Database (PEDro), and PubMed and gray literature databases (Dissertation Abstracts International WorldCat, The Grey Literature Report, and Open Grey) using both Medical Subject Headings and free-text keywords relating to OA, digital-based structured self-management interventions, and outcomes stated below from inception to May 2018. Search strategies for the first 3 databases above and dates on which searches were conducted are listed in Multimedia Appendix 1. Searches of the reference list of the previous review papers and included studies were also conducted.

Eligibility Criteria

Type of Studies

Randomized controlled trials (RCTs) of any design, including parallel-group, crossover, and cluster RCTs published in the English language were included in this review.

Participants

Adults (≥18 years of age) with a confirmed diagnosis of OA, either radiologically or by a health practitioner. All types of OA

XSL•FO
RenderX

at any stage of the disease were considered. We included studies recruiting patients with OA with other conditions only if outcome data for OA patients were provided.

Intervention

The intervention considered structured and coordinated SMPs as defined by Lorig [14] and Osborne [13] in isolation or in combination with other interventions delivered fully or partially via digital technologies (eg, websites, mobile apps, social networking tools, web-based games, animation, and telephone). Self-management is defined as an engagement in activities that promote health and prevent adverse events; interacting with a health care professional; improving self-monitoring; coping with disease; and developing skills in problem-solving, decision making, resource utilization, forming of a patient and health care provider partnership, and taking action.

Control Condition

Any type of control group (ie, waitlist, treatment as usual or minimal interventions, alternative treatment, or other digital-based interventions) was considered.

Outcome

We included any psychometrically sound unidimensional or multidimensional measure as well as the relevant subscales relating to the following outcomes:

Primary Outcomes

1. Pain: Visual Analog Scale, Numerical Rating Scale, and Brief Pain Inventory.
2. Function: patient-specific physical function, physical function subscale of Arthritis Impact Measurement Scale, or Short Form-36.
3. Disability: Oswestry Disability Questionnaire and Roland-Morris Disability Questionnaire.

Secondary Outcomes

1. QoL: European Quality of Life-5 Dimensions, Short Form-36, or WHO Quality of Life–Brief scale.
2. Cost and resource use: We extracted the results of economics reported alongside the effectiveness studies, either full or partial economic evaluation or estimates of resource use and costs associated with interventions and comparators. Change from baseline data relating to any follow-up time points were considered.

Study Selection and Data Extraction

Two reviewers (RS and JJ) independently screened abstracts and full text of the search results. Any disagreements were resolved through discussion. Data were extracted by 2 researchers (JJ and EH; mentioned in Acknowledgments) and cross-checked by a third researcher (RS) using a data extraction tool developed a priori based on the Cochrane Handbook recommendations. Data items included participants' characteristics, guiding theory or rationale, mechanism of effect, the digital medium, method of delivery, who delivered, where it was delivered, adherence, and fidelity. Additional data relating to describing components of the interventions were extracted based on the Template for Intervention Description and Replication (TiDieR) guidance [28] and SMP components [13,14]. The behavior change techniques used were also extracted based on the hierarchical taxonomy by Michie et al [29] to examine if certain techniques were favored or, if there were sufficient data, whether particular techniques yielded bigger effects. The data extraction form was piloted on 2 included papers before full data extraction.

Meta-Analysis

Outcome data were expressed as SMD and were pooled in a pair-wise fixed effects model stratified based on the outcome and type of digital medium. We used the mean (SD) of within-group change from baseline to calculate the SMDs. Per Cohen [30], SMDs <0.2 were classified as small, those between 0.2 and 0.8 were classified as medium, and those >0.8 were classified as large. We converted the SMD to the percentage change in the outcome measure by multiplying the SMD with the SD of the control group of the sufficiently powered study (ie, the trial with the largest sample size).

A sensitivity analysis was conducted for the risk of bias assessment. A subgroup analysis was performed to assess the effect of SMPs on the different digital media used, control conditions, and presence of active exercise component in the intervention. Heterogeneity across pooled studies was examined using the chi-squared test and I^2 statistics. As there were <10 studies, we did not use a funnel plot or Egger test to assess publication bias. Furthermore, a narrative synthesis of the studies was conducted where there were insufficient data to pool studies in a meta-analysis. We did not conduct additional exploratory analyses to explore potential moderators because of an insufficient number of studies. The common study characteristics tables were supplemented by additional summary tables including the TiDieR process information; risk of bias assessment; effect estimates; Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) analysis; and behavior change taxonomy groupings. We used Review Manager (RevMan, version 5.3; Cochrane Collaboration) for all analyses.

Dealing With Missing Data

We contacted the authors of included studies for missing data. Aggregated data were provided by authors for 6 studies. We used the RevMan calculator to impute missing SDs from the test statistics reported for 2 studies.

Risk of Bias and Grading of Recommendations, Assessment, Development, and Evaluations Assessment

Risk of bias was assessed using the risk of bias tool of the Cochrane Handbook for Systematic Reviews [27]. Quality of evidence for outcomes was assessed according to the 5 GRADE domains, including study limitation (risk of bias), inconsistency, indirectness, imprecision, and publication bias [31,32].

Results

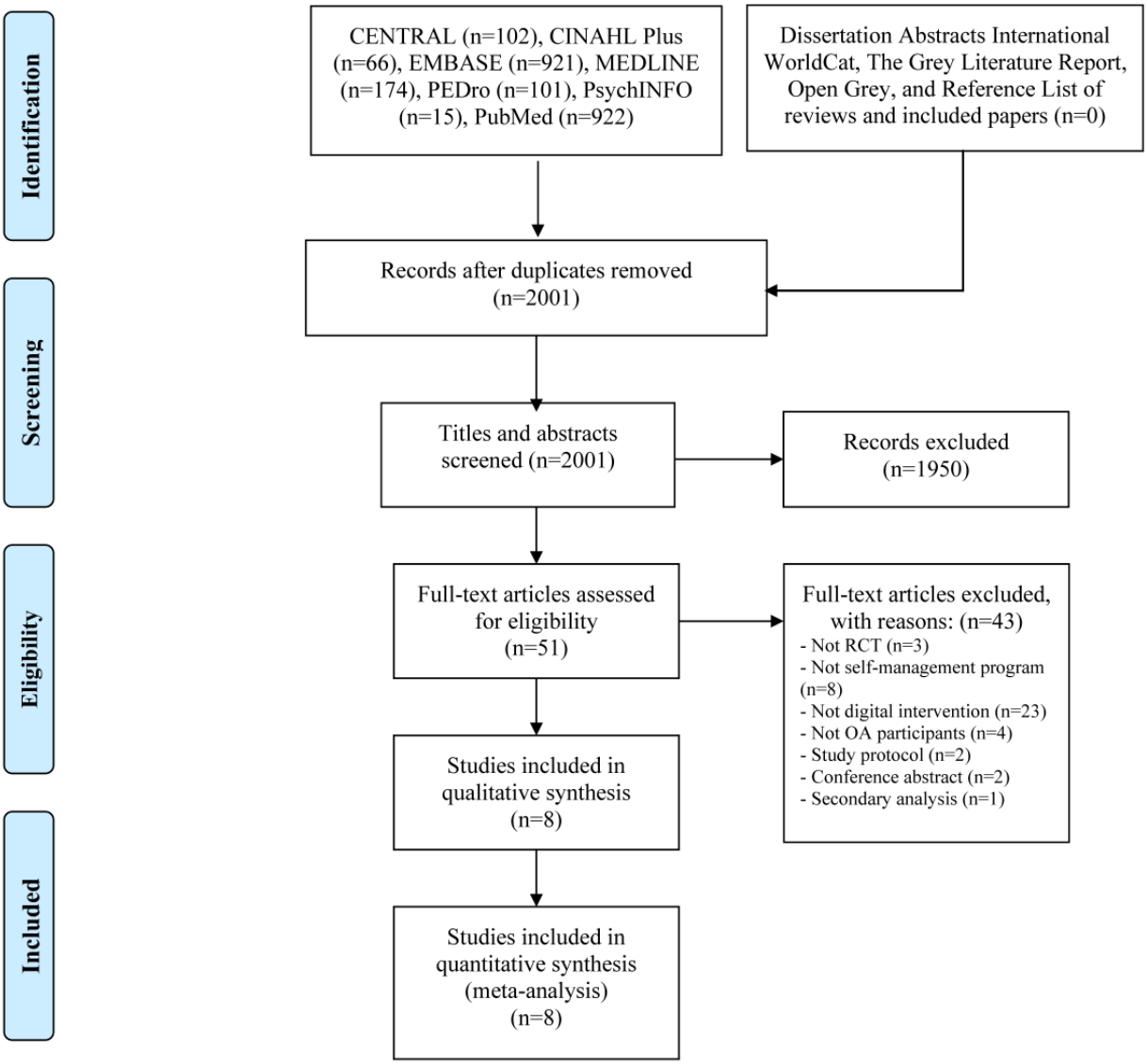
Study Selection and Characteristics

A total of 2001 titles and abstracts were screened after excluding duplicates, of which 1950 records did not meet the inclusion criteria (Figure 1). Full texts of 51 potential eligible records were read, and 8 studies published between 2008 and 2018 were

XSL•FO
RenderX

included. These studies included 2256 (range 113-352) individuals with OA. A total of 8 digital-based structured SMPs were compared with 10 control conditions. Two studies were cluster RCTs and 6 were RCTs. A list of studies excluded at the full text screening stage with reasons for exclusion is presented in Multimedia Appendix 2 [33-75]. A summary of the study characteristics and participants' demographics are presented in Tables 1 and 2.

Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses flow diagram. CENTRAL: Cochrane Central Register of Controlled Trials; CINAHL Plus: Cumulative Index to Nursing and Allied Health Literature Plus; EMBASE: Excerpta Medica Database; MEDLINE: Medical Literature Analysis and Retrieval System Online; PEDro Physiotherapy Evidence Database; OA: osteoarthritis; RCT: randomized controlled trial.



XSL•FO
RenderX

Table 1. Characteristics of the included studies.

Author (year)	Country	Participants, n (%)		Intervention type		Outcomes (primary or secondary; baseline mean [SD])	Outcomes (primary or secondary; 95% CI)	Postintervention (weeks); analysis (ITT[a] or PP[b])	Follow-up (weeks); analysis (ITT or PP)	Attrition at postintervention, n (%)	
		Experiment	Control	Experiment	Control					Experiment	Control
Allen et al (2010) [76]	United States	172 (33.3)	171 (33.1)	Self-management	TAU[c]	AIMS2[d]: pain (P; 5.9 [2.3]); AIMS2: function (S; 2.6 [1.7])	N/A[e]	52; ITT	NR[f]	26 (15.1)	14 (8.1)
Allen et al (2010) [76]	United States	172 (33.3)	172 (33.3)	Self-management	Health education (attention control)	AIMS2: pain (P; 6.0 [2.3]); AIMS2: function (S; 2.7 [1.8])	N/A	52; ITT	NR	26 (15.1)	14 (8.1)
Allen et al (2016) [77]	United States	151 (50.3)	149 (49.6)	Combined patient and provider OA[g] management program	TAU	WOMAC[h]: pain (S; 8.5, NR) WOMAC: function (S; 28.7 [NR])	N/A	52; ITT	NR	15 (9.9)	12 (8.1)
Allen et al (2018) [78]	United States	142 (40.5)	68 (19.4)	Internet-based exercise training	WL[i]	WOMAC: pain (S; 6.0 [3.9]); WOMAC: function (S; 21.8 [2.7])	N/A	16; ITT	52; ITT	30 (21.1)	5 (7)
Allen et al (2018) [78]	United States	142 (40.5)	140 (40.0)	Internet-based exercise training	In-person physical therapy	WOMAC: pain (S; 6.1, 3.5); WOMAC: function (S; 22.6 [12.9])	N/A	16; ITT	52; ITT	30 (21.1)	11 (7.8)
Bossen (2013) [79]	The Netherlands	100 (50.2)	99 (49.8)	Join2move: automated web based	WL	N/A	KOOS[j] and HOOS[k]: function (P; 58.8 [95% CI 51.5-66.0]); KOOS and HOOS: pain (S; 5.4 [95% CI 4.2-6.5]) KOOS and HOOS: QoL (S; 38 [95% CI 30.6-45.5])	12; ITT	52; ITT	16 (16)	15 (15.1)

XSL•FO
RenderX

Author (year)	Country	Participants, n (%)		Intervention type		Outcomes (primary or secondary; baseline mean [SD])	Outcomes (primary or secondary; 95% CI)	Postintervention (weeks); analysis (ITT[a] or PP[b])	Follow-up (weeks); analysis (ITT or PP)	Attrition at postintervention, n (%)	
		Experiment	Control	Experiment	Control					Experiment	Control
Kloek (2018) [80]	The Netherlands	109 (50.2)	99 (49.8)	Internet-based exercise	Usual physical therapy	N/A	KOOS and HOOS: function (P; 50.7 [95% CI 45.1-56.4]); KOOS and HOOS: pain (S; 43.9 [95% CI 35.2-52.7]); KOOS and HOOS: QoL[l] (S; 44.2 [95% CI 38.1-50.4])	12; PP	52; PP	20 (18.3)	(12) 12
Skrepnik (2017) [81]	United States	107 (50.7)	104 (49.3)	Mobile app (OA GO)	TAU	Numeric Pain Rating Scale: (NR; 4.6 [2.3]); 6-min walk test (NR; 402.8 [120.5])	N/A	12; ITT	NR	6 (5.06)	2 (1.9)
Lorig (2008) [82]	United States	433 (50.6)	422 (49.4)	Web-based arthritis self-management program	TAU	Activity Limitation Scale (S; 2.20 [1.03]); Health Assessment Questionnaire; Disability (S; 0.552 [0.402]); Numeric Pain Rating Scale (S; 6.53 [2.23])	N/A	24; ITT and PP	52; ITT and PP	NR	NR

XSL•FO
RenderX

Author (year)	Country	Participants, n (%)		Intervention type		Outcomes (primary or secondary; baseline mean [SD])	Outcomes (primary or secondary; 95% CI)	Postintervention (weeks); analysis (ITT[a] or PP[b])	Follow-up (weeks); analysis (ITT or PP)	Attrition at postintervention, n (%)	
		Experiment	Control	Experiment	Control					Experiment	Control
Rini (2015) [83]	United States	58 (51.3)	55 (48.6)	Pain COACH group	No intervention	AIMS2: pain (P; 4.82 [1.73]); AIMS2: function (S; 1.70 [1.30])	N/A	9-12; ITT	NR	1 (1.7)	3 (5.4)

[a]ITT: intention-to-treat.

[b]PP: per protocol.

[c]TAU: treatment as usual.

[d]AIMS: Arthritis Impact Measurement Scales.

[e]N/A: not applicable.

[f]NR: not reported.

[g]OA: osteoarthritis.

[h]WOMAC: The Western Ontario and McMaster Universities Osteoarthritis Index.

[i]WL: waiting list.

[j]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[k]HOOS: Hip disability and Osteoarthritis Outcome Score.

[l]QoL: quality of life.

XSL•FO
RenderX

Table 2. Study participants' demographics.

Author (year), Study arms	Age (years), mean (SD)	Female, n (%)	Affected joint, n (%)			Time since diagnosis (years), mean (SD)
			Knee	Hip	Both	
Allen et al (2010) [76]						
Self-management	60.3 (10.03)	NR[a] (9)	NR (82)	NR (12)	NR (6)	16.5 (12.7)
TAU[b]	59.7 (10.1)	NR (6)	NR (79)	NR (17)	NR (4)	15.9 (11.9)
Health education (attention control)	60.3 (10.8)	NR (7)	NR (79)	NR (16)	NR (5)	15.8 (12.0)
Allen et al (2016) [77]						
Combined patient and provider OA[c] management program	61.7 (9.0)	20 (13.2)	114 (75.5)	18 (11.9)	19 (12.6)	13.8 (11.1)
TAU	60.4 (9.4)	8 (5.4)	124 (83.2)	14 (9.4)	11 (7.4)	14.6 (12.1)
Allen et al (2018) [78]						
Internet-based exercise training	65.3 (11.5)	98 (69)	142 (100)	N/A[d]	N/A	11.6 (11)
Waiting list	64.3 (12.2)	53 (78)	68 (100)	N/A	N/A	14.2 (13)
In-person physical therapy	65.7 (10.3)	100 (71.4)	140 (100)	N/A	N/A	14.1 (11.6)
Bossen et al (2013) [79]						
Join2move: automated web based	61 (5.9)	60 (60)	67 (67.0)	21 (21.0)	12 (12.0)	NR
Waiting list	63 (5.4)	69 (70)	60 (60.6)	20 (20.2)	19 (19.2)	NR
Kloek et al (2018) [80]						
Internet-based exercise	63.8 (8.5)	74 (67.9)	71 (65.1)	21 (19.3)	17 (15.6)	NR
Usual physical therapy	63.3 (8.9)	67 (67.7)	67 (67)	17 (17)	15 (15)	NR
Skrepnik et al (2017) [81]						
Mobile app (OA GO)	61.6 (9.5)	59 (55.1)	100	N/A	N/A	NR
TAU	63.6 (9.3)	47 (45.2)	100	N/A	N/A	NR
Lorig et al (2008) [82]						
Web-based arthritis self-management program	52.5 (12.2)	NR (90.5)	NR	NR	NR	NR
TAU	52.2 (10.9)	NR (89.8)	NR	NR	NR	NR
Rini et al (2015) [83]						
Pain COACH group	68.52 (7.65)	46 (79)	18 (33)	9 (16)	28 (51)	NR
No intervention	66.67 (11.02)	45 (82)	22 (38)	5 (9)	31 (53)	NR

[a]NR: not reported.

[b]TAU: treatment as usual.

[c]OA: osteoarthritis.

[d]N/A: not applicable.

Participants

A total of 6 studies were conducted in the United States, and 2 were performed in the Netherlands.

The number of participants in the studies ranged from 113 to 855. The median (IQR) age was 63 (3.55) years. Most of the participants in the 6 studies were female (median 74.7%, IQR 6.9%), whereas in the remaining 2 studies, most participants were men (93.0% and 90.7%).

Two studies recorded participants with knee OA only; 5 studies recorded a mixture of knee (median 71.55%, IQR 15.77%), hip (median 16%, IQR 5.75%), or both (median 6%, IQR 10.5%); and 1 study did not record the area affected. No studies recorded OA in the hand or any other type of OA. Five of the studies

XSL•FO
RenderX

recorded the time since diagnosis, with a median of 15 (IQR 2) years.

Intervention Groups

Intervention Delivery Mechanisms

Out of the studies reviewed, 2 used telephone, audio and video, and written materials to deliver an SMP derived from social cognitive theory and 5 studies delivered an internet-based exercise training: 1 study was delivered by a physical therapist, 2 studies were delivered in a combination of face-to-face time with a physical therapist, 1 study had a virtual coach, and 1 study was self-administered only. The remaining study used a mobile app with a wearable monitor in combination with a physician face-to-face. Three studies delivered interventions over 52 weeks, 2 studies had 12-week interventions, and 1 study each had interventions lasting 6, 8, and 9 weeks (Table 3).

XSL•FO
RenderX

Table 3. Detailed intervention delivery mechanisms.

Author (year)	Medium or method (Y or N)[a]							Professional input or support	Timing		Tailoring (Y or N)	Modification (Y or N)	Adherence assessed: Y or N; Completion: n (%)
	Telephone	Audio and/or video	Internet	Mobile app	Wearables	Written material and/or booklet	Face-to-face element		SMP[b] delivery period (weeks)	Number of support sessions			
Allen et al (2010) [76]	Y	Y	N	N	N	Y	N	Telephone calls by health educator	52	Once a month for 12 months	Y	N	Y; NR[c]
Allen et al (2016)[d] [77]	Y	Y	N	N	N	Y	N	Telephone calls by counselor trained in OA[e] and behavior change	52	Twice a week for 6 weeks plus once a week for 6 weeks	Y	N	N; NR
Allen et al (2018) [78]	N	N	Y	N	N	N	N	Physical therapist administered the intervention	52	Up to 8	Y	N	Y; 114 (80.2)
Bossen (2013) [79]	N	N	Y	N	N	N	N	None	9	N/A[f]	Y	Y	Y; 46 (46)
Kloek et al (2018) [80]	N	N	Y	N	N	N	Y	Face-to-face with physical therapist	12	5 over 12 weeks	N	Y	Y; NR (81)
Skrepnik et al (2017) [81]	N	N	N	Y	Y	N	Y	Face-to-face with physician investigator plus trial coordinator demonstrated the app	12	5 over 12 weeks	Y	N	Y; 90 (82.5)
Lorig et al (2008) [82]	N	N	Y	N	N	Y	N	SMP-trained moderator facilitating the program	6 weeks	NR	Y	N	Y; approximately 95%[g]
Rini et al (2015) [83]	N	N	Y	N	N	Y	N	Virtual coach led participants through the program	8	Once a week for 8 weeks	N	N	Y; 53 (91)

[a]Y: yes and N: no.

[b]SMP: self-management program.

[c]NR: not reported.

XSL•FO
RenderX

[d]The intervention also included P*rovider Intervention*, which involved delivery of patient-specific recommendations at the point of care.

[e]OA: osteoarthritis.

[f]N/A: not applicable.

[g]Logged at least once into the program.

Self-Management Components

All studies used health education as a component of self-management. For 5 studies, this was the main component. Additional self-management components included goal setting (n=6); action planning (n=4); and exercise components such as physical activity (n=6), aerobic (n=5), resistive (n=4), flexibility (n=3), and balance (n=1) were recorded.

Additional self-management components included diet and/or weight management (n=5); pain management (n=6); medication (n=3); motivation (n=6); peer support (n=2); patient-therapist communication (n=2); and stress management, relaxation, or sleep (n=4; Tables 4 and 5).

A range of theories were used to inform the intervention, including Social Cognition Theory (n=3), Self-Efficacy Theory (n=2), and behavioral graded activity theory (n=2). One study stated that they used a combination of Social Cognition Theory, adult learning theory, and principles of multimedia instruction (Tables 4 and 5).

Table 4. Detailed self-management program components.

Author (year)	Education	Goal setting	Action planning	Exercise[a] components or PA					Exercise dose; (weeks×frequency×minute)
				Aerobic	Resistive	Flexibility	Balance	Physical activity	
Allen et al (2010) [76]	✓[b]	✓[b]	✓[b]	NR[c]	NR	NR	NR	✓[d]	NR
Allen et al (2016)[e] [77]	✓[d]	✓[b]	✓[b]	✓[d]	✓[d]	✓[d]	—	✓[d]	52×2×75
Allen et al (2018) [78]	✓[d]	—	✓[d]	✓[b]	✓[b]	✓[b]	—	—	Aerobic: 52×7×NR; resistive and flexibility: 52×3×NR
Bossen et al (2013) [79]	✓[b]	✓[b]	—	✓[d]	—	—	—	✓[b]	Varied with gradual increments
Kloek et al (2018) [80]	✓[b]	—	—	✓[d]	✓[b]	—	✓[b]	✓[b]	12×3×NR (with gradual increments)
Skrepnik et al (2017) [81]	✓[d]	✓[b]	—	N/A[f]	N/A	N/A	N/A	✓[b]	12×7×NR
Lorig et al (2008) [82]	✓[b]	✓[d]	✓[b]	✓[d]	✓[d]	✓[d]	—	—	Varied: tailored to individual
Rini et al (2015) [83]	✓[b]	✓[d]	—	N/A	N/A	N/A	N/A	✓[d]	NR

[a]All studies had exercise as part of the intervention, except Skrepnik et al [81] and Rini et al [83].

[b]Main components of the intervention.

[c]NR: not reported.

[d]Other components of the intervention.

[e]The intervention also included *Provider Intervention*, which involved delivery of patient-specific recommendations at the point of care.

[f]N/A: not applicable.

XSL•FO
RenderX

Table 5. Detailed self-management program components (continued).

Author (Year)	Diet or weight management	Pain management	Medication	Motivation	Peer support	Patient-therapist communication	Stress management, relaxation and/or sleep	Theory
Allen et al (2010) [76]	✓[a]	✓[a]	✓[a]	—[b]	—	✓[a]	✓[a]	SCT[c]
Allen et al (2016)[d] [77]	✓[a]	✓[a]	—	✓[a]	—	—	✓[a]	SCT
Allen et al (2018) [78]	✓[a]	✓[a]	—		—	—	—	SET[e]
Bossen et al (2013) [79]	—	—	—	✓[a]	—	—	—	BGAT[f]
Kloek et al (2018) [80]	✓[a]	✓[a]	✓[a]	✓[a]	—	—	—	BGAT
Skrepnik et al (2017) [81]	—	—	—	✓[a]	—	—	—	NR[g]
Lorig et al (2008) [82]	✓[a]	✓[h]	✓[a]	✓[a]	✓[h]	✓[a]	✓[a]	SET
Rini et al (2015) [83]	—	✓[h]	N/A[i]	✓[a]	✓[a]	—	✓[h]	SCT, ALT[j], PMI[k]

[a]Other components of the intervention.

[b]Components not included in the intervention.

[c]SCT: social cognitive theory.

[d]The intervention also included *Provider Intervention*, which involved delivery of patient-specific recommendations at the point of care.

[e]SET: self-efficacy theory.

[f]BGAT: Behavior Graded Activity Theory.

[g]NR: not reported.

[h]Main components of the intervention.

[i]N/A: not applicable.

[j]ALT: adult learning theory.

[k]PMI: principles of multimedia instruction.

Behavior Change Components

A variety of behavior change techniques have been used. The most common techniques involved goal setting and planning (n=8), feedback and monitoring (n=7), and shaping knowledge (n=7). All studies used at least four different groups of techniques, with Rini et al [83] using 10 group techniques [29] (Table 6; Multimedia Appendix 3). Owing to the heterogeneity in the approaches used, a meta-analysis was not appropriate.

XSL•FO
RenderX

Table 6. Behavior change techniques used within included studies.

Author (year)	BCT[a] taxonomy grouping[b]															
	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	10[l]	11[m]	12[n]	13[o]	14[p]	15[q]	16[r]
Allen et al (2010) [76]	1.1[s], 1.2[s], 1.4[s]	—[t]	3.1[s]	4.1[s], 4.2[s]	5.1[s]	—	—	—	—	—	—	—	—	—	—	—
Allen et al (2016) [77]	1.1, 1.4, 1.4, 1.5 [s]	2.3[s]	3.1, 3.2[s], 3.3[s]	4.1	5.1[s]	6.1[s]	—	8.1[s]	—	—	—	—	—	—	—	16.2[s]
Allen et al (2018) [78]	1.7[s]	2.3, 2.4[s]	3.2	4.1, 4.2	5.1	6.1	—	8.7[s]	—	—	—	—	—	—	—	—
Bossen et al (2013) [79]	1.1	2.4	3.1	4.1	—	6.1	—	8.7	—	—	—	—	—	—	15.1 s	—
Kloek et al (2018) [80]	1.1	2.3, 2.6[s]	3.1, 3.2	4.1, 4.2	—	6.1	—	8.7	—	—	—	—	—	—	—	—
Skrepnik et al (2017) [81]	1.1, 1.5	2.2[s], 2.4,2.6	—	—	5.1, 5.4[s]	—	7.1[s]	—	—	—	—	—	—	—	—	—
Lorig et al (2008) [82]	1.2, 1.4	2.2, 2.3	3.2	4.1	—	—	—	—	—	—	11.2[s]	12.6[s]	—	—	15.4[s]	16.3[s]
Rini et al (2015) [83]	1.1, 1.2, 1.4, 1.5	2.2, 2.4	—	4.1	5.4 5.5[s]	6.2[s], 6.3[s]	—	8.1, 8.3[s]	9.2[s]	—	11.2	12.4[s]	—	—	15.1	—

[a]BCT: behavior change technique.

[b]Multimedia Appendix 3 provides an explanation of the grouping and example text from study papers.

[c]1: Goals and planning

[d]2: Feedback and monitoring

[e]3: Social support

[f]4: Shaping knowledge

[g]5: Natural consequences

[h]6: Comparison of behavior

[i]7: Associations

[j]8: Repitition and substitution

[k]9: Comparison of outcomes

[l]10: Rewards and threats

[m]11: Regulation

[n]12: Antecedents

[o]13: Identity

[p]14: Scheduled consequences

[q]15: Self-belief

[r]16: Covert learning

[s]1.1: goal setting (behavior); 1.2: problem-solving; 1.4: action planning; 1.5: review behavior goal(s); 1.7: review outcome goal(s); 2.2: feedback on behavior; 2.3: self-monitoring of behavior; 2.4: self-monitoring of outcome(s) of behavior; 2.6: biofeedback; 3.1: social support (unspecified); 3.2: social support (practical); 3.3: social support (emotional); 4.1: instruction on how to perform the behavior; 4.2: information about antecedents; 5.1: information about health consequences; 5.4: monitoring of emotional consequences; 5.5: anticipated regret; 6.1: demonstration of the behavior; 6.2: social comparison; 6.3: information about others' approval; 7.1: prompts or cues; 11.2: reduce negative emotions; 12.4: distraction; 12.6: body changes; 15.1: verbal persuasion about capability; 15.4: self-talk; 16.2: imaginary reward; 16.3: vicarious consequences.

[t]Behavior change techniques not used in the study.

XSL•FO
RenderX

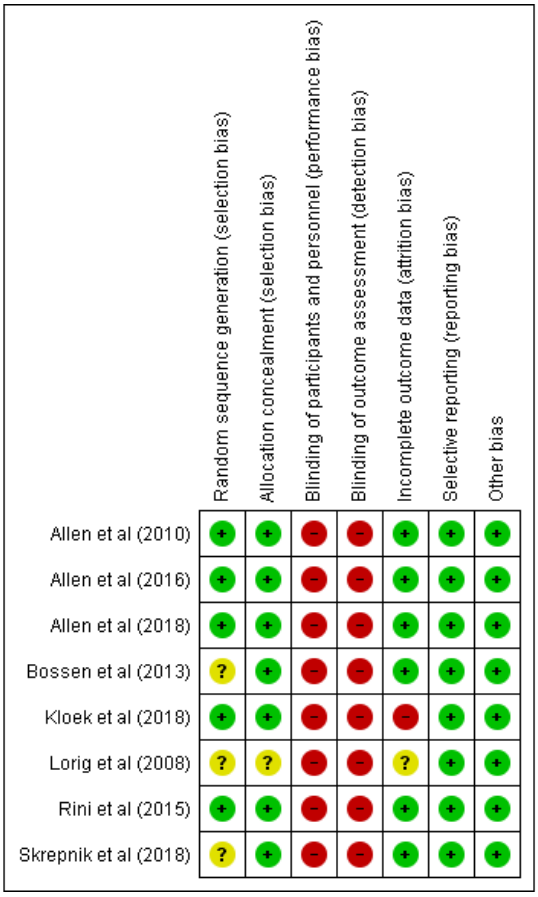
Comparison

Digital self-management interventions were compared with treatment as usual (n=4), wait list control (n=2), in-person physical therapy (n=2), attention group (health education; n=1), and no intervention (n=1).

Risk of Bias

The risk of bias was assessed as unclear for random sequence generation (n=3), allocation concealment (n=1), and incomplete outcome data (n=1). There was a high risk of bias for incomplete outcome data in 1 study (Figure 2). The detection bias and performance bias were rated high in all studies because of the nature of the interventions and patient-reported outcome tools.

Figure 2. Summary of risk of bias assessment.



Effects of Interventions

A summary of findings for the main comparisons of the study with GRADE ratings is presented in Table 7. Detailed meta-analytic forest plots are also presented in Multimedia Appendix 4.

XSL•FO
RenderX

Table 7. Summary of findings and Grading of Recommendations, Assessment, Development, and Evaluations ratings for the main comparisons.

Outcomes	Number of participants (studies)	Quality of the evidence[a] (GRADE[b])	Comments	Illustrative comparative risks (95% CI)	
				Assumed risk control	Corresponding risk intervention (95% CI)
Pain (D-SMP[c] vs TAU[d] control): AIMS2[e], WOMAC[f], HOOS[g]/KOOS[h], NPRS[i]; follow-up: 9-52 weeks	1614 (7 studies)	⊕⊕⊕ Moderate[j]	Limitation (–1)	No risk assumed	The mean pain in the intervention groups was 0.28 SDs lower (0.38-0.18 lower)
Pain (D-SMP vs TAU control): WOMAC, HOOS/KOOS, NRS PAIN; follow-up: ≥52 weeks	716 (3 studies)	⊕⊕⊕ Moderate[j]	Limitation (–1)	No risk assumed	The mean pain in the intervention groups was 0.20 SDs lower (0.35-0.05 lower)
Pain (internet-SMP vs physical therapy): WOMAC, HOOS/KOOS; follow-up: 12-52 weeks	456 (2 studies)	⊕⊕ Low[j,k]	Limitation (–1); imprecision (–1)	No risk assumed	The mean pain in the intervention groups was 0.07 SDs lower (0.25 lower to 0.01 higher)
Pain (internet-SMP vs physical therapy): WOMAC, HOOS/KOOS; follow-up: ≥52 weeks	416 (2 studies)	⊕⊕ Low[j,k]	Limitation (–1); imprecision (-1)	No risk assumed	The mean pain in the intervention groups was 0.12 SDs lower (0.31 lower to 0.07 higher)
Pain (telephone- and video-SMP vs attention control); AIMSS2	306 (1 study)	⊕⊕ Low[l]	Unknown consistency (–2)	No risk assumed	The mean pain in the intervention groups was 0.26 SDs lower (0.49 lower to 0.04 lower)
Physical function (D-SMP vs TAU: AIMS2, WOMAC, HOOS/KOOS, 6MWT[m], ALS[n]; follow-up: 9-52 weeks	1625 (7 studies)	⊕⊕⊕ Moderate[j]	Limitation (-1)	No risk assumed	The mean physical function in the intervention groups was 0.26 SDs higher (0.35-0.16 higher)
Physical function (D-SMP vs TAU control): WOMAC, KOOS/HOOS, ALS; follow-up: ≥52 weeks	707 (3 studies)	⊕⊕⊕ Moderate[j]	Limitation (–1)	No risk assumed	The mean pain in the intervention groups was 0.23 SDs higher (0.38-0.08 higher)
Physical function (internet-SMP vs active control): WOMAC, HOOS/KOOS follow-up: 12-52 weeks	258 (2 studies)	⊕⊕ Low[j,k]	Limitation (–1); imprecision (–1)	No risk assumed	The mean physical function in the intervention groups was 0.05 SDs lower (0.13 higher to 0.23 lower)
Physical function (internet-SMP vs active control): WOMAC, HOOS/KOOS; follow-up: ≥52 weeks	416 (2 studies)	⊕⊕ Low[j,k]	Limitation (–1); imprecision (–1)	No risk assumed	The mean physical function in the intervention groups was 0.03 SDs higher (0.22 higher to 0.16 lower)
Physical function (telephone- and video-SMP vs attention control): AIMS2	306 (1 study)	⊕ Very low[k,l]	Imprecision (–1); unknown consistency (–2)	No risk assumed	The mean physical function in the intervention groups was 0.17 SDs higher (0.39 higher to 0.06 lower)
Disability (internet-SMP vs TAU): HAQ[o]	352 (1 study)	⊕ Very low[j,l]	Limitation (–1); unknown consistency (–2)	No risk assumed	The mean disability in the intervention groups was 0.10 SDs lower (0.17-0.03 lower)
Quality of life (internet-SMP vs TAU): HOOS/KOOS	165 (1 study)	⊕ Very low[j,k,l]	Imprecision (–1); unknown consistency (–2)	No risk assumed	The mean quality of life in the intervention groups was 0.17 SDs higher (0.47 higher to 0.14 lower)

[a]GRADE Working Group grades of evidence: High quality: further research is very unlikely to change our confidence in the estimate of effect; Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; Very low quality: we are very uncertain about the estimate.

[b]GRADE: Grading of Recommendations, Assessment, Development, and Evaluations.

[c]D-SMP: digital-based structured self-management program.

[d]TAU: treatment as usual.

[e]AIMS: Arthritis Impact Measurement Scales.

XSL•FO
RenderX

[f]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[g]HOOS: Hip disability and Osteoarthritis Outcome Score.

[h]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[i]NPRS: numerical pain rating scale.

[j]Majority of the evidence comes from studies with unclear randomization and/or allocation concealment.

[k]Total sample size is small. Total effect size has CIs crossing the no effect line.

[l]Unknown consistency and/or publication bias.

[m]6MWT: 6-minute walk test.

[n]ALS: Activity Limitation Scale.

[o]HAQ: Health Assessment Questionnaire.

Main Outcomes

Pain

Seven studies involving 1614 participants reported the effect of digital-based structured SMPs compared with usual care control condition on pain outcome at the postintervention time point (range 9-52 weeks). Overall, a significant and medium effect in favor of the intervention was observed on pain reduction (SMD –0.28; 95% CI –0.38 to –0.18); heterogeneity (X^2_6=5.1; *P*=.53; I^2=0%). Three studies with 716 participants reported a long-term (≥12 months) effect of digital-based structured self-management program compared with usual care conditions, indicating a significant and small overall effect (SMD –0.20; 95% CI –0.35 to –0.05); heterogeneity (X^2_2=0.4; *P*=.83; I^2=0%). Using the GRADE approach, the quality of evidence was rated *moderate* for both postintervention and long-term follow-up comparisons because of the high risk of bias in most studies (ie, study limitations).

Two studies with 456 participants comparing digital-based structured SMPs (internet) and physical therapy did not show a difference in pain reduction between the 2 intervention conditions (SMD –0.07; 95% CI –0.25 to 0.11); heterogeneity (X^2_1=0.2; *P*=.64; I^2=0%). The effect of the digital-based structured SMP slightly increased at the 12-month follow-up time point but was not significant (SMD –0.12; 95% CI – 0.31 to 0.07; heterogeneity X^2_1=0.1; *P*=.94; I^2=0%). The quality of evidence for both time points was rated as *low* because of the high risk of bias (ie, study limitation) and total small sample size (ie, imprecision).

The only study comparing digital-based structured SMPs with attention control (health education) condition [76] (n=306 participants) showed a medium and significant effect on pain reduction in favor of digital-based structured SMPs (SMD –0.26; 95% CI –0.49 to –0.04). The quality of evidence was rated as *low* because of unknown consistency and/or publication bias.

Physical Function

A total of 7 studies involving 1625 people with OA reported the effect of digital-based structured SMPs compared with usual care control conditions on patient-reported physical function outcomes at the postintervention time point (range 9-52 weeks). Overall effect size indicates a medium and significant improvement in function in favor of the SMP (SMD –0.26; 95% CI –0.35 to –0.16; heterogeneity X^2_6=5.2; *P*=.52; I^2=0%). Studies reporting the 12-month follow-up data (n=3 studies; n=707 participants) showed medium and significant overall effect of the digital-based structured SMP compared with usual care control condition (SMD –0.23; 95% CI –0.38 to –0.08; heterogeneity X^2_2=0.1; *P*=.96; I^2=0%). The quality of evidence was rated as *moderate* for both comparisons because of the high risk of bias in most studies (ie, study limitation).

The 2 studies comparing digital-based structured SMPs (internet) with physical activity did not show a significant difference between the 2 interventions neither at postintervention nor at longer follow-up time points (SMD 0.05; 95% CI –0.13 to 0.23; heterogeneity X^2_1=0.1; *P*=.77; I^2=0%; and SMD –0.03; 95% CI –0.22 to 0.16; heterogeneity X^2_1=0.01; *P*=.90; I^2=0%, respectively). The quality of evidence was rated *low* for both time points due to *limitation* and *imprecision*.

The digital-based structured SMP versus attention control group (health education), compared in 1 study [76], resulted in a small and nonsignificant (*P*=.15) improvement in physical function (SMD –0.17; 95% CI –0.39 to 0.06). The quality of evidence for this intervention was rated *very low* because of *imprecision* (ie, small sample size) and unknown consistency and/or publication bias.

Disability

One study [82] reported disability measures in 352 participants. The results show a minimal effect size in favor of internet-based SMPs compared with the usual care condition (SMD –0.10; 95% CI –0.17 to –0.03). However, the effect of intervention did not persist after 12 months of follow-up (SMD –0.00; 95% CI –0.21 to 0.20). The quality of evidence for this outcome at postintervention was rated as *very low* because of the high risk of bias in the study and unknown consistency and/or publication bias. The quality of evidence for the outcome at the 12-month follow-up was rated *very low* because of *imprecision* (ie, small sample size) and unknown consistency and/or publication bias for the follow-up time point.

Quality of Life

Only 1 study [79] reported the QoL outcome measure in 165 people with OA, indicating that the internet-based SMP did not make a significant improvement in QoL neither at 4 months postintervention nor at the 12-month follow-up (SMD –0.17; 95% CI –0.47 to 0.14 and SMD –0.07; 95% CI –0.39 to 0.26, respectively). The quality of evidence for this intervention at both time points was rated *very low* because of *imprecision* (ie, small sample size) and unknown consistency and/or publication bias.

XSL•FO
RenderX

Cost and Resource Use

Data reported on the cost of these interventions and resources used are limited. Allen et al [76] calculated cost by excluding nonrecurring labor costs, such as time for training a health educator at salary cost and the making of the intervention. The cost calculated included labor intervention delivery costs and indirect nonlabor costs, such as printing educational materials and creating compact discs. It was reported that the per-participant costs were US $107 (range US $100-US $121) for OA self-management and US $51 (range US $47-US $60) for health education. This cost was not compared with that of usual care. In the study by Lorig et al [82], no significant difference was reported between digital-based structured SMPs and usual care in terms of health care utilization (ie, physician, emergency, chiropractic or physical therapists visits or days in hospital).

Subgroup Analysis

The planned subgroup analysis based on the digital medium used to deliver the intervention revealed that the mobile app (n=1 study, n=197 participants) resulted in the largest effect size in pain reduction (SMD –0.38; 95% CI –0.67 to –0.10), followed by the internet medium (n=4 studies, n=841 participants; SMD –0.33; 95% CI –0.46 to –0.19; I^2=0%). The telephone as a medium of intervention delivery resulted in the smallest effect size in pain reduction (n=2 studies, n=576 participants; SMD –0.18; 95% CI –0.35 to –0.02; I^2=0%). However, the difference between the 2 groups was not significant (X^2_2=2.3; *P*=.32; I^2=13.4%). Subgroup differences for physical function outcome were not significant (X^2_2=0.1; *P*=.97; I^2=0%) between telephone, internet, and mobile apps, and resulted in similar small effect sizes (SMD –0.23; 95% CI –0.40 to –0.07, SMD –0.27; 95% CI –0.41 to –0.14, and SMD –0.24; 95% CI –0.51 to 0.03, respectively).

The intervention in most studies had exercise and/or physical activity components, and, in a few studies, it was the main component of the intervention. Therefore, we conducted a subgroup analysis accordingly. The pain reduction in studies with exercise and/or physical activity as the main component of the intervention was greater than the pain reduction in those studies without exercise and/or physical activity as the main component of the intervention (SMD –0.37; 95% CI – 0.54 to –0.20 and SMD –0.23; 95% CI –0.36 to –0.11, respectively). However, the difference was not statistically significant (X^2_1=1.7; *P*=.19; I^2=40.6%). The improvement in physical function was comparable in both groups—(SMD –0.27; 95% CI – 0.44 to –0.11) in studies with exercise and/or physical function as the main component and (SMD –0.24; 95% CI –0.38 to –0.11) in studies without exercise as the main component (X^2_1=0.1; *P*=.76; I^2=0%).

No subgroup analysis could be conducted for the follow-up time points as all 3 studies reporting the 12-month follow-up data used the internet as the intervention delivery medium. Similarly, no subgroup analysis could be conducted on the studies with high rates of completion because intervention was delivered via the telephone in 2 studies that did not report completion rate, or there was no measure of adherence.

Risk of Bias Sensitivity Analysis

Risk of bias sensitivity analysis, excluding studies with no or unclear random allocation, inadequate treatment allocation concealment, and/or incomplete outcome data, reduced the effect size in both pain reduction and improvement in physical function (SMD –0.19; 95% CI –0.32 to –0.05 and SMD –0.19; 95% CI –0.32 to –0.06, respectively).

Discussion

Principal Findings

The findings of this review indicate that digital-based structured SMPs compared with the treatment as usual or no intervention control groups resulted in a significant, homogeneous, and medium reduction in pain (SMD –0.28; 95% CI –0.38 to –0.18) and improvement in physical function (SMD –0.26; 95% CI –0.35 to –0.16). The SMDs reduced slightly at longer follow-up time points but remained significant with medium effect sizes. However, the findings should be interpreted with caution as the overall quality of the body of evidence is moderate because of the risk of bias in the included studies. Using SD from the control group of the largest (most adequately powered) study by Lorig et al [82], the SMDs translate to 5.70% reduction in pain and 5.07% improvement in physical function at postintervention time points. In accordance with Tubach et al [84] and Angst et al [85], we determined a minimal clinically important difference (MCID) of 15% in pain and 8% in physical function. Therefore, these effects are unlikely to be clinically significant. The findings on disability and QoL are less clear, as these outcomes were reported in 1 study with very small effect sizes.

Both immediate and longer-term effects of digital-based structured SMPs on pain in our review (SMD –0.28; 95% CI –0.38 to –0.18 and SMD –0.20; 95% CI –0.35 to –0.5, respectively) were slightly greater than those reported in a Cochrane review comparing standard OA SMPs with usual care (SMD –0.26; 95% CI –0.41 to –0.10 and SMD –0.17; 95% CI –0.26 to –0.08, respectively) [19]. However, the immediate and longer-term effect of digital-based structured SMPs on physical function was greater in our review compared with the immediate effect of SMPs reported in the Cochrane review by Kroon et al (SMD 0.01; 95% CI –019 to 0.18 and SMD –0.16; 95% CI –0.25 to –0.01, respectively) [19]. Despite the moderate quality of evidence reported in both reviews and a lack of direct comparison between the digital-based structured SMP and the SMP intervention, it could be postulated that the digital-based structured SMPs may have similar or greater effects than conventional SMPs on pain and physical function. Considering the potential cost-effectiveness of digital interventions, digital-based structured SMPs have the potential to reach populations reluctant or unable to attend face-to-face appointments. This could have a considerable benefit from a public health perspective despite the modest size of the effect.

Previous reviews indicate that eHealth interventions have the potential to reduce treatment costs [21,86]. We could not assess intervention costs because none of the included studies compared intervention costs across groups. In the study by Kloek et al [80], participants in the intervention group

XSL•FO
RenderX

(internet-based exercise) had 5 in-person physical therapy sessions, whereas those in the control group visited a physical therapist an average of 12 times. Nevertheless, both groups showed a significant improvement in most health outcomes. Reduced numbers of physical therapy visits would likely result in a reduction in health care costs. Future RCTs of digital-based structured SMPs are encouraged to assess the cost-effectiveness of digital interventions in isolation or combination with face-to-face sessions. It should be noted that smartphones and/or internet devices are now owned by the majority of the population; therefore, these interventions are accessible and can be delivered conveniently and easily to the target audience. However, it should be noted that digital-based structured SMPs may not be suitable for all patients, potentially because of age, preference, comorbidities, and/or severity of illness; therefore, the intervention must be tailored to the needs, preferences, and conditions of patients and include face-to-face, group, and digital modes in the intervention package.

The result of the subgroup analysis indicates that there is no significant difference among the different digital modes of SMP delivery. However, mobile app (1 study) and internet SMPs (4 studies) resulted in medium effects on pain outcome, whereas telephone SMPs (2 studies) showed a small effect. Potential reasons for this observation could be that the majority of participants in the telephone studies were male veterans (>90%) who had more comorbidities and/or severe OA symptoms. In addition, the interventions were low intensity, and telephone call sessions were short (average 16.6 min). However, the small number of studies and the fact that the 2 telephone trials had a low risk of bias cautions against this interpretation; inflated effect size in studies with a high risk of bias is expected. The results of further subgroup analysis indicate that studies with exercise or physical activity as the main component of the intervention resulted in a greater improvement in pain but not in perceived physical function. A possible explanation for the reduced pain in this group of studies could be because of the change in pain tolerance and decreased perception of pain after exercise [87,88]. Four studies (n=2 telephone and n=2 internet) were included in the risk of bias sensitivity analysis. The analysis resulted in a reduction in the effect size compared with the main meta-analysis for both pain and physical function outcomes. Notably, the 2 combined telephone and video studies contributed most to the SMD (weight >65%) in the sensitivity analysis.

Attrition at postintervention was modest; median 10% (IQR 7%) in the intervention groups and median 8% (IQR 4%) in the control conditions. Despite common concerns with high dropout rates in digital interventions [89], the reported attrition rates in the included studies appear reasonable compared with behavioral studies. In a recent study, Bennell et al [90] reported that a web-based exercise program improved home exercise adherence and confidence in the ability to undertake exercise compared with a home-based exercise program prescribed by a physical therapist's usual methods [90]. In our review, treatment adherence in 6 studies reporting treatment completers was 46% in 1 study [79], >80% in 3 studies [78,80,81], and >90% in 2 studies (Table 1) [82,83]. There are a few possible reasons for the relatively low dropout and nonadherence rates. First, interventions in all studies were tailored to participants' needs and conditions. Second, researchers used features in web-based or mobile apps to develop reminder and monitoring systems and created higher interactivity of the intervention delivery. Third, in some studies, health professionals maintained contact with the study participants during the study through either face-to-face meetings or telephone calls. However, it should be noted that the participants were highly educated in most studies and, in some studies, self-selected (ie, responded to the study participants' recruitment advertisement). Thus, they may have been enthusiastic about the new intervention. In future research and development, the advanced features of mobile apps and internet interventions could be employed to deliver even more effective and tailored monitoring and/or motivational interventions. Digital interventions can also be used to show patients' symptom improvement, which is a useful and effective way not only to improve retention and compliance but also to increase their effect through self-efficacy [91,92].

There are a few limitations to the current review. First, only studies published in English were considered. Second, blinding of intervention and outcome assessment is not possible; therefore, all studies suffer from performance and detection biases. Moreover, some of the most powered studies have a high risk of bias, lowering the overall quality of evidence. Third, disability and QoL outcomes were only reported in 1 study each. Fourth, this review has limited generalizability because participants in most studies were highly educated and self-selected; thus, they may have been highly motivated. Finally, interventions in some studies were targeted to exercise and/or physical activity as the main component of the program, and behavior change techniques were used to improve exercise adherence [78-81]; the multiplicity and heterogeneity of techniques prohibited meta-analysis of the effects of particular techniques. However, almost all studies employed theoretically driven interventions incorporating important components of the SMPs (ie, education, goal setting, action planning, problem solving, skills acquisition, self-monitoring, understanding illness, and managing emotions).

In conclusion, digital-based structured SMPs resulted in medium improvements in pain and physical function postintervention. However, the effects are below the MCID and may, therefore, not be clinically significant. The quality of the evidence for pain and function was graded as *medium* because of the high risk of bias in the studies; therefore, the true effect is likely to be close to our estimate of the effect, but there is a possibility that it is substantially different. The effect of digital-based structured SMP intervention on pain and physical function was slightly reduced at the 12-month follow-up, but remained at the medium threshold. One study reporting results on disability and one study reporting results on health-related QoL indicated small improvements in both outcomes. We rated the quality of the evidence as *low* and *very low* for disability and health-related QoL, respectively, indicating that the true effects are likely to be substantially different from our estimate of effect.

Conclusions

This review of digital-based structured SMPs on self-reported outcomes including pain, physical function, disability, and

XSL•FO
RenderX

health-related QoL in patients with OA revealed 6 RCTs and 2 cluster RCTs. digital-based structured SMPs resulted in medium improvements in pain and physical function postintervention, but these may not be clinically significant. These effects were slightly reduced at the 12-month follow-up, but remained at the medium threshold. The quality of the evidence for pain and function was graded as *medium* because of a high risk of bias in some studies. More high-quality studies are needed, and the routine assessment of QoL and disability would be useful.

Acknowledgments

The authors would like to thank the following individuals for providing consultation with the project: Ms Jill Boland, Professor Derek Ward, Dr Buk Dhadda, Dr Merryl Watkins, Dr Vinod Chudasama, Mrs Victoria Price, and Mr Graham Joynes. The authors thank Ms Emillie Haspel for helping with data extraction. They also thank the authors who provided data for the meta-analysis.

This study is funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East Midlands. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 43 KB - jmir_v22i7e15365_app1.docx]

Multimedia Appendix 2
Studies excluded with reasons.
[DOCX File , 28 KB - jmir_v22i7e15365_app2.docx]

Multimedia Appendix 3
Behavior change technique taxonomy grouping with explanation and example texts from studies.
[DOCX File , 38 KB - jmir_v22i7e15365_app3.docx]

Multimedia Appendix 4
Forest plots of review comparisons.
[DOCX File , 220 KB - jmir_v22i7e15365_app4.docx]

References

1. Palazzo C, Nguyen C, Lefevre-Colau MM, Rannou F, Poiraudeau S. Risk factors and burden of osteoarthritis. Ann Phys Rehabil Med 2016 Jun;59(3):134-138 [FREE Full text] [doi: 10.1016/j.rehab.2016.01.006] [Medline: 26904959]
2. Hawker GA, Gignac MA, Badley E, Davis AM, French MR, Li Y, et al. A longitudinal study to explain the pain-depression link in older adults with osteoarthritis. Arthritis Care Res (Hoboken) 2011 Oct;63(10):1382-1390 [FREE Full text] [doi: 10.1002/acr.20298] [Medline: 20662042]
3. Litwic A, Edwards MH, Dennison EM, Cooper C. Epidemiology and burden of osteoarthritis. Br Med Bull 2013;105:185-199 [FREE Full text] [doi: 10.1093/bmb/lds038] [Medline: 23337796]
4. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
5. Thune I, Furberg AS. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S530-50; discussion S609. [doi: 10.1097/00005768-200106001-00025] [Medline: 11427781]
6. Williams A, Kamper SJ, Wiggers JH, O'Brien KM, Lee H, Wolfenden L, et al. Musculoskeletal conditions may increase the risk of chronic disease: a systematic review and meta-analysis of cohort studies. BMC Med 2018 Sep 25;16(1):167 [FREE Full text] [doi: 10.1186/s12916-018-1151-2] [Medline: 30249247]
7. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Prevalence and burden of osteoarthritis: results from a population survey in Norway. J Rheumatol 2008 Apr;35(4):677-684. [Medline: 18278832]
8. Hermans J, Koopmanschap MA, Bierma-Zeinstra SM, van Linge JH, Verhaar JA, Reijman M, et al. Productivity costs and medical costs among working patients with knee osteoarthritis. Arthritis Care Res 2012 May 25;64(6):853-861. [doi: 10.1002/acr.21617]

XSL•FO
RenderX

9. Salaffi F, Carotti M, Stancati A, Grassi W. Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with matched healthy controls. Aging Clin Exp Res 2005 Aug;17(4):255-263. [doi: 10.1007/BF03324607] [Medline: 16285189]
10. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: the chronic osteoarthritis management initiative of the US bone and joint initiative. Semin Arthritis Rheum 2014 Jun;43(6):701-712. [doi: 10.1016/j.semarthrit.2013.11.012] [Medline: 24387819]
11. Osteoarthritis: Care and Management. The National Institute for Health and Care Excellence. 2014. URL: https://www.nice.org.uk/guidance/cg177/chapter/1-Recommendations#education-and-self-management-2 [accessed 2020-06-11]
12. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014 Mar;22(3):363-388 [FREE Full text] [doi: 10.1016/j.joca.2014.01.003] [Medline: 24462672]
13. Osborne RH, Spinks JM, Wicks IP. Patient education and self-management programs in arthritis. Med J Aust 2004 Mar 1;180(S5):S23-S26. [Medline: 14984359]
14. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
15. Warsi A, LaValley MP, Wang PS, Avorn J, Solomon DH. Arthritis self-management education programs: a meta-analysis of the effect on pain and disability. Arthritis Rheum 2003 Aug;48(8):2207-2213 [FREE Full text] [doi: 10.1002/art.11210] [Medline: 12905474]
16. Walsh N, Mitchell H, Reeves B, Hurley M. Integrated exercise and self-management programmes in osteoarthritis of the hip and knee: a systematic review of effectiveness. Phys Ther Rev 2013 Jul 19;11(4):289-297. [doi: 10.1179/108331906X163432]
17. Devos-Comby L, Cronan T, Roesch SC. Do exercise and self-management interventions benefit patients with osteoarthritis of the knee? A metaanalytic review. J Rheumatol 2006 Apr;33(4):744-756. [Medline: 16583478]
18. Smith T, Davies L, McConnell L, Cross J, Hing C. Self-management programmes for people with osteoarthritis: a systematic review and meta-analysis. Curr Rheumatol Rev 2014 Jan 31;9(3):165-175. [doi: 10.2174/15733971090314013012185 9]
19. Kroon F, van der Burg LR, Buchbinder R, Osborne R, Johnston R, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database Syst Rev 2014 Jan 15(1):CD008963. [doi: 10.1002/14651858.CD008963.pub2] [Medline: 24425500]
20. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
21. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of eHealth technologies to enable the implementation of musculoskeletal models of care: evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502 [FREE Full text] [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
22. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an internet-based chronic pain self-management program. Int J Med Inform 2017 Jan;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]
23. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019.
24. Schäfer AG, Zalpour C, von Piekartz H, Hall TM, Paelke V. The efficacy of electronic health-supported home exercise interventions for patients with osteoarthritis of the knee: systematic review. J Med Internet Res 2018 Apr 26;20(4):e152 [FREE Full text] [doi: 10.2196/jmir.9465] [Medline: 29699963]
25. O'Brien K, Hodder R, Wiggers J, Williams A, Campbell E, Wolfenden L, et al. Effectiveness of telephone-based interventions for managing osteoarthritis and spinal pain: a systematic review and meta-analysis. PeerJ 2018;6:e5846 [FREE Full text] [doi: 10.7717/peerj.5846] [Medline: 30397549]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions The Cochrane Collaboration. Chichester, West Sussex: Wiley-Blackwell; 2019.
28. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
29. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. New York, USA: Academic Press; 1997.
31. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

XSL•FO
RenderX

32. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. Br Med J 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
33. Ackerman IN, Bucknill A, Page RS, Broughton NS, Roberts C, Cavka B, et al. Preferences for disease-related education and support among younger people with hip or knee osteoarthritis. Arthritis Care Res (Hoboken) 2017 Apr;69(4):499-508 [FREE Full text] [doi: 10.1002/acr.22950] [Medline: 27273912]
34. Osteras N, Moe RH, Fernandes L. Un programme de soutien téléphonique à la prise en charge personnelle pourrait être efficace pour réduire la douleur arthrosique. Article in French. Kinésithérapie, la Revue 2015;15(167):8-9. [doi: 10.1016/j.kine.2015.07.017]
35. Rippey RM, Bill D, Abeles M, Day J, Downing DS, Pfeiffer CA, et al. Computer-based patient education for older persons with osteoarthritis. Arthritis Rheum 1987 Aug;30(8):932-935. [doi: 10.1002/art.1780300814] [Medline: 3307793]
36. Allen KD, Sanders LL, Olsen MK, Bowlby L, Katz JN, Mather RC, et al. Internet versus DVD Decision aids for hip and knee osteoarthritis. Musculoskeletal Care 2016 Jun;14(2):87-97. [doi: 10.1002/msc.1116] [Medline: 26220784]
37. de Achaval S, Fraenkel L, Volk RJ, Cox V, Suarez-Almazor ME. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. Arthritis Care Res (Hoboken) 2012 Feb;64(2):229-237 [FREE Full text] [doi: 10.1002/acr.20646] [Medline: 21954198]
38. René J, Weinberger M, Mazzuca SA, Brandt KD, Katz BP. Reduction of joint pain in patients with knee osteoarthritis who have received monthly telephone calls from lay personnel and whose medical treatment regimens have remained stable. Arthritis Rheum 1992 May;35(5):511-515. [doi: 10.1002/art.1780350504] [Medline: 1575787]
39. Srikesavan CS, Shay B, Szturm T. Task-oriented training with computer games for people with rheumatoid arthritis or hand osteoarthritis: a feasibility randomized controlled trial. Games Health J 2016 Oct;5(5):295-303. [doi: 10.1089/g4h.2016.0005] [Medline: 27623532]
40. Blixen CE, Bramstedt KA, Hammel JP, Tilley BC. A pilot study of health education via a nurse-run telephone self-management programme for elderly people with osteoarthritis. J Telemed Telecare 2004;10(1):44-49. [doi: 10.1258/135763304322764194] [Medline: 15006216]
41. Cuperus N, Hoogeboom TJ, Kersten CC, den Broeder AA, Vlieland TP, van den Ende CH. Randomized trial of the effectiveness of a non-pharmacological multidisciplinary face-to-face treatment program on daily function compared to a telephone-based treatment program in patients with generalized osteoarthritis. Osteoarthritis Cartilage 2015 Aug;23(8):1267-1275 [FREE Full text] [doi: 10.1016/j.joca.2015.04.007] [Medline: 25887365]
42. Cuperus N, van den Hout WB, Hoogeboom TJ, van den Hoogen FH, Vliet Vlieland TP, van den Ende CH. Cost-utility and cost-effectiveness analyses of face-to-face versus telephone-based nonpharmacologic multidisciplinary treatments for patients with generalized osteoarthritis. Arthritis Care Res (Hoboken) 2016 Apr;68(4):502-510 [FREE Full text] [doi: 10.1002/acr.22709] [Medline: 26314289]
43. Sciamanna CN, Harrold LR, Manocchia M, Walker NJ, Mui S. The effect of web-based, personalized, osteoarthritis quality improvement feedback on patient satisfaction with osteoarthritis care. Am J Med Qual 2005;20(3):127-137. [doi: 10.1177/1062860605274518] [Medline: 15951518]
44. Barlow JH, Turner AP, Wright CC. A randomized controlled study of the Arthritis Self-Management Programme in the UK. Health Educ Res 2000 Dec;15(6):665-680. [doi: 10.1093/her/15.6.665] [Medline: 11142075]
45. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94 [FREE Full text] [doi: 10.1002/acr.22915] [Medline: 27111441]
46. Brosseau L, Wells G, Brooks S, Bell M, Angelis G, Loew L. The knowledge-to-action cycle (KTAC): a Canadian model to guide the dissemination and the implementation of the bilingual people getting a grip on arthritis self-management program. J Rheumatol 2014;41:1520 [FREE Full text]
47. Crossley KM, Vicenzino B, Lentzos J, Schache AG, Pandy MG, Ozturk H, et al. Exercise, education, manual-therapy and taping compared to education for patellofemoral osteoarthritis: a blinded, randomised clinical trial. Osteoarthritis Cartilage 2015 Sep;23(9):1457-1464 [FREE Full text] [doi: 10.1016/j.joca.2015.04.024] [Medline: 25960116]
48. Crotty M, Prendergast J, Battersby MW, Rowett D, Graves SE, Leach G, et al. Self-management and peer support among people with arthritis on a hospital joint replacement waiting list: a randomised controlled trial. Osteoarthritis Cartilage 2009 Nov;17(11):1428-1433 [FREE Full text] [doi: 10.1016/j.joca.2009.05.010] [Medline: 19486959]
49. da Silva FS, de Melo FE, do Amaral MM, Caldas VV, Pinheiro IL, Abreu BJ, et al. Efficacy of simple integrated group rehabilitation program for patients with knee osteoarthritis: single-blind randomized controlled trial. J Rehabil Res Dev 2015;52(3):309-322 [FREE Full text] [doi: 10.1682/JRRD.2014.08.0199] [Medline: 26237073]
50. Edworthy SM, Devins GM. Improving medication adherence through patient education distinguishing between appropriate and inappropriate utilization. Patient Education Study Group. J Rheumatol 1999 Aug;26(8):1793-1801. [Medline: 10451079]
51. Farr JN, Going SB, McKnight PE, Kasle S, Cussler EC, Cornett M. Progressive resistance training improves overall physical activity levels in patients with early osteoarthritis of the knee: a randomized controlled trial. Phys Ther 2010 Mar;90(3):356-366 [FREE Full text] [doi: 10.2522/ptj.20090041] [Medline: 20056719]

XSL•FO
RenderX

52. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care Res (Hoboken) 2012 Feb;64(2):238-247 [FREE Full text] [doi: 10.1002/acr.20642] [Medline: 21954131]
53. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E, et al. Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. Arthritis Rheum 2007 Oct 15;57(7):1211-1219 [FREE Full text] [doi: 10.1002/art.22995] [Medline: 17907147]
54. Jessep SA, Walsh NE, Ratcliffe J, Hurley MV. Long-term clinical benefits and costs of an integrated rehabilitation programme compared with outpatient physiotherapy for chronic knee pain. Physiotherapy 2009 Jun;95(2):94-102. [doi: 10.1016/j.physio.2009.01.005] [Medline: 19627690]
55. Kao MJ, Wu MP, Tsai MW, Chang WW, Wu SF. The effectiveness of a self-management program on quality of life for knee osteoarthritis (OA) patients. Arch Gerontol Geriatr 2012;54(2):317-324. [doi: 10.1016/j.archger.2011.05.018] [Medline: 21726907]
56. Laforest S, Nour K, Gignac M, Gauvin L, Parisien M, Poirier M. Short-term effects of a self-management intervention on health status of housebound older adults with arthritis. J Appl Gerontol 2008 Aug 29;27(5):539-567. [doi: 10.1177/0733464808319712]
57. Marconcin P, Espanha M, Teles J, Bento P, Campos P, André R, et al. A randomized controlled trial of a combined self-management and exercise intervention for elderly people with osteoarthritis of the knee: the PLENO program. Clin Rehabil 2018 Feb;32(2):223-232. [doi: 10.1177/0269215517718892] [Medline: 28714343]
58. McDonald DD, Walsh S, Vergara C, Gifford T, Weiner DK. The effect of a Spanish virtual pain coach for older adults: a pilot study. Pain Med 2012 Nov;13(11):1397-1406. [doi: 10.1111/j.1526-4637.2012.01491.x] [Medline: 22994410]
59. McKnight PE, Kasle S, Going S, Villanueva I, Cornett M, Farr J, et al. A comparison of strength training, self-management, and the combination for early osteoarthritis of the knee. Arthritis Care Res (Hoboken) 2010 Jan 15;62(1):45-53 [FREE Full text] [doi: 10.1002/acr.20013] [Medline: 20191490]
60. Nour K, Laforest S, Gauvin L, Gignac M. Behavior change following a self-management intervention for housebound older adults with arthritis: an experimental study. Int J Behav Nutr Phys Act 2006 May 30;3:12 [FREE Full text] [doi: 10.1186/1479-5868-3-12] [Medline: 16734904]
61. Patel A, Buszewicz M, Beecham J, Griffin M, Rait G, Nazareth I, et al. Economic evaluation of arthritis self management in primary care. BMJ 2009 Sep 22;339:b3532 [FREE Full text] [doi: 10.1136/bmj.b3532] [Medline: 19773329]
62. Solomon DH, Warsi A, Brown-Stevenson T, Farrell M, Gauthier S, Mikels D, et al. Does self-management education benefit all populations with arthritis? A randomized controlled trial in a primary care physician network. J Rheumatol 2002 Feb;29(2):362-368. [Medline: 11838857]
63. Somers TJ, Blumenthal JA, Guilak F, Kraus VB, Schmitt DO, Babyak MA, et al. Pain coping skills training and lifestyle behavioral weight management in patients with knee osteoarthritis: a randomized controlled study. Pain 2012 Jun;153(6):1199-1209. [doi: 10.1016/j.pain.2012.02.023] [Medline: 22503223]
64. Wu SF, Kao MJ, Wu MP, Tsai MW, Chang WW. Effects of an osteoarthritis self-management programme. J Adv Nurs 2011 Jul;67(7):1491-1501. [doi: 10.1111/j.1365-2648.2010.05603.x] [Medline: 21375573]
65. Yip YB, Sit JW, Fung KK, Wong DY, Chong SY, Chung LH, et al. Impact of an Arthritis Self-Management Programme with an added exercise component for osteoarthritic knee sufferers on improving pain, functional outcomes, and use of health care services: An experimental study. Patient Educ Couns 2007 Jan;65(1):113-121. [doi: 10.1016/j.pec.2006.06.019] [Medline: 17010554]
66. Yip YB, Sit JW, Wong DY, Chong SY, Chung LH. A 1-year follow-up of an experimental study of a self-management arthritis programme with an added exercise component of clients with osteoarthritis of the knee. Psychol Health Med 2008 Aug;13(4):402-414. [doi: 10.1080/13548500701584030] [Medline: 18825579]
67. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
68. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C, et al. Moderators of effects of internet-delivered exercise and pain coping skills training for people with knee osteoarthritis: exploratory analysis of the impact randomized controlled trial. J Med Internet Res 2018 May 9;20(5):e10021 [FREE Full text] [doi: 10.2196/10021] [Medline: 29743149]
69. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
70. Wilson M, Roll JM, Corbett C, Barbosa-Leiker C. Empowering patients with persistent pain using an internet-based self-management program. Pain Manag Nurs 2015 Aug;16(4):503-514. [doi: 10.1016/j.pmn.2014.09.009] [Medline: 26088940]
71. Bennell KL, Rini C, Keefe F, French S, Nelligan R, Kasza J, et al. Effects of adding an internet-based pain coping skills training protocol to a standardized education and exercise program for people with persistent hip pain (HOPE trial): randomized controlled trial protocol. Phys Ther 2015 Oct;95(10):1408-1422 [FREE Full text] [doi: 10.2522/ptj.20150119] [Medline: 26023213]

XSL•FO
RenderX

72. Dobson F, Hinman RS, French S, Rini C, Keefe F, Nelligan R, et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskelet Disord 2014 Aug 13;15:279 [FREE Full text] [doi: 10.1186/1471-2474-15-279] [Medline: 25125068]
73. Lopez-Olivo M, Barbo A, Rizvi T, Volk R, Lin H, Suarez-Almazor M. A randomized controlled trial to evaluate a multimedia patient education tool in patients with knee osteoarthritis. Six-month results. Arthritis Rheumatol 2015 Oct;67(Suppl 10):-Abstract [FREE Full text] [doi: 10.1002/art.39448] [Medline: 26540142]
74. Williamson OD. An online exercise program plus automated coping skills training improved pain and function in chronic knee pain. Ann Intern Med 2017 Jun 20;166(12):JC67. [doi: 10.7326/ACPJC-2017-166-12-067] [Medline: 28630979]
75. Sperber NR, Bosworth HB, Coffman CJ, Lindquist JH, Oddone EZ, Weinberger M, et al. Differences in osteoarthritis self-management support intervention outcomes according to race and health literacy. Health Educ Res 2013 Jun;28(3):502-511. [doi: 10.1093/her/cyt043] [Medline: 23525779]
76. Allen KD, Oddone EZ, Coffman CJ, Datta SK, Juntilla KA, Lindquist JH, et al. Telephone-based self-management of osteoarthritis: a randomized trial. Ann Intern Med 2010 Nov 2;153(9):570-579. [doi: 10.7326/0003-4819-153-9-201011020-00006] [Medline: 21041576]
77. Allen KD, Yancy WS, Bosworth HB, Coffman CJ, Jeffreys AS, Datta SK, et al. A combined patient and provider intervention for management of osteoarthritis in veterans: a randomized clinical trial. Ann Intern Med 2016 Jan 19;164(2):73-83 [FREE Full text] [doi: 10.7326/M15-0378] [Medline: 26720751]
78. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
79. Bossen D, Veenhof C, van Beek KE, Spreeuwenberg PM, Dekker J, de Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
80. Kloek CJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 1;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
81. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 9;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
82. Lorig K, Ritter P, Laurent D, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
83. Rini C, Porter L, Somers T, McKee D, DeVellis R, Smith M, et al. Automated internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
84. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005 Jan;64(1):29-33 [FREE Full text] [doi: 10.1136/ard.2004.022905] [Medline: 15208174]
85. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001 Aug;45(4):384-391 [FREE Full text] [doi: 10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0] [Medline: 11501727]
86. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
87. O'Reilly SC, Muir KR, Doherty M. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. Ann Rheum Dis 1999 Jan;58(1):15-19 [FREE Full text] [doi: 10.1136/ard.58.1.15] [Medline: 10343535]
88. Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. Br Med J 2002 Oct 5;325(7367):752 [FREE Full text] [doi: 10.1136/bmj.325.7367.752] [Medline: 12364304]
89. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
90. Bennell KL, Marshall CJ, Dobson F, Kasza J, Lonsdale C, Hinman RS. Does a web-based exercise programming system improve home exercise adherence for people with musculoskeletal conditions?: a randomized controlled trial. Am J Phys Med Rehabil 2019 Oct;98(10):850-858. [doi: 10.1097/PHM.0000000000001204] [Medline: 31021823]

XSL•FO
RenderX

91. Hurley M, Dickson K, Hallett R, Grant R, Hauari H, Walsh N, et al. Exercise interventions and patient beliefs for people with hip, knee or hip and knee osteoarthritis: a mixed methods review. Cochrane Database Syst Rev 2018 Apr 17;4:CD010842 [FREE Full text] [doi: 10.1002/14651858.CD010842.pub2] [Medline: 29664187]
92. Williamson W, Kluzek S, Roberts N, Richards J, Arden N, Leeson P, et al. Behavioural physical activity interventions in participants with lower-limb osteoarthritis: a systematic review with meta-analysis. BMJ Open 2015 Aug 10;5(8):e007642 [FREE Full text] [doi: 10.1136/bmjopen-2015-007642] [Medline: 26260348]

Abbreviations

eHealth: electronic health
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
MCID: minimal clinically important difference
NIHR: National Institute for Health Research
OA: osteoarthritis
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference
SMP: self-management program
TiDieR: template for intervention description and replication
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.07.19; peer-reviewed by A Briggs, S Ahmed, M Sewitch; comments to author 23.09.19; revised version received 13.11.19; accepted 26.02.20; published 20.07.20.

Please cite as:
Safari R, Jackson J, Sheffield D
Digital Self-Management Interventions for People With Osteoarthritis: Systematic Review With Meta-Analysis
J Med Internet Res 2020;22(7):e15365
URL: http://www.jmir.org/2020/7/e15365/
doi:10.2196/15365
PMID:32706657

©Reza Safari, Jessica Jackson, David Sheffield. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review

Siqi Liu[1,2], BEng; Kay Choong See[3], MBBS; Kee Yuan Ngiam[4], MBBS, MRCS, MMed, FRCS; Leo Anthony Celi[5,6], MD, MS, MPH; Xingzhi Sun[7], PhD; Mengling Feng[2], PhD

[1]NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore
[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[3]Division of Respiratory & Critical Care Medicine, National University Hospital, Singapore, Singapore
[4]Group Chief Technology Office, National University Health System, Singapore, Singapore
[5]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States
[6]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States
[7]Ping An Health Technology, Beijing, China

Corresponding Author:
Mengling Feng, PhD
Saw Swee Hock School of Public Health
National University of Singapore
12 Science Drive 2, #10-01
Singapore, 117549
Singapore
Phone: 65 65164988
Email: ephfm@nus.edu.sg

Abstract

Background: Decision support systems based on reinforcement learning (RL) have been implemented to facilitate the delivery of personalized care. This paper aimed to provide a comprehensive review of RL applications in the critical care setting.

Objective: This review aimed to survey the literature on RL applications for clinical decision support in critical care and to provide insight into the challenges of applying various RL models.

Methods: We performed an extensive search of the following databases: PubMed, Google Scholar, Institute of Electrical and Electronics Engineers (IEEE), ScienceDirect, Web of Science, Medical Literature Analysis and Retrieval System Online (MEDLINE), and Excerpta Medica Database (EMBASE). Studies published over the past 10 years (2010-2019) that have applied RL for critical care were included.

Results: We included 21 papers and found that RL has been used to optimize the choice of medications, drug dosing, and timing of interventions and to target personalized laboratory values. We further compared and contrasted the design of the RL models and the evaluation metrics for each application.

Conclusions: RL has great potential for enhancing decision making in critical care. Challenges regarding RL system design, evaluation metrics, and model choice exist. More importantly, further work is required to validate RL in authentic clinical environments.

(J Med Internet Res 2020;22(7):e18477) doi:10.2196/18477

KEYWORDS

artificial intelligence; reinforcement learning; critical care; decision support systems, clinical; intensive care unit; machine learning

Introduction

Background

In the health care domain, clinical processes are dynamic because of the high prevalence of complex diseases and dynamic changes in the clinical conditions of patients. Existing treatment recommendation systems are mainly implemented using rule-based protocols defined by physicians based on evidence-based clinical guidelines or best practices [1-3]. In addition, these protocols and guidelines may not consider multiple comorbid conditions [4]. In an intensive care unit

XSL•FO
RenderX

(ICU), critically ill patients may benefit from deviation from established treatment protocols and from personalizing patient care using means not based on rules [5,6].

When physicians need to adapt treatment for individual patients, they may take reference from randomized controlled trials (RCTs), systemic reviews, and meta-analyses. However, RCTs may not be available or definitive for many ICU conditions. Many patients admitted to ICUs might also be too ill for inclusion in clinical trials [6]. Furthermore, only 9% of treatment recommendations in the ICU are based on RCTs [7], and the vast majority of RCTs in critical care have negative findings [8]. To aid clinical decisions in ICUs, we need other methods, including the use of large observational data sets. ICU data can be useful for learning about patients as they were collected in a data-rich environment. A large amount of data can then be fed into artificial intelligence (AI) systems (using computers to mimic human cognitive functions) and machine learning methods (using computer algorithms to perform clinical tasks without the need for explicit instructions). AI and machine learning can then help with diagnosis [9,10], treatment [11,12], and resource management [13,14] in the ICU. Given the dynamic nature of critically ill patients, one machine learning method called reinforcement learning (RL) is particularly suitable for ICU settings.

Fundamentals of Reinforcement Learning

RL is a goal-oriented learning tool where a computer *agent*, acting as a decision maker, analyzes available data within its defined environment [15], derives a rule for taking actions, and optimizes long-term rewards. The agent is the RL model that we wish to develop. In general, an RL agent receives evaluative feedback about the performance of its action in each time step, allowing it to improve the performance of subsequent actions by trial and error [16]. Mathematically, this sequential decision-making process is called the Markov decision process (MDP) [17]. An MDP is defined by 4 major components: (1) a state that represents the environment at each time; (2) an action the agent takes at each time that influences the next state; (3) a transition probability that provides an estimate for reaching different subsequent states, which reflects the environment for an agent to interact with; (4) a reward function is the observed feedback given a state-action pair. The solution of the MDP is an optimized set of rules and is termed the policy.

RL has already emerged as an effective tool to solve complicated control problems with large-scale, high-dimensional data in some application domains, including video games, board games, and autonomous control [18-20]. In these domains, RL has been proven to achieve human-level capacity for learning complex sequential decisions. For instance, Alpha Go is an RL agent for playing the strategy board game Go. On the basis of Alpha Go's learned policy, and given the current position of the Go stones, it is possible to decide where the next white/black stone should be placed on the board to maximize its chance of winning.

Analogies to Critical Care

For critical care, given the large amount and granular nature of recorded data, RL is well suited for providing sequential treatment suggestions, optimizing treatments, and improving outcomes for new ICU patients. RL also has the potential to expand our understanding of existing clinical protocols by automatically exploring various treatment options. The RL agent analyzes the patient trajectories, and through trial and error, derives a policy, a personalized treatment protocol that optimizes the probability of favorable clinical outcomes (eg, survival). As this computerized process is an attempt to mimic the human clinician's thought process, RL has also been called the AI clinician [21].

We can consider the state as the well-being/condition of a patient. The state of the patients could depend on static traits (eg, patient demographics including age, gender, ethnicity, pre-existing comorbidity) and longitudinal measurements (eg, vital signs, laboratory test results). An action is a treatment or an intervention that physicians do for patients (eg, prescription of medications and ordering of laboratory tests). The transition probability is the likelihood of state transitions, and it is viewed as a prognosis. If the well-being in the new state is improved, we assign a reward to the RL agent, but we penalize the agent if the patient's condition worsens or stays stagnant after the intervention.

As illustrated in Figure 1, if we take a snapshot of the current well-being of a patient as his/her state, the physician would provide a treatment or an intervention (an action) to the patient. This action would lead the patient to the next state depending on his/her current state and the action performed on him/her. While knowing the next state of the patient, the physician would need to take another action according to the new state. These state-action pairs would continue to rollout over time, and the resultant trajectory of state-action pairs could represent the changes in the patients' conditions and the sequential treatment decisions that were performed by the physicians. We can define the length of the trajectory for each patient as fixed (eg, during the first 24 hours of the ICUs stay) or as dynamic (eg, different patients could be discharged from the ICUs at different times).

XSL•FO
RenderX

Figure 1. Illustration of reinforcement learning in critical care.



The main objective of the RL algorithm is to train an agent that can maximize the cumulative future reward from the state-action pairs given the patients' state-action trajectories. When a new state is observed, the agent is able to perform an action, which could choose the action for the greatest long-term outcome (eg, survival). When the RL agent is well-trained, it is possible to pick the best action given the state of a patient, and we describe this process as acting according to an optimal policy.

A policy is analogous to a clinical protocol. Nonetheless, a policy has advantages over a clinical protocol because it is capable of capturing more personalized details of individual patients. A policy can be represented by a table where it maps all possible states with actions. Alternatively, a policy could also be represented by a deep neural network (DNN) where given the input of a patient's state, the DNN model outputs the highest probability of an action. An optimal policy can be trained using various RL algorithms. Some widely applied RL algorithms include the fitted-Q-iteration (FQI) [22], deep Q network (DQN) [23], actor-critic network [24], and model-based RL [25]. More technical details about various RL models have been explained [26,27].

As RL in critical care is a relatively nascent field, we therefore aimed to review all the existing clinical applications that applied RL in the ICU setting for decision support over the past 10 years (2010-2019). Specifically, we aimed to categorize RL applications and summarize and compare different RL designs. We hope that our overview of RL applications in critical care can help reveal both the advances and gaps for future clinical development of RL. A detailed explanation of the concept of RL and its algorithms is available in Multimedia Appendix 1 [28].

Methods

Search Strategy

A review of the literature was conducted using the following 7 databases: PubMed, Institute of Electrical and Electronics Engineers (IEEE), Google Scholar, Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica Database (EMBASE), ScienceDirect, and Web of Science. The search terms *reinforcement learning*, *critical care*, *intensive care*, *intensive care units*, and *ICUs* were combined. The search phrases listed in Textbox 1 were used to identify articles in each database.

XSL•FO
RenderX

Textbox 1. Queries used to retrieve records.

EMBASE (Excerpta Medica Database)

- #1 'reinforcement learning'
- #2 'intensive care unit' OR 'critical care' OR 'ICU'
- #1 AND #2

Google Scholar

- (conference OR journal) AND ("intensive care unit" OR "critical care" OR ICU) AND "reinforcement learning" -survey -reviews -reviewed -news

IEEE (Institute of Electrical and Electronics Engineers)

- (("Full Text Only": "reinforcement learning") AND "Full Text Only": "intensive care units") OR (("Full Text Only": "reinforcement learning") AND "Full Text Only": "critical care")

MEDLINE (Medical Literature Analysis and Retrieval System Online)

- multifield search=reinforcement learning, critical care, intensive care

PubMed

- ("reinforcement learning") AND (("ICU") OR ("critical care") OR ("intensive care unit") OR ("intensive care"))

ScienceDirect

- "reinforcement learning" AND ("critical care" OR "intensive care" OR "ICU")

Web of Science

- ALL=(intensive care unit OR "critical care" OR "ICU") AND ((ALL=("reinforcement learning")) AND LANGUAGE: (English))

Inclusion Criteria

To be eligible for inclusion in this review, the primary requirement was that the article needed to focus on the implementation, evaluation, or use of an RL algorithm to process or analyze patient information (including simulated data) in an ICU setting. Papers published from January 1, 2010, to October 19, 2019 were selected. General review articles and articles not published in English were excluded. Only papers that discussed sufficient details on the data, method, and results were included in this review.

Data Synthesis

Data were manually extracted from the articles included in the review. A formal quality assessment was not conducted, as relevant reporting standards have not been established for articles on RL. Instead, we extracted the following characteristics from each study: the purpose of the study, data source, number of patients included, main method, evaluation metrics, and related outcomes. The final collection of articles was divided into categories to assist reading according to their application type in the ICUs.

Results

Selection Process and Results Overview

The selection process of this review was demonstrated using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (Figure 2). From the full text of 269 distinct articles, an independent assessment for eligibility was performed by 2 authors (SL and MF). Disagreements were discussed to reach consensus. During the full-text review, 249 articles were excluded, and 21 articles were eventually included. The reasons for exclusion during the review process are outlined in Table 1.

XSL•FO
RenderX

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the search strategy.

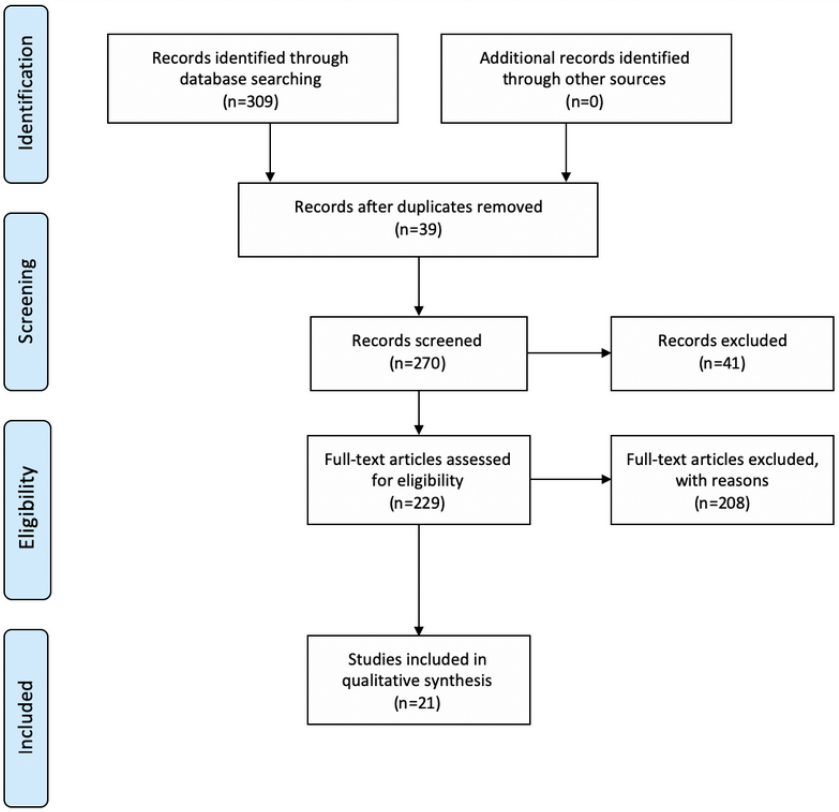


Table 1. Exclusion criteria used to exclude papers.

Criterion number	Exclusion criteria	Justification	Excluded articles, n
1	Duplicates	The papers have duplicate titles	39
2	Not a research article	The papers were blog articles, reports, comments, or views	23
3	Not written in English	The papers were not written in English	6
4	Review	The papers were review articles regarding general methods on big data, deep learning, and clinical applications	12
5	Not applied in the field of critical care	The papers did not focus on applications in critical care or intensive care	92
6	Not using RL[a] as the approach in critical care	The papers discussed issues in the critical care setting, but not using RL as an approach	115
7	No clear description of the method and result	The methods and results were not clearly described and thus not qualified for this review	1

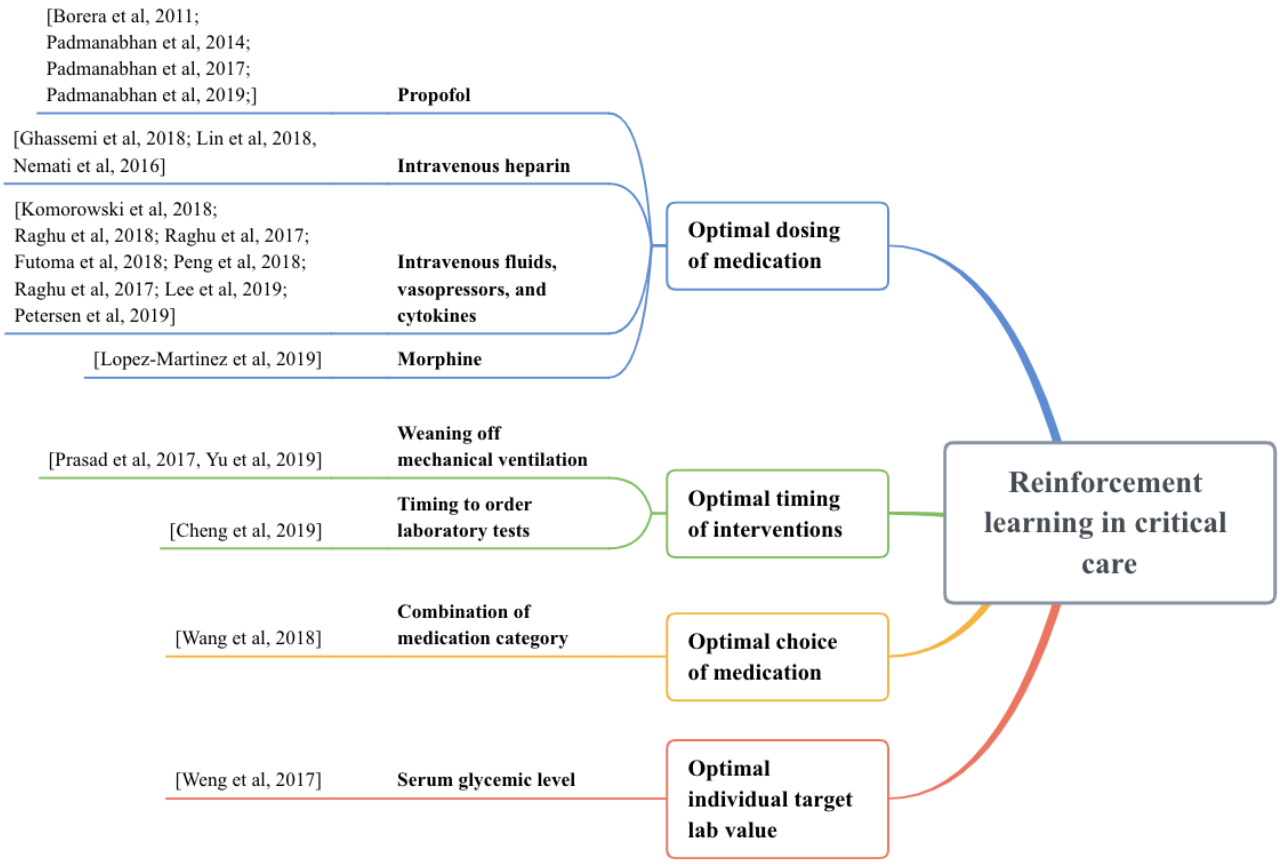
[a]RL: reinforcement learning.

In this section, we organized the reviewed articles into 4 categories, which reflect clinically relevant domains: (1) optimal individualized target laboratory value; (2) optimal choice of medication; (3) optimal timing of an intervention; and (4) optimal dosing of medication.

We plotted the number of articles reviewed by their category and year of publication in Figure 3. We found that the majority of the papers were published in the past 3 years (n=17), indicating an increasing trend of applying RL-based approaches to assist physicians in decision making in critical care. In each of the 4 categories, we further organized the articles into subgroups based on their clinical questions (Figure 3). The figure shows that most of the applications used RL to find optimal drug dosing (n=16) [6,21,29-42], followed by the timing of an intervention (n=3) [43-45]. Only a few applications were looking at the individualized laboratory value (n=1) [46] and the optimal choice of medication (n=1) [47].

XSL•FO
RenderX

Figure 3. Mapping of reinforcement learning studies in critical care by application type.



Next, we discuss the details for each category with the methods and outcomes for each application. In particular, we further grouped the studies based on specific medication or treatment type in categories 3 and 4 to assist readers. A summary of all study details is found in Multimedia Appendix 2.

Optimal Individualized Target Laboratory Value

Even after decades of routine use of laboratory value ranges, reference standards may need to be reconsidered, especially for individual patients [48]. Personalized targets for laboratory values in ICU patients could account for disease severity, comorbidities, and other patient-specific differences. Weng et al [46] tried to identify individualized targeted blood glucose levels as a reference for physicians. They applied an RL-based approach, *policy iteration*, to learn the target glycemic range at an hourly interval for severely ill patients with sepsis using real ICU data. Their approach was tested using the Medical Information Mart for Intensive Care III (MIMIC III), a large, publicly available ICU database [49]. MIMIC III contains information for hospital admissions of 43,000 patients in critical care units during 2001 and 2012, from which the authors extracted hourly data for 5565 patients with sepsis.

Weng et al [46] constructed their RL model as follows: First, they represented the patients' states from 128 variables. These variables included patient demographics, comorbid conditions, vital sign changes, and laboratory value changes. They used a spare autoencoder [50] to reduce the high dimensionality of the raw features (128 dimensions) to only 32 dimensions so that the RL model could be trained more efficiently with limited observational data. Second, they chose to act upon 1 of 11 discrete ranges of serum glucose at each time step. Third, they designed the reward function so that the RL agent could recommend an hourly target glucose level to optimize long-term survival. A positive 100 was assigned to the end state if patients survived 90 days after admission, and a negative 100 was assigned if the patients died. For each state-action pair, the *value* of the pair was iteratively estimated using the reward from the training data.

To understand how the reward value was related to mortality, the authors assigned values to discrete buckets using separate test data. In each value bucket, if the state-action pair is part of a trajectory where a patient died, a label of 1 was assigned to that bucket; otherwise, a label of 0 was assigned. After assigning all the state-action pairs from the test data with the labels in the corresponding value bucket, the mortality rate could be estimated for each value bucket. The authors plotted the estimated mortality rate with respect to the value-buckets and found an inverse relationship between them, where the highest value was associated with the lowest mortality. This result suggested that the learnt value represented the relationship between the state-action pair and mortality and that the learnt value of the state-action pairs from training data was validated on the test data.

To validate the RL policy, the author calculated the frequency of state transitions from the training data and generated new

XSL•FO
RenderX

trajectories. Starting from the observed state in the test data, the RL policy would recommend an action with the highest value, and the subsequent state was estimated with the transition probability. By averaging the value for all state-action pairs in the simulated trajectory, the mortality for simulated trajectories could be estimated by mapping this value in the mortality-value plot. Compared with the actual mortality rate in the test data, the author claimed that if physicians could control patients' hourly blood glucose levels within the range recommended by the RL model, the estimated 90-day mortality would be lowered by 6.3% (from 31% to 24.7%).

Optimal Choice of Medications

Apart from some clinical decision support systems, commonly used systems such as computerized prescriber order entry and bar-coded medication administration lack personalized recommendations to optimize medication effectiveness and minimize side effects [51]. Wang et al [47] applied a deep learning network based on RL to exploit medication recommendations with a data-driven strategy. Their approach accounted for individual patient demographics, laboratory values, vital signs, and diagnoses from the MIMIC III database. They selected the top 1000 out of 4127 medications and the top 2000 out of 6695 diseases (represented by the International Classification of Diseases, Ninth Revision codes), which covered 85.4% of all medication records and 95.3% of all diagnosis records, respectively. To reduce the problem complexity, the authors further categorized the 1000 medications into 180 drug categories using anatomical therapeutic chemical codes and aggregated patients' drug prescriptions into 24-hour windows.

The authors defined RL action as the medication combinations from the 180 drug categories. They adopted an actor-critic RL agent that suggested a daily medication prescription set, and aimed to improve patients' hospital survival. The details of the actor-critic RL algorithm are explained in Multimedia Appendix 1 [28]. For each patient's ICU day, the *actor* network would recommend one medication combination by considering state variables such as demographics, laboratory results, and vital signs. A reward value of positive 15 would be given to the end state if a patient survived until hospital discharge and negative 15 if the patient died. The reward was designated as 0 for all other time steps. To counterbalance the *actor* network, the *critic* network was applied to evaluate the consistency of actual physician prescriptions and the RL agent's recommendations. The net effect of the actor-critic RL agent was to optimize the long-term outcomes of patients (hospital mortality) while minimizing deviations of RL-recommended actions from actual prescription patterns. In addition to the actor-critic network, the authors also applied long short-term memory [52] to represent a patient's current state by incorporating the long sequence of all historical states. Wang et al [47] suggested that hospital mortality would be reduced by 4.4% if clinicians adhered to the RL agent's recommendations.

Optimal Timing of Intervention

Weaning of Mechanical Ventilation

Mechanical ventilation (MV) is a life-saving treatment applied in approximately a third of all critically ill patients [53]. Prematurely discontinuing MV (premature weaning) and excessively prolonged MV (late weaning) are both associated with higher mortality [54]. The best time to wean may be uncertain [55].

To optimize the timing of ventilation discontinuation, Prasad et al [43] applied the RL-based FQI (the details of the FQI algorithm are explained in Multimedia Appendix 1 [28]) on the MIMIC III database for all patients who were kept under ventilator support for more than 24 hours and extracted their records every 10 min from ICU admission to discharge. Patient states included a number of factors that could affect extubation, such as demographics, pre-existing conditions, comorbidities, and time-varying vital signs. The action for the ventilation setting was binary, that is, for each 10-min time step, the RL agent needed to decide whether the ventilation should be set on (continued MV) or off (weaned from MV). For reward design, Prasad et al [43] followed an existing weaning protocol from the Hospital of University of Pennsylvania. They assigned reward values to the RL agent at each time step according to 3 major considerations: (1) the RL agent should penalize each additional hour spent on the ventilator, (2) the RL agent should be assigned a positive reward value to a weaning action if the patient's vital signs and laboratory results were steady and within normal ranges after extubation, and (3) there was no reward value for failed spontaneous breathing trial or for reintubation after the first extubation. For RL policy evaluation, the authors calculated the proportion of weaning actions from the RL policy, referencing the total number of weaning actions from the clinician's policy at each time step, and calculated the overall consistency of weaning transitions. The recommend actions from the RL agent could match 85% of those from clinicians. The authors categorized the degree of consistency into 5 bins, and plotted the distribution of the number of reintubations with respect to the discrete consistency levels. Their results showed that when the consistency was high, vital sign fluctuations were fewer, laboratory results were more in-range, and reintubations were minimized.

Yu et al [45] studied the same clinical issue as Prasad et al [43] and used the same data set, but designed a different reward function using inverse RL. The inverse RL model directly learnt reward mapping from data for each state-action pair and inferred what clinicians would wish to achieve as a reward. Similar to Prasad et al [43], the RL recommendations by Yu et al [45] were associated with shorter weaning times and fewer reintubations compared with clinician decision making.

Timing to Order Laboratory Tests

The timing of ordering a laboratory test can be challenging. Delayed testing would lead to continued uncertainty over the patient's condition and possible late treatment [56]. However, excessively early ordering of laboratory tests can cause unnecessary discomfort to the patient, increase the risk of anemia, and increase health care cost.

Cheng et al [44] applied the FQI method to find the optimal timing for ordering laboratory tests among patients with sepsis in the MIMIC III data set. They examined the timing of 4 types of laboratory tests: white blood cell count (WBC), creatinine, blood urea nitrogen (BUN), and lactate. They sampled the

XSL•FO
RenderX

patients' data at hourly intervals and constructed the state of a patient by considering the predictive variables of severe sepsis or acute kidney failure, including respiratory rate, heart rate, mean blood pressure, temperature, creatinine, BUN, WBC, and lactate. The missing values were predicted by a multioutput Gaussian process [57,58]. In their RL model, they chose to design the reward function with the combination of 4 factors: (1) a positive reward should be given only if the ordering of test was necessary, while penalizing over or under ordering; (2) the RL agent should be encouraged to order laboratory tests when there was a sudden change in laboratory results or vital signs; (3) negative reward should be given if the laboratory results were similar to the last measurements (no information gain); (4) a penalty would be added to a reward whenever a test was ordered, to reflect the testing cost. Their RL agent, compared with clinicians, was able to reduce the number of laboratory tests by 27% for lactate and 44% for WBC, while maintaining high information gain.

Optimal Dosing of a Drug

Recommendations for dosing regimens in ICU patients are often extrapolated from clinical trials in healthy volunteers or noncritically ill patients. This extrapolation assumes similar drug behavior (pharmacokinetics and pharmacodynamics) in the ICU and other patients or healthy volunteers. However, it is well known that many drugs used in critically ill patients may have alterations in pharmacokinetic and pharmacodynamic properties because of pathophysiological changes or drug interactions [59]. Therefore, critically ill patients bring unique challenges in drug dosing.

Dosing of Propofol

Critically ill patients in ICUs often require sedation to facilitate various clinical procedures and to comfort patients during treatment. Propofol is a widely used sedative medication [60], but titration of propofol is challenging, and both over sedation and under sedation can have adverse effects [32]. Of the studies reviewed, 6 studies have focused on applying RL to determine the optimal dosage for propofol while maintaining the physiological stability of the patient. The bispectral index (BIS) was used to monitor sedation level and to determine the effect of propofol.

Borera et al [29] was the first to apply RL to a pharmacokinetic model [61] to describe the time-dependent distribution of propofol in human surgical patients. The RL agent was a neural network aimed at optimizing the propofol dose to achieve the target BIS value. The patient's state and state transition were modeled using a mathematical pharmacokinetic model with predefined parameters such as the concentration at half maximal effect of BIS, degree of nonlinearity of BIS, and time-lag coefficient to estimate the BIS value for simulated patients. The action was a discrete range of propofol infusion rate. The reward function was the error rate between the target BIS value and the current simulated BIS value, where a larger negative reward was given when the current simulated BIS value was further away from the predefined target value. They measured the performance of the RL agent by looking at the time to reach the target BIS value (steady time). The evaluation was conducted on 1000 simulated patients. On average, the steady time was 3.25 min for the BIS value to reach target.

To ensure patient safety, propofol dosing should consider the concurrent stability of vital parameters. For instance, Padmanabhan et al [30] chose mean arterial pressure (MAP) as the secondary control variable. The authors combined the error rates for both BIS and MAP when designing the reward. The target for the RL agent was to infuse propofol so that the target BIS would be reached in a short time, whereas MAP was kept within a desired range. In subsequent studies, Padmanabhan et al [31,32] modified their methods with different RL training algorithms (Q-learning and policy iteration). In all their studies, the RL agent was able to suggest accurate propofol doses and achieve target BIS values within a few minutes.

In contrast to fixed pharmacokinetic models in the RL model environment, Yu et al [45] applied FQI and Bayesian inverse RL on the MIMIC III database. They considered patients' demographic characteristics, pre-existing conditions, comorbidities, and time-varying vital signs to construct the state of the patient. Their inverse RL model interpreted clinician preference as a reward for different patient states. The learned reward function from the inverse RL model suggested that clinicians may pay more attention to patients' cardiorespiratory stability rather than oxygenation when making decisions about propofol dosage.

Dosing of Intravenous Heparin

Anticoagulant agents are often used to prevent and treat a wide range of cardiovascular diseases. Heparin is commonly used in critical care [62], yet its precise dosing is complicated by a narrow therapeutic window. Overdosing of heparin results in bleeding whereas under dosing risks clotting. To guide heparin dosing, activated partial thromboplastin time (aPTT) is often used as a measure of the anticoagulant effect of heparin.

Nemati et al [6] applied FQI with a neural network to optimize and individualize heparin dosing. Their study was conducted on the MIMIC II database, with the reward function based on aPTT levels following heparin dosing [63]. The reward to the RL agent will be high if the aPTT value is between 60 and 100 seconds. After training, they plot the state-action value with respect to the level of consistency between the RL policy and clinician practice. Their results showed that, on average, following the recommendations of the RL agent resulted in higher state-action values.

Ghassemi et al [33] and Lin et al [34] focused on a personalized optimal heparin dosing using different RL algorithms. In addition to the MIMIC III data set, Lin et al [34] applied an actor-critic network on the Emory Healthcare data set from Emory University. For RL policy evaluation, Lin et al [34] regressed the discordance between RL policy and physician practice over the number of clotting and bleeding complications, adjusting for covariates such as history of clot or bleed, weight, age, and sequential organ failure assessment score. The regression coefficient suggested that following the RL agent's recommendations would have likely resulted in improved clinical outcomes with a reduced number of clotting and bleeding complications.

XSL•FO
RenderX

Intravenous Fluids, Vasopressors, and Cytokine Therapy for Treating Sepsis

Sepsis is the third leading cause of death and is expensive to treat [64]. Besides antibiotics and source control, challenges remain with the use of intravenous (IV) fluids to correct hypovolemia and administration of vasopressors to counteract sepsis-induced vasodilation. Raghu et al [36] suggested a data-driven RL approach to recommend personalized optimal dosage for IV fluids and vasopressors to improve hospital mortality. Their RL model was double DQN with dueling, which can minimize the overestimation problem of previous Q-learning models. The details of the Q-learning and double DQN algorithms are explained in Multimedia Appendix 1 [28]. The authors considered patients' demographics, laboratory values, vital signs, and intake/output events as state features in the RL model. Action was designed as a combination of 5 discrete bins for IV fluid dosing and 5 bins for vasopressor dosing to treat patients with sepsis. The reward was issued at the terminal time step of the patient's trajectory, with a positive reward if the patient survived. Data were extracted from the MIMIC III database for all patients who fulfilled sepsis-3 criteria [65]. For policy evaluation, Raghu et al [36] plotted the estimated hospital mortality with respect to the difference between dosages recommended by the RL agent and by clinicians. The plot showed that the mortality was lowest when there was no discrepancy between RL policy and physician decision making. Six other groups of researchers also focused on the same research question and applied various RL algorithms with slightly different designs of the state space, reward function, and evaluation metrics [21,35,37-40]. The findings from these studies all suggest that the RL agent would be able to learn from the data and if physicians followed the RL policy, the estimated hospital mortality could be improved.

Among the aforementioned studies, Komorowski et al [21] were the pioneers of applying RL in the ICU, using data from patients with sepsis in the MIMIC III database. They inferred a patient's health status using an array of inputs, which included demographics, vital signs, laboratory tests, illness severity scores, medications, procedures, fluid intake and output, physician notes, and diagnostic coding. Patient data were aggregated and averaged every 4 hours to represent patient states. Using a k-means algorithm, these patient states were then simplified into 750 discrete mutually exclusive clusters. A sequence of these clustered states would describe a particular patient's trajectory. The authors estimated the state transition probability by counting how many times each transition was observed and converted the counts to a stochastic matrix. This transition matrix contained the probability for each patient going to a new state, given a previous action taken in the current state. The entire trajectory of a patient's state can be estimated using the transition matrix. The authors applied a policy iteration RL algorithm that learnt the optimal dosing policy for IV fluids and vasopressors to maximize the probability of 90-day survival.

Nevertheless, the study by Komorowski et al [21]. had several limitations. First, their study only considered fluid and vasopressor management, ignoring other important treatments such as source control, correction of hypovolemia, and management of secondary organ failures [21]. Second, 90-day mortality is affected by factors outside of the ICU, which the study did not take into account. Third, clinical decision making considers both short-term outcomes (eg, physiological stability) and long-term outcomes (eg, kidney failure or mortality), but the study only considered mortality as the single goal for training the RL algorithm [66]. Fourth, discretizing patient health status into discrete clusters loses data granularity and may limit the ability to detect changes in patient status. These limitations also occur in other studies, which we will elaborate in the Discussion section.

Other than using IV fluids and vasopressors for treating sepsis. Petersen et al [42] investigated cytokine therapy using the deep deterministic policy gradient [67] method. The details of the policy gradient RL algorithm are explained in Multimedia Appendix 1 [28]. They evaluated the RL model by using an agent-based model, the innate immune response agent-based model [68], that simulated the immune response to infection. The RL policy was able to achieve a very low mortality rate of 0.8% over 500 simulated patients, and suggested that personalized multicytokine treatment could be promising for patients with sepsis.

Dosing of Morphine

Critically ill patients may experience pain as a result of disease or certain invasive interventions. Morphine is one of the most commonly used opioids for analgesia [69]. Similar to sedation, the dosing of analgesia is subject to uncertainty. Lopez-Martinez et al [41] collected data for patients who had at least one pain intensity score and at least one dose of IV morphine in the MIMIC III database. They applied double DQN with dueling as their RL model and constructed the state space to be continuous with features including the patient's self-reported pain intensity and their measured physiological status. The action was a choice of 14 discrete dosing ranges of IV morphine. The reward was determined by considering both the patients' cardiorespiratory stability and their pain intensity. The highest reward was given when pain was absent and both heart rate and respiration rate were within the acceptable range. By comparing the RL policy with physicians' choices, Lopez-Martinez et al [41] found that RL policy tended to prescribe higher doses of morphine. This result was consistent with previous studies: continuous dosing provided similar or even better pain relief with no increase in acute adverse effects [70,71].

Discussion

Principal Findings

Our comprehensive review of the literature demonstrates that RL has the potential to be a clinical decision support tool in the ICU. As the RL algorithm is well aligned with sequential decision making in ICUs, RL consistently outperformed physicians in simulated studies. Nonetheless, challenges regarding RL system design, evaluation metrics, and model choice exist. In addition, all current applications have focused on using retrospective data sets to derive treatment algorithms and require prospective validation in authentic clinical settings.

XSL•FO
RenderX

RL System Design

The majority of applications were similar in their formulation of the RL system design. The state space is usually constructed by features including patient demographics, laboratory test values, and vital signs, whereas some studies applied encoding methods to represent the state of the patients instead of using raw features. The action space was very specific to each application. For instance, in terms of the dosing category, the action space would be discretized ranges of medication dosage. For other categories, such as timing of an intervention, the action space would be the binary indicator of an intervention for each time step. The number of action levels differed among the studies. For some studies, the action levels could be as many as a dozen or a hundred (eg, optimal medication combination), whereas for other studies, the action levels were limited to only 2 (eg, on/off MV). The design of the reward function is central to successful RL learning. Most of the reward functions were designed a priori with guidance from clinical practice and protocols, but 2 studies [40,45] managed to directly learn the reward function from the data using inverse RL.

Evaluation Metrics

The only metric that matters is if the adoption of an RL algorithm leads to improvement in some clinical outcomes. Most studies calculated the estimated mortality as the long-term outcome and drew plots to show the relationship between the estimated mortality versus the learnt value of patients' state-action trajectories, where the higher value function was associated with lower mortality. The RL agent would provide treatment suggestions for those actions with higher values, thus leading to a lower estimated mortality. Estimated mortality is a popular metric for RL policy evaluation. However, the problem with the estimated mortality is that it is calculated from simulated trajectories with observational data, and may not be the actual mortality.

Mortality is not always the most relevant and appropriate outcome measure. For instance, in the study by Weng et al [46], they tried to identify individualized targeted blood glucose levels as a reference for physicians. In their study, 90-day mortality was used to evaluate the RL policy. However, a more relevant measure could be considered, such as short-term changes in the blood glucose level, physiological stability, and development of complications.

Several studies that focused on propofol titration have considered BIS as the evaluation metric to monitor the sedation level and hence to determine the effect of propofol. Although BIS monitoring is fairly objective, assessing sedation is usually performed by health care providers with clinically validated behavioral assessment scales such as the Richmond Agitation-Sedation Scale score [72]. In addition, EEG-based technologies, such as BIS and M-entropy, have been validated more in the operating room than in the ICU [73]. Furthermore, BIS cannot be used as the sole monitoring parameter for sedation, as it is affected by several other factors, including the anesthetic drugs used, muscle movement, or artifacts from surgical equipments [74].

To date, there has been no prospective evaluation of an RL algorithm. Moreover, the observational data itself may not truly reflect the underlying condition of patients. This is known as the partially observable MDP [75] problem, where we are only able to represent a patient's state by the observed physiological features, which are solved by mathematical approximation.

Model Choice

FQI and DQN seem to be the top RL approaches among the reviewed studies. FQI is not a deep learning–based RL model, which guarantees convergence for many commonly used regressors, including kernel-based methods and decision trees. On the other hand, DQN leverages the representational power of DNNs to learn optimal treatment recommendations, mapping the patient state-action pair to the value function. Neural networks hold an advantage over tree-based methods in iterative settings in that it is possible to simply update the network weights at each iteration, rather than rebuilding the trees entirely.

Both FQI and DQN are off-policy RL models. Off-policy refers to learning about one way of behaving from the data generated by another way of selecting actions [76]. For instance, an off-policy RL model tries to train a policy X to select actions in each step, but it estimates the Q-values from state-action pairs where the action was chosen by following another policy Y. In contrast to off-policy learning, on-policy learning uses the same policy X to choose actions and to evaluate the returns in each step during training. Most of the included studies adopted off-policy RL models because the RL models aim to learn policy X from the data, which was generated by following real actions of physicians (policy Y). The data generated by policy Y is the actual physicians' policy, where the RL models try to learn and improve from. This is the fundamental idea of applying off-policy RL models.

In addition, both FQI and DQN are value-based RL models that aim to learn the value functions. In value-based RL, a policy can be derived by following the action with the highest value at each time step. Another type of RL is called policy-based RL, which aims to learn the policy directly without worrying about the value function. Policy-based methods are more useful in continuous space. When the data volume is insufficient to train a DQN model, the DQN is not guaranteed to achieve a stable RL policy. As there is an infinite number of actions or states to estimate the values for, value-based RL models are too computationally expensive in the continuous space. However, policy-based RL models demand more data samples for training. Otherwise, the learned policy is not guaranteed to converge to an optimal one. Both value-based and policy-based RL models can be grouped in a more general way as *model-free* RL. Here the word *model-free* means the environment is unknown to an agent. The RL agent makes use of the trajectories generated from the environment, rather than explicitly knowing the rule or the transition probability. In contrast to model-free RL, *model-based* RL requires the agent to know the transition probability for all the state-action combinations explicitly and hence impractical as the state space and action space grow. In the critical care context, patients' conditions and prognosis are very complex to apply model-based RL because we are not exactly sure about the probability of all state transitions. In

XSL•FO
RenderX

addition, most studies in critical care could only use limited retrospective data to train the model offline. Therefore, we found that most of the studies have applied a value-based RL model to utilize the available observational data.

Common Data Sets

We found that 71% (15/21) of applications utilized the MIMIC II or MIMIC III database to conduct their experiments. We conjecture that such popularity might be due to public availability and high quality of MIMIC data. However, data collected from a single source may introduce potential bias to the research findings. There are inherent biases in the medical data sets obtained at various institutions due to multiple factors, including operation strategy, hospital protocol, instrument difference, and patient preference. Therefore, the RL models trained on a single data set, regardless of the data volume, cannot be confidently applied to another data set. The findings from the reviewed articles may not be generalizable to other institutions and populations. In addition to the MIMIC database, one of the studies also utilized the eICU Research Institute (eRI) database to test their RL model [77]. The eRI database has a larger volume of data compared with the MIMIC database, and it is also publicly available. We suggest that future applications could cross-validate their models on both the MIMIC and eRI databases. In addition, all current applications have focused on using retrospective data sets to derive treatment algorithms and require prospective validation in authentic clinical settings.

Strengths and Limitations of This Study

The strengths of this paper include the comprehensive and extensive search for all available publications that applied RL as an approach in the critical care context. Nonetheless, we acknowledge the limitations. We included papers (eg, those on arXiv) that have not been peer-reviewed *before* publication but these papers have undergone a postpublication peer review. According to the search phrases applied in this review, we may miss out certain papers that applied RL in critical care, but did not specify the phrase *intensive care* nor *ICU* in their full text papers.

Challenges and Future Directions

A number of challenges must be overcome before RL can be implemented in a clinical setting. First, it is important to have a meaningful reward design. The RL agent would be vulnerable in case of reward misspecification, and might not be able to produce any meaningful treatment suggestion. Inverse RL can be an alternative to a priori–specified reward functions. However, inverse RL assumes that the given data represent the experts' demonstrations and the recommendations from the data were already optimal; these may not be true.

Second, medical domains present special challenges with respect to data acquisition, analysis, interpretation, and presentation of these data in a clinically relevant and usable format. Addressing the question of censoring in suboptimal historical data and explicitly correcting for the bias that arises from the timing of interventions or dosing of medication is crucial to fair evaluation of learnt policies.

Third, another challenge for applying the RL model in the clinical setting is exploration. Unlike other domains such as game playing, where one can repeat the experiments as many times, in the clinical setting, the RL agent has to learn from a limited set of data and intervention variations that were collected offline. Using trial and error to explore all possible scenarios may conflict with medical ethics, thereby limiting the ability of the RL agent to attempt new behaviors to discover ones with higher rewards and better long-term outcomes.

In comparison with other machine learning approaches, there is an absence of acceptable performance standards in RL. This problem is not unique to RL but seems harder to address in RL compared with other machine learning approaches, such as prediction and classification algorithms, where accuracy and precision recall are more straightforward to implement. However, it is worth noting that RL has a distinct advantage over other machine learning approaches, that one can choose which outcome to optimize by specifying the reward function. This provides an opportunity to involve patient preferences and shared decision making. This becomes more relevant when learned policies change depending on the reward function. For example, an RL algorithm that optimizes survival may recommend a different set of treatments versus an RL algorithm that optimizes neurologic outcome. In such situations, patient preference is elicited to guide the choice of the RL algorithm.

RL has the potential to offer considerable advantages in supporting the decision making of physicians. However, certain key issues need to be addressed, such as clinical implementation, ethics, and medico-legal limitations in health care delivery [78]. In fact, any machine learning model would need to address these limitations carefully to serve as truly effective tools. In clinical practice, the RL models need to be refined iteratively throughout the time to include newly generated data from electronic health systems in hospitals, and the model must produce robust results for physicians to interpret and understand. Besides, patients' understanding and willingness to use the RL model as a supporting tool in their care would be another important consideration. Another important ethical consideration would be the liability in case of medical error when the RL model recommendation differs from the physician. It has an impact on the autonomy of both the physician and patient. The problem of medical error works in both ways when there is a poor outcome: (1) if the physician follows the RL model recommendation, can the clinician then blame the model and the personnel who maintain the model; (2) if the clinician does not follow the RL model recommendation, can the clinician then be said to have made the wrong decision and be penalized.

Possible directions for future work include (1) modeling the RL environment as a partially observable MDP, in which observations from the data are mapped to some state space that truly represents patients' underlying well-being; (2) extending the action space to be continuous, suggesting more precise and practical treatment recommendations to physicians; and (3) improving the interpretability of the RL models so that physicians can have more confidence in accepting the model results. With further efforts to tackle these challenges, RL methods could play a crucial role in helping to inform patient-specific decisions in critical care.

XSL•FO
RenderX

Conclusions

In this comprehensive review, we synthesized data from 21 articles on the use of RL to process or analyze retrospective data from ICU patients. With the improvement of data collection and advancement in reinforcement learning technologies, we see great potential in RL-based decision support systems to optimize treatment recommendations for critical care.

Acknowledgments

SL was funded by the National University of Singapore Graduate School for Integrative Sciences and Engineering Scholarship. This research was supported by the National Research Foundation Singapore under its AI Singapore Programme (award no. AISG-GC-2019-002), the National University Health System joint grant (WBS R-608-000-199-733), and the National Medical Research Council health service research grant (HSRG-OC17nov004).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Introduction to reinforcement learning.
[PDF File (Adobe PDF File), 617 KB - jmir_v22i7e18477_app1.pdf]

Multimedia Appendix 2
Summary of study characteristics.
[PDF File (Adobe PDF File), 167 KB - jmir_v22i7e18477_app2.pdf]

References

1. Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. Stat Med 2012 Jul 30;31(17):1887-1902 [FREE Full text] [doi: 10.1002/sim.4512] [Medline: 22438190]
2. Chen Z, Marple K, Salazar E, Gupta G, Tamil L. A physician advisory system for chronic heart failure management based on knowledge patterns. Theor Pract Log Prog 2016 Oct 14;16(5-6):604-618. [doi: 10.1017/S1471068416000429]
3. Hannes K, Leys M, Vermeire E, Aertgeerts B, Buntinx F, Depoorter A. Implementing evidence-based medicine in general practice: a focus group based study. BMC Fam Pract 2005 Sep 9;6:37 [FREE Full text] [doi: 10.1186/1471-2296-6-37] [Medline: 16153300]
4. Hutchinson A, Baker R. Making use of guidelines in clinical practice. Br Med J 1999 Oct 16;319(7216):1078 [FREE Full text] [doi: 10.1136/bmj.319.7216.1078] [Medline: 10521225]
5. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf 2013 Sep;9(3):122-128. [doi: 10.1097/PTS.0b013e3182948a69] [Medline: 23860193]
6. Nemati S, Ghassemi MM, Clifford GD. Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:2978-2981. [doi: 10.1109/EMBC.2016.7591355] [Medline: 28268938]
7. Zhang Z, Hong Y, Liu N. Scientific evidence underlying the recommendations of critical care clinical practice guidelines: a lack of high level evidence. Intensive Care Med 2018 Jul;44(7):1189-1191. [doi: 10.1007/s00134-018-5142-8] [Medline: 29564478]
8. Laffey JG, Kavanagh BP. Negative trials in critical care: why most research is probably wrong. Lancet Respir Med 2018 Sep;6(9):659-660. [doi: 10.1016/S2213-2600(18)30279-0] [Medline: 30061048]
9. Burke AE, Thaler KM, Geva M, Adiri Y. Feasibility and acceptability of home use of a smartphone-based urine testing application among women in prenatal care. Am J Obstet Gynecol 2019 Nov;221(5):527-528. [doi: 10.1016/j.ajog.2019.06.015] [Medline: 31300161]
10. Laserson J, Lantsman CD, Cohen-Sfady M, Tamir I, Goz E, Brestel C, et al. TextRay: Mining Clinical Reports to Gain a Broad Understanding of Chest X-Rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2018 Presented at: MICCAI'18; September 16-20, 2018; Granada, Spain. [doi: 10.1007/978-3-030-00934-2_62]
11. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
12. Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys 2019 Jan;46(1):370-381. [doi: 10.1002/mp.13271] [Medline: 30383300]
13. Dagan N, Elnekave E, Barda N, Bregman-Amitai O, Bar A, Orlovsky M, et al. Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization. Nat Med 2020 Jan;26(1):77-82. [doi: 10.1038/s41591-019-0720-z] [Medline: 31932801]

XSL•FO
RenderX

14. Watanabe AT, Lim V, Vu HX, Chim R, Weise E, Liu J, et al. Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. J Digit Imaging 2019 Aug;32(4):625-637 [FREE Full text] [doi: 10.1007/s10278-019-00192-5] [Medline: 31011956]
15. Montague PR. Reinforcement learning: an introduction, by Sutton, RS and Barto, AG. Trends Cogn Sci 1999 Sep;3(9):360. [doi: 10.1016/S1364-6613(99)01331-5]
16. Kiumarsi B, Vamvoudakis KG, Modares H, Lewis FL. Optimal and autonomous control using reinforcement learning: a survey. IEEE Trans Neural Netw Learn Syst 2018 Jun;29(6):2042-2062. [doi: 10.1109/TNNLS.2017.2773458] [Medline: 29771662]
17. Howard RA. Dynamic Programming and Markov Process. New York, USA: MIT Press and Wiley; 1960.
18. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. arXiv preprint 2013:- epub ahead of print(1312.5602) [FREE Full text]
19. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016 Jan 28;529(7587):484-489. [doi: 10.1038/nature16961] [Medline: 26819042]
20. Ng A, Coates A, Diel M, Ganapathi V, Schulte J, Tse B, et al. Autonomous inverted autonomous helicopter flight via reinforcement learning. In: Experimental Robotics IX. New York, USA: Springer; 2006:363-372.
21. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
22. Riedmiller M. Neural Fitted Q Iteration – First Experiences with a Data Efficient Neural Reinforcement Learning Method. In: Proceedings of the European Conference on Machine Learning. 2005 Presented at: ECML'05; October 3-7, 2005; Porto, Portugal URL: https://doi.org/10.1007/11564096_32 [doi: 10.1007/11564096_32]
23. van Hasselt H, Guez A, Silver D. Deep Reinforcement Learning With Double Q-learning. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016 Presented at: AAAI'16; February 12-17, 2016; Phoenix, Arizona, USA. [doi: 10.5555/3016100.3016191]
24. Mnih V, Puigdomenech A, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. Arxiv 2016:- epub ahead of print(1602.01783) [FREE Full text]
25. Doya K, Samejima K, Katagiri K, Kawato M. Multiple model-based reinforcement learning. Neural Comput 2002 Jun;14(6):1347-1369. [doi: 10.1162/089976602753712972] [Medline: 12020450]
26. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: a brief survey. IEEE Signal Process Mag 2017 Nov;34(6):26-38. [doi: 10.1109/MSP.2017.2743240]
27. Wiering M, van Otterlo M, editors. Reinforcement Learning: State-of-the-Art. Berlin, Heidelberg: Springer-Verlag; 2012.
28. Siqi L. Reinforcement-Learning. GitHub. URL: https://github.com/nus-mornin-lab/Reinforcement-Learning [accessed 2020-01-01]
29. Borera EC, Moore BL, Doufas AG, Pyeatt LD. An Adaptive Neural Network Filter for Improved Patient State Estimation in Closed-Loop Anesthesia Control. In: 23rd International Conference on Tools with Artificial Intelligence. 2011 Presented at: ICTAI'11; November 7-9, 2011; Boca Raton, FL, USA. [doi: 10.1109/ictai.2011.15]
30. Padmanabhan R, Meskin N, Haddad WM. Closed-loop Control of Anesthesia and Mean Arterial Pressure Using Reinforcement Learning. In: Symposium on Adaptive Dynamic Programming and Reinforcement Learning. 2014 Presented at: ADPRL'14; December 9-12, 2014; Orlando, FL, USA. [doi: 10.1109/ADPRL.2014.7010644]
31. Padmanabhan R, Meskin N, Haddad WM. Reinforcement Learning-Based Control for Combined Infusion of Sedatives and Analgesics. In: 4th International Conference on Control, Decision and Information Technologies. 2017 Presented at: CoDIT'17; April 5-7, 2017; Barcelona, Spain. [doi: 10.1109/codit.2017.8102643]
32. Padmanabhan R, Meskin N, Haddad WM. Optimal adaptive control of drug dosing using integral reinforcement learning. Math Biosci 2019 Mar;309:131-142. [doi: 10.1016/j.mbs.2019.01.012] [Medline: 30735696]
33. Ghassemi MM, Alhanai T, Westover MB, Mark TG, Nemati S. Personalized Medication Dosing Using Volatile Data Streams. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018 Presented at: AAAI'18; February 2-7, 2018; New Orleans, Louisiana, USA URL: https://aaai.org/ocs/index.php/WS/AAAIW18/paper/view/17234
34. Lin R, Stanley MD, Ghassemi MM, Nemati S. A deep deterministic policy gradient approach to medication dosing and surveillance in the ICU. Conf Proc IEEE Eng Med Biol Soc 2018 Jul;2018:4927-4931 [FREE Full text] [doi: 10.1109/EMBC.2018.8513203] [Medline: 30441448]
35. Raghu A, Komorowski M, Ahmed I, Celi L, Szolovits P, Ghassemi M. Deep reinforcement learning for sepsis treatment. arXiv 2017:- epub ahead of print(1711.09602) [FREE Full text]
36. Raghu A, Komorowski M, Celi L, Szolovits P, Ghassemi M. Continuous state-space models for optimal sepsis treatment-a deep reinforcement learning approach. arXiv 2017:- epub ahead of print(1705.08422) [FREE Full text]
37. Raghu A, Komorowski M, Singh S. Model-based reinforcement learning for sepsis treatment. arXiv 2018:- epub ahead of print-1811.09602 [FREE Full text]
38. Futoma J, Lin A, Sendak M, Bedoya A, Clement M, O'Brien C, et al. Learning to Treat Sepsis with Multi-Output Gaussian Process Deep Recurrent Q-Networks. OpenReview. 2018. URL: https://openreview.net/forum?id=SyxCqGbRZ [accessed 2020-06-10]

XSL•FO
RenderX

39. Peng X, Ding Y, Wihl D, Gottesman O, Komorowski M, Lehman LH, et al. Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. AMIA Annu Symp Proc 2018;2018:887-896 [FREE Full text] [Medline: 30815131]
40. Lee D, Srinivasan S, Doshi-Velez F. Truly Batch Apprenticeship Learning with Deep Successor Features. 2019 Presented at: International Joint Conferences on Artificial Intelligence Organization; August 10-16, 2019; Macao, China. [doi: 10.24963/ijcai.2019/819]
41. Lopez-Martinez D, Eschenfeldt P, Ostvar S, Ingram M, Hur C, Picard R. Deep reinforcement learning for optimal critical care pain management with morphine using dueling double-deep Q networks. Conf Proc IEEE Eng Med Biol Soc 2019 Jul;2019:3960-3963. [doi: 10.1109/EMBC.2019.8857295] [Medline: 31946739]
42. Petersen B, Yang J, Grathwohl WS, Cockrell C, Santiago C, An G, et al. Precision medicine as a control problem: Using simulation and deep reinforcement learning to discover adaptive, personalized multi-cytokine therapy for sepsis. arXiv preprint. arXiv:1802.10440 2018 [FREE Full text]
43. Prasad N, Cheng LF, Chivers C, Draugelis M, Engelhardt B. A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. Arxiv 2017:- epub ahead of print(1704.06300) [FREE Full text]
44. Cheng L, Prasad N, Engelhardt BE. An optimal policy for patient laboratory tests in intensive care units. Pac Symp Biocomput 2019;24:320-331 [FREE Full text] [Medline: 30864333]
45. Yu C, Liu J, Zhao H. Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Med Inform Decis Mak 2019 Apr 9;19(Suppl 2):57 [FREE Full text] [doi: 10.1186/s12911-019-0763-6] [Medline: 30961594]
46. Weng W, Gao M, He Z, Yan S, Szolovits P. Representation and reinforcement learning for personalized glycemic control in septic patients. arXiv 2017:- epub ahead of print(1712.00654) [FREE Full text]
47. Wang L, Zhang W, He X, Zha H. Supervised Reinforcement Learning with Recurrent Neural Network for Dynamic Treatment Recommendation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD'18; August 19-23, 2018; London, UK. [doi: 10.1145/3219819.3219961]
48. Manrai AK, Patel CJ, Ioannidis JP. In the era of precision medicine and big data, who is normal? J Am Med Assoc 2018 May 15;319(19):1981-1982. [doi: 10.1001/jama.2018.2009] [Medline: 29710130]
49. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
50. Ng A. CS294A Lecture Notes: Sparse Autoencoder. Stanford University. 2011. URL: https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf [accessed 2020-06-09]
51. Kane-Gill SL, Dasta JF, Buckley MS, Devabhakthuni S, Liu M, Cohen H, et al. Clinical practice guideline: safe medication use in the ICU. Crit Care Med 2017 Sep;45(9):e877-e915. [doi: 10.1097/CCM.0000000000002533] [Medline: 28816851]
52. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
53. Sinderby C, Breck J, Brander L. Bedside monitoring of diaphragm electrical activity during mechanical ventilation. In: Vincent JL, editor. Yearbook of Intensive Care and Emergency Medicine. New York, NY: Springer; 2009:385-393.
54. Patel SB, Kress JP. Sedation and analgesia in the mechanically ventilated patient. Am J Respir Crit Care Med 2012 Mar 1;185(5):486-497. [doi: 10.1164/rccm.201102-0273CI] [Medline: 22016443]
55. Cohen J, Shapiro M, Grozovski E, Singer P. Automatic tube compensation-assisted respiratory rate to tidal volume ratio improves the prediction of weaning outcome. Chest 2002 Sep;122(3):980-984. [doi: 10.1378/chest.122.3.980] [Medline: 12226043]
56. Kashyap R, Loftsgard TO. Clinicians role in reducing lab order frequency in ICU settings. J Perioper Crit Intens Care Nurs 2015;2(1):320-331 [FREE Full text] [doi: 10.4172/jpcic.1000112]
57. Osborne MA, Roberts SJ, Rogers A, Ramchurn SD, Jennings NR. Towards Real-Time Information Processing of Sensor Network Data Using Computationally Efficient Multi-output Gaussian Processes. In: International Conference on Information Processing in Sensor Networks. 2008 Presented at: IPSN'08; April 22-24, 2008; St Louis, MO, USA. [doi: 10.1109/ipsn.2008.25]
58. Kent JT. Information gain and a general measure of correlation. Biometrika 1983 Apr;70(1):163-173. [doi: 10.2307/2335954]
59. Liu X, Kruger P, Roberts MS. Optimizing drug dosing in the ICU. In: Yearbook of Intensive Care and Emergency Medicine. New York, USA: Springer; 2009.
60. Smith M. Monitoring and managing raised intracranial pressure after traumatic brain injury. In: Vincent JL, editor. Intensive Care Medicine. New York, NY: Springer; 2009:801-808.
61. Iannuzzi E, Iannuzzi M, Viola G, Sidro L, Cardinale A, Chiefari M. BIS - AAI and clinical measures during propofol target controlled infusion with Schnider's pharmacokinetic model. Minerva Anestesiol 2007;73(1-2):23-31 [FREE Full text] [Medline: 17115013]
62. Levi M. Emergency reversal of antithrombotic treatment. Intern Emerg Med 2009 Apr;4(2):137-145. [doi: 10.1007/s11739-008-0201-8] [Medline: 19002653]

XSL•FO
RenderX

63. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 2011 May;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]
64. Cohen J, Vincent J, Adhikari NK, Machado FR, Angus DC, Calandra T, et al. Sepsis: a roadmap for future research. Lancet Infect Dis 2015 May;15(5):581-614. [doi: 10.1016/S1473-3099(15)70112-X] [Medline: 25932591]
65. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). J Am Med Assoc 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
66. Jeter R, Josef C, Shashikumar S, Nemati S. Does the artificial intelligence clinician learn optimal treatment strategies for sepsis in intensive care? arXiv 2019:- epub ahead of print(1902.03271) [FREE Full text]
67. Lillicrap T, Hunt J, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. arXiv 2015:- epub ahead of print(1509.02971) [FREE Full text]
68. An G. In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. Crit Care Med 2004 Oct;32(10):2050-2060. [doi: 10.1097/01.ccm.0000139707.13729.7d] [Medline: 15483414]
69. Barr J, Fraser GL, Puntillo K, Ely EW, Gélinas C, Dasta JF, American College of Critical Care Medicine. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306. [doi: 10.1097/CCM.0b013e3182783b72] [Medline: 23269131]
70. Yu G, Zhang F, Tang S, Lai M, Su R, Gong Z. Continuous infusion versus intermittent bolus dosing of morphine: a comparison of analgesia, tolerance, and subsequent voluntary morphine intake. J Psychiatr Res 2014 Dec;59:161-166. [doi: 10.1016/j.jpsychires.2014.08.009] [Medline: 25193460]
71. Rutter PC, Murphy F, Dudley HA. Morphine: controlled trial of different methods of administration for postoperative pain relief. Br Med J 1980 Jan 5;280(6206):12-13 [FREE Full text] [doi: 10.1136/bmj.280.6206.12] [Medline: 6986940]
72. Nagaraj SB, McClain LM, Zhou DW, Biswal S, Rosenthal ES, Purdon PL, et al. Automatic classification of sedation levels in ICU patients using heart rate variability. Critical Care Medicine 2016;44(9):e782-e789. [doi: 10.1097/ccm.0000000000001708] [Medline: 27035240]
73. Carrasco G. Instruments for monitoring intensive care unit sedation. Crit Care 2000;4(4):217-225 [FREE Full text] [doi: 10.1186/cc697] [Medline: 11094504]
74. Lewis SR, Pritchard MW, Fawcett LJ, Punjasawadwong Y. Bispectral index for improving intraoperative awareness and early postoperative recovery in adults. Cochrane Database Syst Rev 2019 Sep 26;9(6):CD003843 [FREE Full text] [doi: 10.1002/14651858.CD003843.pub4] [Medline: 31557307]
75. Williams JD, Young S. Partially observable Markov decision processes for spoken dialog systems. Comput Speech Lang 2007 Apr;21(2):393-422. [doi: 10.1016/j.csl.2006.06.008]
76. Maei H, Szepesv´ari C, Bhatnagar S, Sutton R. Toward Off-Policy Learning Control with Function Approximation. In: The 27th International Conference on Machine Learning. 2010 Presented at: ICML'10; June 21-24, 2010; Haifa, Israel URL: https://icml.cc/Conferences/2010/papers/627.pdf
77. McShea M, Holl R, Badawi O, Riker RR, Silfen E. The eICU research institute - a collaboration between industry, health-care providers, and academia. IEEE Eng Med Biol Mag 2010;29(2):18-25. [doi: 10.1109/MEMB.2009.935720] [Medline: 20659837]
78. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4]

Abbreviations

AI: artificial intelligence
aPTT: activated partial thromboplastin time
BIS: bispectral index
BUN: blood urea nitrogen
DNN: deep neural network
DQN: deep Q network
eRI: eICU Research Institute
FQI: fitted-Q-Iteration
ICU: intensive care unit
IV: intravenous
MAP: mean arterial pressure
MDP: Markov decision process
MIMIC III: Medical Information Mart for Intensive Care III
MV: mechanical ventilation
RCT: randomized controlled trial
RL: reinforcement learning

XSL•FO
RenderX

WBC: white blood cell

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by A Hallawa, Z Zhang, D Maslove; comments to author 08.04.20; revised version received 05.05.20; accepted 13.05.20; published 20.07.20.

Please cite as:
Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M
Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review
J Med Internet Res 2020;22(7):e18477
URL: https://www.jmir.org/2020/7/e18477
doi:10.2196/18477
PMID:32706670

©Siqi Liu, Kay Choong See, Kee Yuan Ngiam, Leo Anthony Celi, Xingzhi Sun, Mengling Feng. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Implementing Web-Based Therapy in Routine Mental Health Care: Systematic Review of Health Professionals' Perspectives

Fiona Davies[1], BA(Hons), MPsyc, PhD; Heather L Shepherd[1], BA(Hons), Dip HE, PhD; Lisa Beatty[2], BPsych(Hons), PhD; Brodie Clark[1], BPsych (Hons); Phyllis Butow[1], BA(Hons)DipEd, MClin Psych, MPH, PhD; Joanne Shaw[1], BAppl Sc, BPsych(Hon), PhD

[1]Psycho-Oncology Co-operative Research Group, School of Psychology, The University of Sydney, Sydney, Australia

[2]Flinders Centre for Innovation in Cancer, College of Medicine & Public Health, Flinders University South Australia, Adelaide, Australia

Corresponding Author:
Heather L Shepherd, BA(Hons), Dip HE, PhD
Psycho-Oncology Co-operative Research Group
School of Psychology
The University of Sydney
Lifehouse Level 6-North (C39Z)
Sydney, 2006
Australia
Phone: 61 2 8627 0828
Email: heather.shepherd@sydney.edu.au

Abstract

Background: Web-based therapies hold great promise to increase accessibility and reduce costs of delivering mental health care; however, uptake in routine settings has been low.

Objective: Our objective in this review was to summarize what is known about health care professionals' perceptions of the barriers to and facilitators of the implementation of web-based psychological treatments in routine care of adults in health care settings.

Methods: We searched 5 major databases (MEDLINE, EMBASE, PsycINFO, CINAHL, and the Cochrane Library) for qualitative, quantitative, or mixed-methods studies exploring health professionals' views on computer- or internet-based psychological treatment programs. We coded included articles for risk of bias and extracted data using a prepiloted extraction sheet.

Results: We identified 29 eligible articles: 14 qualitative, 11 quantitative, and 4 mixed methods. We identified the following themes: patient factors, health professional factors, the therapeutic relationship, therapy factors, organizational and system factors, and models of care. Health professionals supported web-based therapies only for patients with relatively straightforward, low-risk diagnoses, strong motivation and engagement, high computer literacy and access, and low need for tailored content. They perceived flexibility with timing and location as advantages of web-based therapy, but preferred blended therapy to facilitate rapport and allow active monitoring and follow-up of patients. They emphasized the need for targeted training and organizational support to manage changed workflows. Health professionals were concerned about the confidentiality and security of client data for web-based programs, suggesting that clear and transparent protocols need to be in place to reassure health professionals before they will be willing to refer.

Conclusions: Without health professionals' support, many people will not access web-based therapies. To increase uptake, it is important to ensure that health professionals receive education, familiarization, and training to support them in incorporating web-based therapies into their practice, and to design systems that support health professionals in this new way of working with patients and addressing their concerns.

Trial Registration: PROSPERO CRD42018100869; https://tinyurl.com/y5vaoqsk

(J Med Internet Res 2020;22(7):e17362) doi:10.2196/17362

KEYWORDS

health professional views; implementation; online psychological therapy; online CBT; barriers; facilitators; models of care; cognitive behavioral therapy; internet-based intervention

XSL•FO
RenderX

Introduction

Background

Internationally there is a move toward using digital technologies in mental health care, including the development of an increasing number of Web-based therapy programs [1,2]. The reasons for this interest in electronic mental health commonly include accessibility, flexibility (in terms of when and where they can be accessed), lack of mental health care professionals to cover need, and cost considerations [3,4]. Randomized controlled trials of web-based therapies show promise, with efficacy reported to be similar to that of face-to-face interventions [5]. A systematic review of internet-based cognitive behavioral therapy [6], for example, found that this therapy has been tested for 25 different clinical disorders, with large effect sizes meeting the highest level of criteria for evidence for disorders such as depression, anxiety, and severe health anxiety.

Despite this interest, research on referral and uptake of web-based therapies in routine care is limited [7]. In a rapid review of literature on this topic, we found no studies set in hospital-based medical clinics and only 1 study in primary care, which reported extremely low rates of referral (14% of the expected number) and uptake (1% of those referred) [8]. A major implementation project in Europe, known as MasterMind, is studying web-based therapies in routine mental health care [9,10], but no data are published at this stage.

Nor is there a large body of evidence on effective processes for implementation of web-based therapies in routine clinical care [7]. A recent review of staff-reported barriers to and facilitators of hospital-based interventions of any kind (web-based or face-to-face) [11] identified 3 key domains that may affect implementation success—system (eg, workload and workflow, physical structure and resources, culture, communication, and external pressures), staff (eg, attitudes, understanding and awareness, role identity, skills, and confidence), and intervention (eg, ease of integration, validity and evidence base, safety and ethics, supporting resources)—with similar barriers likely in community settings [12]. Several conceptual frameworks have been developed to further explicate and guide health care implementation, including the Consolidated Framework for Implementation Research (CFIR) [13]. CFIR is composed of 5 major domains: intervention characteristics, the inner setting, the outer setting, characteristics of individuals involved in implementation, and the implementation process. As health professionals often act as gatekeepers to web-based programs, they are key stakeholders and may be very influential in determining whether web-based programs are disseminated widely. Thus, in this systematic review, we decided to focus on health professionals' perspectives of web-based therapies, which falls within the fourth CFIR domain.

Objective

Several individual studies reported data on health professionals' perceptions of web-based therapy programs, which revealed concerns about therapeutic alliance and the quality and effectiveness of programs [14,15], patient commitment and compliance [15,16], patient barriers such as internet literacy and access [16,17], and suicide risk [16]. However, we found no synthesis of these studies to provide an overall picture of health professionals' perspectives to date. Thus, we sought to fill this gap by conducting a systematic review. Our research question was what are health professionals' perceptions of the barriers to and facilitators of the implementation of web-based psychological treatments in routine care in health settings?

Methods

Review Registration

We registered the review with the International Prospective Register of Systematic Reviews (PROSPERO CRD42018100869).

Searches

We searched 5 bibliographic databases: MEDLINE, EMBASE, PsycINFO, CINAHL, and the Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register). The search strategy included terms relating to web-based approaches, psychological therapies, and views of barriers to and facilitators of dissemination.

Qualitative, quantitative, or mixed-methods studies exploring health professionals' views on computer- or internet-based psychological treatment programs were eligible if the treatment contained a component designed to change psychological symptoms or behaviors either associated with a diagnosable mental health condition (using *International Classification of Diseases, Tenth Revision*, Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition], or equivalent) or secondary to a physical health condition such as cancer. This could include anxiety, depression, substance misuse, suicidality, insomnia, complex pain, fatigue (eg, chronic fatigue syndrome or cancer-related fatigue), or medication adherence. Studies of views on all models of web-based therapy, from entirely self-directed to blended with face-to-face care, were eligible. Additionally, the patients targeted (and the health professionals) could be situated either in a hospital or clinic setting or in the general community.

We included only studies investigating views on web-based programs for adults (>18 years of age); we excluded programs targeting adults who were completing the program on behalf of someone else, for example, parents completing a program to help their children with anxiety or behavioral problems.

We included studies of the views of all health professionals involved in referring, triaging, or providing psychosocial care to patients with the above conditions (eg, doctors, nurses, psychosocial health professionals, allied health practitioners). We excluded nonclinical professional groups who do the above (eg, schoolteachers, religious advisers).

We limited included studies to peer-reviewed articles, excluding conference abstracts. We searched studies published from 1986 (when the first known digital mental health site, Ask Uncle Ezra, at Cornell University was launched) to the end of October 2019.

XSL•FO
RenderX

Titles or abstracts, or both, of articles retrieved during the search strategy or from reference lists were screened by 1 reviewer to identify studies that met the inclusion criteria described above. As an initial step, 10% of the titles or abstracts and full-text articles were independently reviewed by a second reviewer to confirm inclusion and exclusion decisions. Following this, any titles or abstracts or full-text articles where the reviewer had any uncertainty about the decision was put aside and discussed with a second reviewer. Any continuing uncertainty was resolved through retrieving the full text of the study or discussion with other members of the authorship team. We hand searched reference lists of included articles for relevant articles.

Data Extraction

We extracted information from articles meeting the inclusion criteria, with 2 team members completing extraction for each article. Any uncertainty was resolved through discussion with additional members of the review team. Data extracted were author, country, study design, sample characteristics, web-based therapy characteristics, measures, results, and limitations.

Risk-of-Bias (Quality) Assessment

We assessed the selected articles for risk of bias using checklists appropriate to quantitative and qualitative studies from Kmet et al [18] and to mixed-methods studies from Pluye et al [19]. All articles were graded by 2 team members and discussed with the author team if there was any ambiguity in grading.

Strategy for Data Synthesis

We analyzed the articles using a 3-stage synthesis process that combined qualitative, quantitative, and mixed-methods studies comprehensively and rigorously to answer the review question, following the methods of Thomas et al [20]. First, we summarized quantitative data using descriptive statistics. Second, we synthesized qualitative data by themes [21]. All authors coded 2 qualitative studies line by line to develop a draft coding tree and themes, which we then refined after coding 2 additional studies. We then applied this coding structure to the remaining qualitative studies, with emerging themes created and discussed by all authors. Similarities and differences between codes and their content were examined to develop higher-order themes. (See Multimedia Appendix 1 for the final coding tree.) Third, we used a matrix to compare the qualitative, quantitative, and mixed-methods findings from the first 2 stages to answer the review research question.

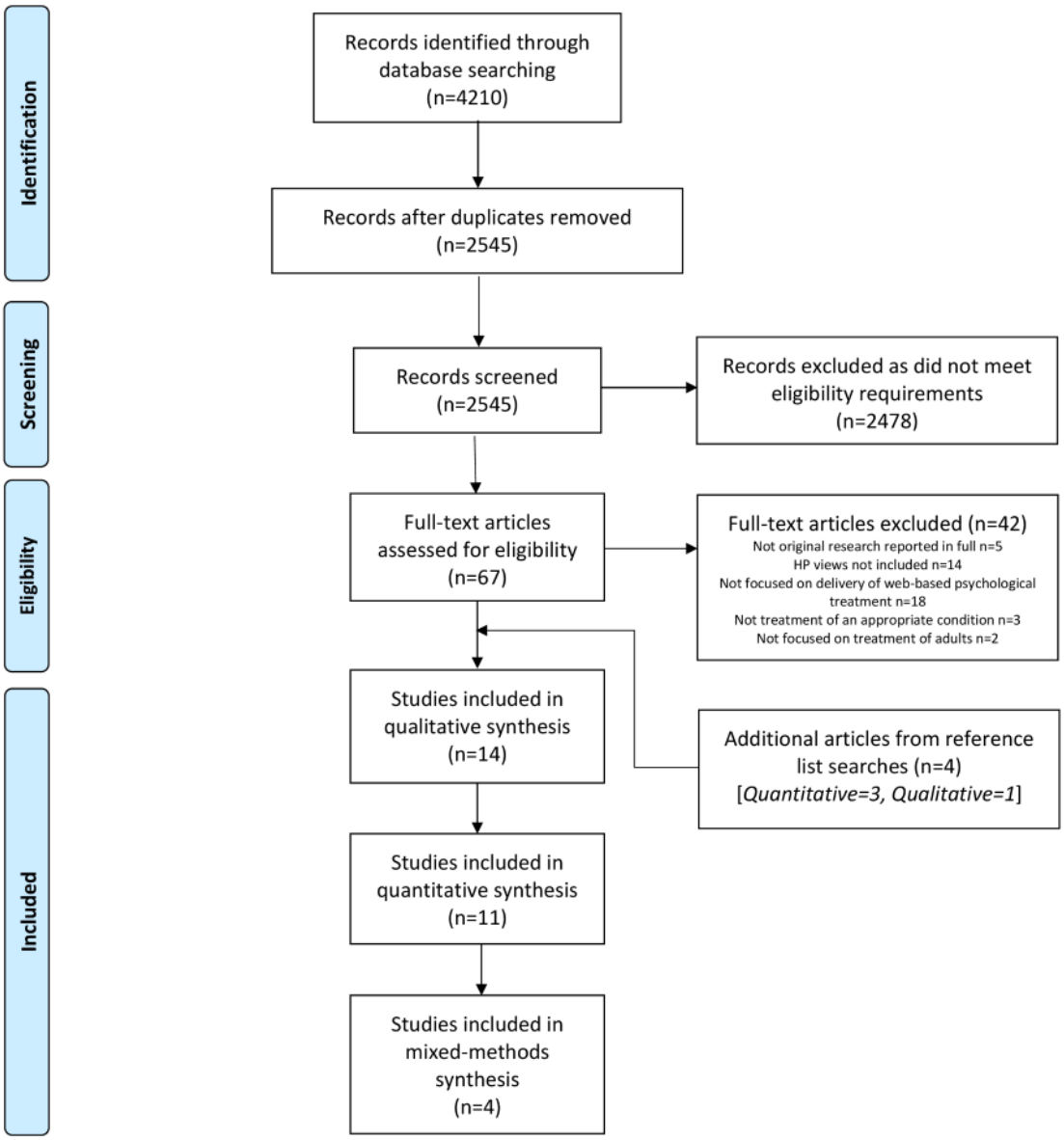
Results

Search Results

We initially identified 4210 potentially eligible articles from the search, with 2545 remaining after deduplication. Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram, with 29 articles meeting full inclusion criteria: 14 qualitative, 11 quantitative, and 4 mixed methods. Most articles were published since 2009 and reported results from the United Kingdom (n=5), the United States (n=6), Australia (n=4), Canada (n=4), Europe (n=10), and Israel (n=1). A total of 16 studies specified the professional background of the included health professionals (psychologists, n=12 studies; social workers, n=9 studies; general practitioners, n=4 studies; nurses, n=3 studies; and psychiatrists, n=2 studies), and the remaining 13 studies used general terms (eg, “psychotherapist” or “counsellor” or “practitioner”). Where studies included both health professionals and patients, we included only data from the health professionals in the review. See Multimedia Appendix 2 for study characteristics. All included articles were of acceptable quality, with reflexivity and use of verification procedures to establish credibility the most poorly done (see Multimedia Appendix 3 for quantitative studies, Multimedia Appendix 4 for qualitative studies, and Multimedia Appendix 5 for mixed-methods studies).

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. HP: health professional.



Thematic Synthesis

We identified 6 themes, which mapped well to the CFIR: therapy factors, organizational (or inner setting) and system (or outer setting) factors, patient factors, health professional factors, the therapeutic relationship, and models of care.

Therapy Factors

In total, 3 of 11 quantitative studies (n=677) [22-24], 5 of 14 qualitative studies (n=78) [14,25-28], and 3 of 4 mixed-methods studies (n=111) [17,29,30] explored the impact of therapy factors on the uptake of web-based therapies. Findings suggested that health professionals believed there was evidence [29,30] supporting the utility of web-based therapies but were generally more comfortable using web technologies as an *adjunct* to therapy, rather than as a *stand-alone* intervention [17,24,26] (see Models of Care subsection below for more detail). Overall, health professionals were neutral to positive toward web-based programs and identified the following consistent therapeutic factors that facilitate or hinder web-based therapy uptake.

The therapeutic content and approach of therapies were frequently raised by study participants as being more or less appropriate for a web-based format, with therapists more likely to refer patients for web-based psychoeducation, monitoring, and skills practice [17,23,26] within a positive psychology, mindfulness, acceptance and commitment therapy, or cognitive behavioral therapy (CBT) framework [17,23]. They were less likely to consider web-based therapy appropriate for psychoanalytic, interpersonal psychotherapy, or schema therapies. Nor were they comfortable with serious gaming [17,22], chat, or video components [17,24]. Concerns were raised regarding privacy of data, ethical considerations for suicidal patients, and levels of evidence available for intervention effectiveness [22,23].

The structured nature of web-based programs, the most commonly raised theme, was extensively discussed as both a facilitator of and barrier to uptake or recommendation. Whereas structure provided focus and aided navigation of programs [27,28], other studies noted that this needs to be balanced with flexible tailoring of content [29,30], a challenge that is

XSL•FO
RenderX

inadequately addressed [14,27]. Some health professionals in the qualitative studies reported wanting to see more transdiagnostic options [25]. Hadjistavropoulos and colleagues partially addressed this recommendation in their mixed-methods study of a transdiagnostic web-based CBT intervention, with feedback specifically citing the transdiagnostic nature as a strength of the program [29].

Organizational and System Factors

A total of 6 of 11 quantitative studies (n=4171) [23,24,31-34], 10 of 14 qualitative studies (n=158) [14,25-28,35-39], and 2 of 4 mixed-methods studies (n=99) [17,29] reported organizational and system concerns, which fall within the second and third domains of the CFIR.

Training

Training was highlighted as driving confidence in recommending web-based therapy options to patients [28]. Health professionals wanted more information, training, clear guidelines, and information on liability to be comfortable with web-based approaches [34].

Training to improve computer skills and stay up to date [23] and the need for better information technology support were also identified [17]. Qualitative studies [14,25,28] highlighted training needs in email or electronic communication skills in a therapeutic context [14,25] and, from an organizational perspective, the time required to obtain skills and to respond to patients in this format. Notably, 1 study suggested that the opportunity to pause and reflect before responding to emails was a strength of web-based therapy [37].

Time Saver or Time Changer

Health professionals reported varying views regarding whether web-based approaches made therapy delivery harder or easier. In the study by Bengtsson et al [27], therapists who were experienced in delivering both web-based and face-to-face therapy noted that work time was more flexible with web-based CBT, enabling work during cancellations, and flexibility in appointment times and place of work. Other advantages included therapy being unaffected by illness, emergencies, or other work; allowing colleagues to take over each other's work if needed; and the potential to work with more clients at once [27]. In contrast, another study noted that web-based therapy was time consuming for both patients and therapists, with experienced web-based therapists spending more time overall on web-based treatment than with face-to-face treatment [25].

Managerial and Organizational Support

In 2 studies, support and leadership from organization staff and local management was noted as valuable in encouraging use of web-based CBT [17,25]. In another quantitative study [31], perceived supportiveness of the organizational environment predicted intention to use web-based CBT. A large study found small effects of workplace factors, such that working in a community treatment center (rather than private practice) and having good access to technology at work increased perceptions of the efficacy of blended treatment [32].

Accessibility and Integration

Some studies reported that health professionals found it easier to integrate written materials into care, rather than referring to web-based programs where specific content is difficult to access easily [38,39]. In the study by Batka and colleagues, health professionals expressed mixed views regarding the ability of web-based therapy to bridge the gap between primary care and behavioral health [36]. One study supported the idea of web-based therapies identifying those in need of additional support [35]. Several studies noted the need to supplement patient self-referral and management with health professional monitoring and follow-up, to ensure the program matched the patient's needs [14,28] and to encourage adherence. Rural health professionals in 1 study [26] expressed concern that an increase in web-based treatment may further isolate rural patients without internet access:

> *Often online services have been looked at as, you know, the great hope for areas where there aren't real services for people, but where there isn't adequate internet access, you are further marginalizing people who live in remote areas who now don't have access to two different services...* [pg 5]

Data Security and Privacy

In 1 study, confidentiality of patient information was listed as the primary concern by 48% of health professionals, with 81% stating they had some concern [33]. Two quantitative studies highlighted health professionals' concerns about information security [17,23], ethical or clinical guidelines [17,23], and legal issues or liability [17], before supporting use of web-based therapies. Concern about confidentiality was raised in 1 study, where health professionals expressed preferences for using secure email or an app over chat or video formats [24]. Qualitative studies identified protocols to address privacy and ethical practice in an information technology context [37,38].

Patient Factors

A total of 4 of 11 quantitative studies [22,23,33,34] incorporating data from 3144 health professionals, 9 of 14 qualitative studies (n=158) [25,26,35-38,40-42], and 3 of 4 mixed-methods studies (n=374) [29,30,43] raised issues related to patient factors. Most health professionals in a quantitative study [22] believed that web-based therapies would be appropriate for 20% of patients; health professionals in another study perceived barriers in 22% to 48% of patients [23]. One study argued that web-based therapies allowed patients to become active participants in their own recovery [40].

Nature and Severity of Symptoms

Patient factors identified in both quantitative and qualitative studies included the nature and severity of symptoms, with anxiety disorders commonly cited as a good fit for web-based therapy [26,34,43], and more complex or severe disorders viewed as a poor fit [26,37,40]. As anxiety and depression symptom severity increased from mild to severe, the disorders were seen as having decreasing fit for web-based therapy [34].

XSL•FO
RenderX

Individual Characteristics

Health professionals were concerned that comorbidity [25] or suicidal ideation required risk management [35]. It was also noted [30] that web-based therapy cannot easily take into account patient characteristics including comorbidity, needs and motivation, written expression, skills, and personality (including self-management). For example [37]:

> *I found for my anxiety clients, it worked quite well. For the depression clients, who are more severe, I found that they tend to take much longer...there's less motivation...those clients, maybe it would be better for them to like see somebody in person because there's a lot of other issues.* [pg 44]

Accessibility

Health professionals identified access to face-to-face treatments as a factor, with several studies noting that web-based therapies were particularly attractive and useful to patients who could *not* readily access therapy [30] (eg, those with limited mobility, time, or access to local services [37,38], with perceived stigma [29], or in rural or remote locations [33]).

Practical Barriers

Health professionals expressed concern about practical barriers such as access to technology [23] or the internet [35,41]. Additional barriers included having poor literacy, not being "psychologically minded," or having difficulties with vision or hearing [41]. Health professionals commented that their patients often had grade 6 to 8 reading levels and could not cope with the reading required [36]. This could also impact ability to complete written tasks [37]:

> *If they don't provide the information or if they're having trouble expressing exactly what the problem was, then it's obviously going to be more difficult.* [pg 46]

Health Professional Factors

Health professional factors relevant to the uptake of web-based therapies were raised in 9 of 11 quantitative studies (n=4536) [22,23,31-33,44-47], 6 of 14 qualitative studies (n=96) [25-28,38,48], and 3 of 4 mixed-methods studies (n=407) [17,30,43]. Quantitative studies raised few health professional issues. A total of 3 studies suggested that psychodynamic therapists are less positive toward web-based therapy [22,32,46], but another found no difference [44]. One study found that openness to evidence-based practice predicted more favorable attitudes [32], whereas another found that expectations of ease of use predicted intentions to use web-based CBT [31]. The main issue raised was lack of familiarity with web-based therapy, consistent with the low uptake of these approaches generally [17,23,31,45,47].

Several barriers related to health professionals were raised in the qualitative studies. There were concerns about the technological skills and comfort level of health professionals [26,48]. Significant concerns were raised about workload issues: finding time to incorporate web-based treatment [28] and managing the different demands of electronic work [25]. A final factor raised by therapists was that they saw web-based treatment as both less engaging and less taxing than face-to-face treatment. For example [27,38]:

> *I just felt like I was in a call centre rather than being a clinician working with patients who had difficulties.* [pg 8]

> *You feel very much less burdened by [internet-based CBT] than in regular outpatient care.* [pg 473]

The mixed-methods studies raised concerns related to health professionals' knowledge and training in web-based therapy. Motivation and use of web-based therapy were low, with reported rates of 2% [33] and 2.4% [43], in part reflecting misunderstandings about confidentiality and liability [33]. For example [30]:

> *My concern is not to get patients motivated to use the online modules, but to get therapists to use them. That is a much larger bump. Everything that is new is seen as more work, and everybody is already loaded with work and doesn't want more work. I'm afraid that pressure from management is the only way to get therapists to work with it.* [pg 7]

The Therapeutic Relationship

A total of 4 of 11 quantitative studies (n=3755) [22,23,32,33], 6 of 14 qualitative studies (n=751) [25,27,28,38,39,48], and no mixed-methods studies explored the impact of web-based therapy on the quality of the health professional-patient therapeutic relationship. Across studies, health professionals who were experienced in web-based therapy held more positive views of the potential for a therapeutic relationship in web-based therapy [33,38,39].

Quantitative findings highlighted minimal concern among health professionals about establishing and maintaining a therapeutic relationship through mobile apps [23] or blended face-to-face and web-based therapy, although concerns were raised in 1 study that including serious games may lead to neglect of relationships and communication during therapy [22]. Health professionals in another large quantitative study indicated that they believed integrating web-based approaches in therapy would interfere with rapport [32].

Among qualitative studies, health professionals perceived the therapeutic relationship as *different from that in face-to-face therapy but not necessarily worse*, and some reported being surprised by their ability to develop relationships online [25]. Bengtsson et al [27] highlighted that the web context may extend the time needed to develop a therapeutic relationship but, as it is less confronting, may be particularly helpful in building relationships with patients with social anxiety.

Two studies reported the views of primary care physicians [28,48], who were ambivalent about the impact of web-based therapy on therapeutic relationships, noting that psychodynamic approaches relied on *open and active dialogue* between health professional and patient.

Models of Care

A total of 3 of 11 quantitative studies (n=677) [22-24], 8 of 14 qualitative studies (n=129) [14,25,28,35,36,38,39,42], and all

XSL•FO
RenderX

4 mixed-methods studies (n=440) [17,29,30,45] explored health professionals' views and preferences regarding models of care delivery incorporating web-based components. In addition, 9 studies [14,25,28,29,35,38,39,42,43] investigated use of web-based therapies as a stand-alone treatment, 2 studies [22,23] discussed preferences for serious gaming and mobile apps as an adjunct to therapy, and 3 studies [17,24,30] explored blending face-to-face and web-based components as part of an integrated approach to therapy. One study [36] included the views of mental health and primary care providers as 1 model of care delivery in a suite of telehealth options.

Among web-based therapy studies, views differed based on whether the therapy was guided or self-directed (low intensity). Although it was acknowledged that self-referral to self-help modules increased ease of access for patients [39] and made it more likely that patients were motivated to engage with treatment [14], there were concerns that this model of care did not provide sufficient health professional support, especially for challenging tasks [29,38]. Health professionals were more comfortable with guided self-help, as they perceived this as fitting more within the therapeutic role [38]. General practitioners trained to administer guided web-based CBT highlighted the challenges in integrating web-based therapy process issues into consultations [28]. Overall, there was a preference to use web-based therapy flexibly, possibly as an adjunct to face-to-face treatment, [38,42,43], as an option to support those on a waiting list for treatment [29] or as part of a stepped-care model [25].

Blended therapy was perceived as a viable alternative to a web-based model, as compatible with current clinical practice, and as raising few concerns about treatment efficacy with the inclusion of web-based components as part of a wider face-to-face model [17,24]. Models of care that incorporated serious gaming and use of mobile apps were perceived as potentially useful adjuncts to traditional therapy approaches to support skills development both during and after therapy [23,30].

Discussion

Principal Findings

Despite good evidence of efficacy in randomized controlled trials [5], the potential of web-based psychological therapies to increase access to mental health care has not been well realized in nontrial settings to date [49]. This is, to our knowledge, the first synthesis of quantitative and qualitative data on health professionals' perceptions of web-based therapy for mental health conditions (both primary and secondary to physical disorders). We found that health professionals perceived many barriers to routine use of web-based therapy and had strong views about appropriate models of care. While health professionals' concerns may not all be evidence based, health professionals are in many cases the gatekeepers to referral. Therefore, addressing their concerns is a key strategy to improve implementation, alongside ongoing education of health professionals so that they are familiar with the content, processes, and evidence base of web-based therapies and are comfortable with incorporating them into their model of care. The importance of thinking carefully about models of care specifically in the context of web-based therapies has not been addressed adequately in the literature to date.

Health professionals' concerns corresponded closely with the 5 domains of the CFIR [13], with issues raised pertaining to intervention characteristics (eg, flexibility), the inner setting (eg, managerial support), the outer setting (eg, data security and privacy), characteristics of individuals (eg, their health and internet literacy), and the implementation process (eg, health professional education). Furthermore, they also accord with the much broader findings of a systematic review of reviews concerning factors that influence implementation of electronic health [50]. We found, as did the review of reviews, that the factors influencing implementation are multilevel and complex, with no single factor acting as a key barrier or facilitator.

This review suggests that for web-based treatments to maximize their acceptability to health professionals, they should incorporate the following features.

Consider Tailoring

Health professionals noted that the highly structured nature of many web-based therapies, while easier to navigate [27], may contribute to patients' sense of being funnelled through a preset and rigid program [25,27]. This accords well with surveys of patients regarding reasons for dropout from web-based programs. For example, in a qualitative study of patients who had dropped out of a transdiagnostic web-based therapy program, they noted the lack of specificity of content to their own particular problems [51]. Indeed, dropout rates for web-based compared with face-to-face programs are reported to be 10% to 15% higher, reinforcing the need for careful consideration of tailoring, as well as treatment credibility and engaging content, which also predict dropout [52].

Target a Subset of Patients

Even successful online clinics do not see web-based therapy as suitable for all individuals [49], and patients also report a preference for face-to-face treatment over web based [53]. This was reflected in our findings, as health professionals supported the potential utility of web-based therapies for a subset of patients only [22,23]. The findings suggest that, to increase referrals and uptake, web-based treatments should be targeted toward individuals with mild to moderate anxiety and depression, with limited comorbidity and low risk. Further patient factors that may lead to more referrals include sufficient computer literacy and access, good motivation, good health literacy, needing more flexible access to treatment, and concerns about stigma.

Use an Approach That Blends Web-Based Therapy With Face-to-Face Therapist Contact

Health professionals reported concern about wholly self-directed treatments, with a clear preference for blended treatments including some therapist face-to-face contact, as it affords a better opportunity to establish rapport and allows active monitoring and follow-up of patients as they move through online tasks [29,38]. This is in line with patient research, which showed a preference for face-to-face treatment over web based [53], and that adherence increased with therapist support [54].

XSL•FO
RenderX

Whereas health professionals who have no experience with web-based treatments are concerned about the therapeutic alliance, those who were experienced in supported web-based treatment were much less concerned. This suggests that models of care that incorporate some level of therapist support are likely to become more acceptable over time as health professionals become accustomed to the model.

Select Therapeutic Approaches and Content That Are Appropriate for a Web-Based Format

Prioritizing psychoeducation, monitoring, and skills practice, with more complex or interpersonal interventions reserved for face-to-face treatment, may make health professionals more confident in web-based treatments. Considering the balance between providing a clear structure and yet allowing content to be tailored to individual needs may also be important, although our review does not offer any clear guidance on how to optimally achieve that.

Educate Health Professionals in Web-Based Treatment

The vast majority of health professionals are unfamiliar with web-based therapies [17,23,31,45,47], and this needs to be addressed in order to increase rates of referral and use. Implementation of any model of care would need to consider how to educate health professionals about web-based therapy, including how to make appropriate referral decisions, find high-quality evidence-based resources, and incorporate web-based treatment as part of routine care.

Design a Supportive System

The review suggests that the broader setting for health professionals is important in successful implementation. This has 4 aspects.

First, the effect of web-based therapy on the work of frontline mental health care workers would need to be carefully considered to ensure that therapists have adequately engaging work and appropriate supervision in skills specific to web-based therapy such as written communication. Providing web-based therapy requires a different workflow from face-to-face treatment, and this would need to be managed. Web-based therapy was seen as both less rewarding and less draining than face-to-face work; the former was a strong disincentive to some health professionals, who felt that their engagement in, and sense of reward from, delivering expert and successful care was being taken from them [27,38]. A sense of reward is a key dimension of stress and burnout in occupational health [55], and work would need to be designed in a way that is satisfying for health professionals.

Second, the provision of adequate technological resources and support is important, as well as good data security.

Third, clear policies and procedures are needed regarding confidentiality, risk management, and liability issues.

Fourth, clear support and expectations from management regarding the use of web-based therapy are required, including impact on workflow and referral pathways for face-to-face therapy.

Limitations

This review identified only 29 articles examining health professionals' perspectives on web-based therapy. These articles varied considerably in terms of the population of health professionals included as participants, as well as the model of care that was being examined (eg, blended, self-directed), and are unlikely to give a complete picture. The main limitation in the literature so far is the inclusion of a very small proportion of health professionals who are experienced in web-based therapy, with most health professionals responding to questions in the absence of direct experience.

Comparison With Prior Work

This is, to our knowledge, the first systematic review of health professionals' perspectives on web-based therapy and provides a synthesis of the research to date.

Conclusions

The findings of this review echo the factors found in a recent article documenting the experiences of 2 successful online clinics [49], which identified the importance of considering the patient, therapists and their training, comprehensive organizational and systemic support, and the place of web-based therapies in models of care. These results have important implications for the implementation of web-based therapy, emphasizing that important preparatory work is required if implementation is to be successful, in the way the intervention is designed, the training and support health professionals receive, and the engagement of senior management in supporting the transition to a different model of care.

Further research is needed to examine the acceptability of specific models of care [2], which may vary in level of therapist support, timing of the intervention (waitlist, posttherapy, etc), positioning in stepped care, or the way it is incorporated in routine care. As health professionals gain experience with web-based therapies, it will be important for their perspectives to be sought again, to elicit perspectives grounded in real interactions and patient feedback. Interventions to educate and support health professionals in using web-based methods will need to be developed and evaluated. Finally, large-scale implementation studies that document the implementation strategies employed, uptake and retention rates, and effectiveness outcomes will be needed to provide a solid evidence base for this therapeutic approach moving forward.

Acknowledgments

The research was supported by the ADAPT Program, funded by a Cancer Institute NSW (Eveleigh, Australia) Translational Program Grant (14/TPG/1-02).

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding tree.
[DOCX File , 15 KB - jmir_v22i7e17362_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 33 KB - jmir_v22i7e17362_app2.docx]

Multimedia Appendix 3
Risk-of-bias assessment for quantitative studies.
[DOCX File , 20 KB - jmir_v22i7e17362_app3.docx]

Multimedia Appendix 4
Risk-of-bias assessment for qualitative studies.
[DOCX File , 19 KB - jmir_v22i7e17362_app4.docx]

Multimedia Appendix 5
Risk-of-bias assessment for mixed-methods studies.
[DOCX File , 14 KB - jmir_v22i7e17362_app5.docx]

References

1. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
2. Reynolds J, Griffiths KM, Cunningham JA, Bennett K, Bennett A. Clinical practice models for the use of e-mental health resources in primary health care by health professionals and peer workers: a conceptual framework. JMIR Ment Health 2015;2(1):e6 [FREE Full text] [doi: 10.2196/mental.4200] [Medline: 26543912]
3. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: a model for dissemination into primary care. Clin Psychol Rev 2015 Nov;41:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]
4. Karageorge A, Murphy MJ, Newby JM, Kirsten L, Andrews G, Allison K, et al. Acceptability of an internet cognitive behavioural therapy program for people with early-stage cancer and cancer survivors with depression and/or anxiety: thematic findings from focus groups. Support Care Cancer 2017 Jul;25(7):2129-2136. [doi: 10.1007/s00520-017-3617-8] [Medline: 28213818]
5. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110. [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
6. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
7. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
8. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011 Sep;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]
9. Mol M, Dozeman E, van Schaik JF, Vis CPCD, Riper H, Smit JH. The therapist's role in the implementation of internet-based cognitive behavioural therapy for patients with depression: study protocol. BMC Psychiatry 2016 Sep 30;16(1):338 [FREE Full text] [doi: 10.1186/s12888-016-1045-9] [Medline: 27716108]
10. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
11. Geerligs L, Rankin NM, Shepherd HL, Butow P. Hospital-based interventions: a systematic review of staff-reported barriers and facilitators to implementation processes. Implement Sci 2018 Feb 23;13(1):36 [FREE Full text] [doi: 10.1186/s13012-018-0726-9] [Medline: 29475440]

XSL•FO
RenderX

12. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of internet interventions for depression: a scoping review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
13. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
14. Folker AP, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: a comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70. [doi: 10.1016/j.invent.2018.02.001]
15. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
16. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, et al. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9. [doi: 10.1016/j.invent.2017.01.001]
17. Dijksman I, Dinant G, Spigt M. The perception and needs of psychologists toward blended care. Telemed J E Health 2017 Dec;23(12):983-995. [doi: 10.1089/tmj.2017.0031] [Medline: 28556693]
18. Kmet LM, Lee RC, Cook LS. Standard Quality Assessment Criteria for Evaluating Primary Research Papers From a Variety of Fields. Edmonton, AB: Health Technology Assessment Unit, Alberta Heritage Foundation for Medical Research; 2004.
19. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
20. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, et al. Integrating qualitative research with trials in systematic reviews. BMJ 2004 Apr 24;328(7446):1010-1012 [FREE Full text] [doi: 10.1136/bmj.328.7446.1010] [Medline: 15105329]
21. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
22. Eichenberg C, Grabmayer G, Green N. Acceptance of serious games in psychotherapy: an inquiry into the stance of therapists and patients. Telemed J E Health 2016 Dec;22(11):945-951. [doi: 10.1089/tmj.2016.0001] [Medline: 27045452]
23. Pierce B, Twohig MP, Levin ME. Perspectives on the use of acceptance and commitment therapy related mobile apps: results from a survey of students and professionals. J Contextual Behav Sci 2016 Oct;5(4):215-224. [doi: 10.1016/j.jcbs.2016.08.001]
24. Shalom JG, Israel R, Shalom N. Combining face-to-face therapy with computerized techniques: a therapists' attitudes survey. Stud Health Technol Inform 2015;219:69-74. [Medline: 26799882]
25. Hadjistavropoulos HD, Alberts NM, Nugent M, Marchildon G. Improving access to psychological services through therapist-assisted, Internet-delivered cognitive behaviour therapy. Can Psychol 2014;55(4):303-311. [doi: 10.1037/a0037716]
26. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013 Sep 05;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]
27. Bengtsson J, Nordin S, Carlbring P. Therapists' experiences of conducting cognitive behavioural therapy online vis-a-vis face-to-face. Cogn Behav Ther 2015;44(6):470-479. [doi: 10.1080/16506073.2015.1053408] [Medline: 26090947]
28. Wilhelmsen M, Høifødt RS, Kolstrup N, Waterloo K, Eisemann M, Chenhall R, et al. Norwegian general practitioners' perspectives on implementation of a guided web-based cognitive behavioral therapy for depression: a qualitative study. J Med Internet Res 2014 Sep 10;16(9):e208 [FREE Full text] [doi: 10.2196/jmir.3556] [Medline: 25208886]
29. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of internet-delivered cognitive behavior therapy within community mental health clinics: a process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]
30. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJEWC. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
31. IJzerman RVH, van der Vaart R, Evers AWM. Internet-based cognitive behavioral therapy among psychologists in a medical setting: a survey on implementation. J Med Internet Res 2019 Aug 09;21(8):e13432 [FREE Full text] [doi: 10.2196/13432] [Medline: 31400101]
32. Becker EM, Jensen-Doss A. Computer-assisted therapies: examination of therapist-level barriers to their use. Behav Ther 2013 Dec;44(4):614-624. [doi: 10.1016/j.beth.2013.05.002] [Medline: 24094787]
33. Wells M, Mitchell KJ, Finkelhor D, Becker-Blease KA. Online mental health treatment: concerns and considerations. Cyberpsychol Behav 2007 Jun;10(3):453-459. [doi: 10.1089/cpb.2006.9933] [Medline: 17594270]
34. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]

XSL•FO
RenderX

35. Alberts NM, Hadjistavropoulos HD, Titov N, Dear BF. Patient and provider perceptions of internet-delivered cognitive behavior therapy for recent cancer survivors. Support Care Cancer 2018 Feb 15;26(2):597-603. [doi: 10.1007/s00520-017-3872-8] [Medline: 28913732]
36. Batka C, Tanielian T, Woldetsadik MA, Farmer C, Jaycox LH. Stakeholder experiences in a stepped collaborative care study within U.S. army clinics. Psychosomatics 2016 Nov;57(6):586-597. [doi: 10.1016/j.psym.2016.05.008]
37. Friesen LN, Hadjistavropoulos HD, Pugh NE. A qualitative examination of psychology graduate students' experiences with guided internet-delivered cognitive behaviour therapy. Internet Interv 2014 Apr;1(2):41-48. [doi: 10.1016/j.invent.2014.04.001]
38. Gellatly J, Pedley R, Molloy C, Butler J, Lovell K, Bee P. Low intensity interventions for Obsessive-Compulsive Disorder (OCD): a qualitative study of mental health practitioner experiences. BMC Psychiatry 2017 Feb 22;17(1):77 [FREE Full text] [doi: 10.1186/s12888-017-1238-x] [Medline: 28222706]
39. Lovell K, Bower P, Gellatly J, Byford S, Bee P, McMillan D, et al. Clinical effectiveness, cost-effectiveness and acceptability of low-intensity interventions in the management of obsessive-compulsive disorder: the Obsessive-Compulsive Treatment Efficacy randomised controlled Trial (OCTET). Health Technol Assess 2017 Jun;21(37):1-132 [FREE Full text] [doi: 10.3310/hta21370] [Medline: 28681717]
40. Advocat J, Lindsay J. Internet-based trials and the creation of health consumers. Soc Sci Med 2010 Feb;70(3):485-492. [doi: 10.1016/j.socscimed.2009.10.051] [Medline: 19926185]
41. Middlemass J, Davy Z, Cavanagh K, Linehan C, Morgan K, Lawson S, et al. Integrating online communities and social networks with computerised treatment for insomnia: a qualitative study. Br J Gen Pract 2012 Dec;62(605):e840-e850 [FREE Full text] [doi: 10.3399/bjgp12X659321] [Medline: 23211265]
42. Kivi M, Eriksson MC, Hange D, Petersson E, Björkelund C, Johansson B. Experiences and attitudes of primary care therapists in the implementation and use of internet-based treatment in Swedish primary care settings. Internet Interv 2015 Sep;2(3):248-256. [doi: 10.1016/j.invent.2015.06.001]
43. Whitfield G, Williams C. If the evidence is so good – why doesn't anyone use them? A national survey of the use of computerized cognitive behaviour therapy. Behav Cogn Psychother 2004 Feb 18;32(1):57-65. [doi: 10.1017/s1352465804001031]
44. Donovan CL, Poole C, Boyes N, Redgate J, March S. Australian mental health worker attitudes towards cCBT: What is the role of knowledge? Are there differences? Can we change them? Internet Interv 2015 Nov;2(4):372-381. [doi: 10.1016/j.invent.2015.09.001]
45. Carper MM, McHugh RK, Barlow DH. The dissemination of computer-based psychological treatment: a preliminary analysis of patient and clinician perceptions. Adm Policy Ment Health 2013 Mar;40(2):87-95. [doi: 10.1007/s10488-011-0377-5] [Medline: 22001968]
46. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]
47. Buti AL, Eakins D, Fussell H, Kunkel LE, Kudura A, McCarty D. Clinician attitudes, social norms and intentions to use a computer-assisted intervention. J Subst Abuse Treat 2013 Apr;44(4):433-437 [FREE Full text] [doi: 10.1016/j.jsat.2012.08.220] [Medline: 23021495]
48. Montero-Marín J, Prado-Abril J, Botella C, Mayoral-Cleries F, Baños R, Herrera-Mercadal P, et al. Expectations among patients and health professionals regarding Web-based interventions for depression in primary care: a qualitative study. J Med Internet Res 2015 Mar 10;17(3):e67 [FREE Full text] [doi: 10.2196/jmir.3985] [Medline: 25757358]
49. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1-14 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
50. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
51. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. Internet Interv 2017 Dec;10:29-38 [FREE Full text] [doi: 10.1016/j.invent.2017.09.001] [Medline: 30135750]
52. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J Consult Clin Psychol 2015 Dec;83(6):1108-1122. [doi: 10.1037/ccp0000044] [Medline: 26302248]
53. Wallin EEK, Mattsson S, Olsson EMG. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
54. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
55. Maslach C. Stress, burnout, and workaholism. In: Kilburg RR, Nathan PE, Thoreson RW, editors. Professionals in Distress: Issues, Syndromes, and Solutions in Psychology. Washington, DC: American Psychological Association; 1986:53-71.

XSL•FO
RenderX

Abbreviations

CBT: cognitive behavioral therapy
CFIR: Consolidated Framework for Implementation Research

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by H Hadjistavropoulos, L Cillessen; comments to author 10.03.20; revised version received 05.06.20; accepted 11.06.20; published 23.07.20.

Please cite as:
Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J
Implementing Web-Based Therapy in Routine Mental Health Care: Systematic Review of Health Professionals' Perspectives
J Med Internet Res 2020;22(7):e17362
URL: http://www.jmir.org/2020/7/e17362/
doi:10.2196/17362
PMID:32706713

©Fiona Davies, Heather L Shepherd, Lisa Beatty, Brodie Clark, Phyllis Butow, Joanne Shaw. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Digital Communication Technology to Increase HIV Testing Among Men Who Have Sex With Men and Transgender Women: Systematic Review and Meta-Analysis

Vanessa Veronese[1], PhD; Kathleen Elizabeth Ryan[1], PhD; Chad Hughes[1], MPH; Megan SC Lim[1], PhD; Alisa Pedrana[1], PhD; Mark Stoové[1], PhD

Disease Elimination Program, Burnet Institute, Melbourne, Australia

Corresponding Author:
Vanessa Veronese, PhD
Disease Elimination Program
Burnet Institute
85 Commerical Road
Melbourne, 3004
Australia
Phone: 61 0404387154
Email: vanessa.veronese@burnet.edu.au

Abstract

Background: HIV continues to disproportionately affect men who have sex with men (MSM) and transgender women (TW). Undiagnosed HIV is a major driver of HIV transmission rates, and increasing the uptake of regular HIV testing and facilitating timely initiation of HIV treatment is a global HIV prevention priority. However, MSM and TW experience a range of barriers that limit their access to testing and other prevention services. Given their growing ubiquity, digital communication technologies are increasingly being used to support HIV prevention efforts, and a growing number of studies have trialed the use of digital technology to promote HIV testing among MSM and TW.

Objective: We undertook a systematic review and meta-analysis to assess the impact of digital communication technology on HIV testing uptake among MSM and TW. Subanalyses aimed to identify the features and characteristics of digital interventions associated with greater impact.

Methods: A systematic literature review was undertaken using select databases and conference repositories. Studies describing the use of a digital technology—internet-enabled devices, including phones, tablets, and computers—to increase HIV testing uptake among MSM or TW using either randomized or observational cohort design with measurement of HIV testing rates measured pre- and postintervention, and published in English between 2010 and 2018 were included. Pooled effect estimates were calculated using a random effects meta-analysis. Subanalyses calculated effect estimates grouped by selected features of digital interventions.

Results: A total of 13 randomized or observational studies were included in the final review. Digital interventions most commonly used mainstream, existing social media platforms (n=7) or promotion through online peer educators (n=5). Most interventions (n=8) were categorized as interactive and allowed user engagement and most directly facilitated testing (n=7) either by providing self-testing kits or referral to testing services. A total of 1930 participants were included across the 13 studies. HIV testing uptake among MSM and TW exposed to digital interventions was 1.5 times higher than that of unexposed MSM and TW (risk ratio [RR] 1.5; 95% CI 1.3-1.7). Subanalyses suggested an increased impact on HIV testing uptake among interventions that were delivered through mainstream social media–based platforms (RR 1.7; 95% CI 1.3-2.1), included direct facilitation of HIV testing (RR 1.6; 95% CI 1.4-1.9), were interactive (RR 1.6; 95% CI 1.4-1.8), and involved end users in the design process (RR 1.6; 95% CI 1.3-2.0).

Conclusions: These findings provide broad support for the integration of technology with existing approaches to promote and facilitate HIV testing among MSM and TW. Our findings identified key features that may be associated with greater impact on HIV testing uptake and can be used to inform future development efforts given the growing interest and application of digital technologies in HIV prevention.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42017070055; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017070055.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e14230) doi:10.2196/14230

KEYWORDS

digital technology; men who have sex with men; transgender women; HIV testing; HIV prevention

Introduction

Globally, men who have sex with men (MSM) and transgender women (TW) are disproportionately affected by HIV [1-4]. Evidence of expanding epidemics among MSM have been noted, with new infections among this group comprising up to half of all incident cases in some regions, including North America [5-9], and up to approximately 70% in specific countries, such as the Philippines [7]. This burden of HIV occurs against a background of expanded access to HIV testing and treatment and the emergence and growing coverage of biomedical HIV prevention strategies [5].

The timely diagnosis of HIV plays an important role in preventing transmission, both by prompting reductions in risk behaviors to prevent onward transmission [10,11] and by facilitating early access to treatment and viral suppression. It is now known that people with suppressed HIV cannot transmit the virus to others [12-15]; promoting regular HIV testing is therefore a key global prevention strategy [16]. However, a range of barriers faced by MSM and TW limits their access to HIV testing and other preventive services. Stigma and discrimination remains a key deterrent to HIV service utilization in many parts of the world, particularly in settings where legislation prohibits same-sex relationships and legitimizes discriminatory behavior toward sexual minorities [17-19]. Other structural barriers, such as accessibility of services, costs, waiting time, and confidentiality concerns [20-24], also limit access to HIV testing among MSM and TW and prevent levels of testing coverage required to impact HIV incidence [25].

The use of digital communication technology—internet-enabled technology such as mobile phones, computers, and tablets that allow access to digital platforms and apps—to promote HIV testing among MSM and TW is an area of growing interest. Digital technology has assumed a prominent role in contemporary gay culture [26], and to a lesser extent in TW communities [27], as a tool to meet sex partners [28-30]. Evidence pointing to higher rates of condomless sex [31-33] and diagnoses of sexually transmitted infections (STI) [27,28,34,35] among MSM and TW who use online platforms to find sex partners suggests that online platforms may be appropriate targets for sexual health promotion. It has been suggested that internet-enabled technology can provide a discrete means for HIV health promotion in locations where same-sex behaviors are highly stigmatized or illegal [36], while the acceptability of HIV prevention interventions delivered through digital communication technologies has been noted among MSM and TW [37-43]. Together, these findings enhance the potential of using digital technology to reach those at high risk of HIV and serve as an important platform for health promotion outreach.

We performed a systematic review and meta-analysis to describe the impact of digital communication technology interventions on the uptake of HIV testing among MSM and TW. Subanalyses were undertaken to identify intervention characteristics associated with increased impact.

Methods

Overview

We systematically searched for current literature that describes the impact of digital communication technology interventions on the uptake of HIV testing among MSM and TW. We defined digital communication technology as technologies that were internet-enabled (through devices such as computers, mobile phones, or tablets) and which provided access to digital platforms such as social media sites, websites, apps, and email. The systematic review was conducted in accordance with the Preferred Reporting for Systematic Reviews and Meta-Analyses guidelines [44] and was registered on PROSPERO (registration number CRD42017070055).

Eligible studies were defined as those that:

1. Utilized at least one digital communication platform to deliver an intervention to promote HIV testing
2. Reported uptake of HIV testing as a result of a digital communication intervention for MSM and TW participants
3. Measured impact either by prospectively comparing testing rates pre- and postintervention exposure within a single cohort or through a randomized study design.

Studies that measured HIV testing outcomes but utilized digital communication technologies primarily for a purpose not directly related to improving HIV testing uptake (eg, to facilitate data collection or recruitment) and those reporting intention to test outcomes only were excluded.

Uptake of HIV testing was defined as any quantitative count of HIV testing events among MSM and TW measured using either self-report or clinic records.

For randomized controlled studies, we report on HIV testing uptake among participants in the intervention and control groups at the study endpoint. For nonrandomized studies, we report on HIV testing uptake at pre-intervention baseline and postintervention study endpoints.

Search Strategy

We conducted a systematic search of the literature published in English using the Medical Literature Analysis and Retrieval System Online, or MEDLARS Online (MEDLINE), Cumulative Index of Nursing and Allied Health Literature (CINHAL), EMBASE, PubMed, and PsychInfo databases. We limited our search to studies published between January 1, 2010, and May 1, 2018, to account for the redundancy of older platforms and technologies. Our search strategy comprised key terms; Medical Subject Headings (MeSH) terms; and subject headings related to participant (eg, *MSM*), intervention (eg, *internet*), and outcome (eg, *HIV testing and counseling*) variables (Multimedia

XSL•FO
RenderX

Appendix 1 shows the illustration of search strategy and MeSH terms). Electronic repositories of the International AIDS Society and the International AIDS Conferences were manually searched for abstracts from January 2010 onward. Reference sections of the identified papers were also searched for additional papers. No restrictions were placed on participant age or other demographic characteristics of the study population or geographical location or setting of the intervention.

Data Extraction

Following the literature search, the first author (VV) removed duplicate records and assessed the remaining abstracts on the basis of the eligibility criteria. The second author (KR) reviewed a random sample (equivalent to 10%) of discarded abstracts to ensure accuracy. Both authors then conducted a full-text review of the remaining abstracts to determine their inclusion in the final analysis.

The following domains were extracted for final analysis by two authors (VV and KR) using a standardized, Excel-based tool: study identification, study design (data collection period, recruitment, and sampling method), intervention characteristics (aim of intervention, mode of intervention delivery, and duration), study population (inclusion and exclusion criteria, sample size, primary HIV testing outcome used and time frame for the outcome, the mean age of sample, proportion reporting previous testing), and results and analysis (number of testing outcomes, effect size measurement, and reported effect size).

Any discrepancies identified between the two authors during the review and data extraction process were discussed with a third author (MS).

Analysis

Qualitative Synthesis of Study Aims, Intervention Characteristics, and Study Outcomes

A qualitative synthesis was undertaken to characterize included studies by their study participants, location of study, study design, digital communication intervention platform, intervention features, length and frequency of exposure, sample size, and rated study quality. For each study, we then described the proportion of MSM and TW reporting previous HIV testing at baseline, how testing uptake was defined and measured during the study period, whether testing was provided or offered as part of the intervention, and the proportion of MSM and TW reporting or receiving HIV testing at the end of the study period.

The primary outcome of this study was uptake of HIV testing among MSM and TW participants, which we defined as the number of individual MSM and TW reporting or receiving an HIV test divided by the total number of MSM and TW exposed to a digital intervention. Intervention effectiveness was determined by manually calculating risk ratios (RRs) comparing testing uptake between exposed groups and unexposed groups for randomized controlled trials (RCTs) or between pre- and postintervention commencement for nonrandomized studies. HIV testing uptake among controls or at preintervention time point was used as the reference group so that an RR greater than one demonstrates a higher chance of HIV testing uptake following exposure to a digital communication intervention.

Meta-Analysis

We performed a meta-analysis of intervention effectiveness to generate pooled RR using a DerSimonian and Laird random effects model to account for the anticipated heterogeneity between studies. We identified a range of characteristics common to digital communication interventions or regarded as features that potentially enhance the intervention effectiveness and performed submeta-analyses to generate pooled RR to examine the impact of these characteristics on overall intervention effectiveness:

1. Intervention interactivity (yes/no): interventions that permitted end users to interact or engage, for example, by chatting with peer educators or other participants, as opposed to passive viewing of an online video
2. End-user involvement in the design process (yes, no, or not reported): any reported involvement of the intended end users in the intervention design process (eg, consultation and pilot testing)
3. HIV testing facilitated as part of the intervention (yes/no): interventions that directly provided (eg, provided self-tests) or facilitated (eg, direct referral) HIV testing to participants, as opposed to simply promoting HIV testing
4. Social media platform (yes/no): interventions implemented through an established social media platform (eg, Facebook) that facilitates social networking among the general population or specifically among gay and other MSM, as opposed to nonsocial media platforms
5. Single dose exposure (yes/no): interventions that delivered a single, time-bound exposure as opposed to multiple exposures over time.

The presence and magnitude of heterogeneity were assessed in meta-analyses using the χ^2 and I^2 tests, respectively. All statistical analyses were performed using Stata version 14 (StataCorp LP).

Assessing Study Quality

Study quality was assessed by two authors (VV and KR) using the quality assessment tool for quantitative studies [45]. This tool critiqued studies on the basis of selection bias, study design, confounders, blinding, data collection methods, and withdrawal and dropouts. Each criterion was rated as strong, moderate, or weak. On the basis of the combined scores, studies were given a final, global rating of strong (no weak ratings), moderate (one weak rating), or weak (two or more weak ratings). As the final global rating was determined by the number of subcategories scored as *weak*, the tool was modified to include an *N/A* option for blinding and withdrawal and dropout fields to account for nonrandomized studies. The first and second authors (VV and KR) met after completing the first two quality assessments to ensure consistency in the use of the tools. After consistency was confirmed, the two authors completed the remaining quality assessments, and the final results were compared and discussed. Similar to the process for data extraction, any disagreements between the assessments of the first and second author were discussed with a third senior author (MS).

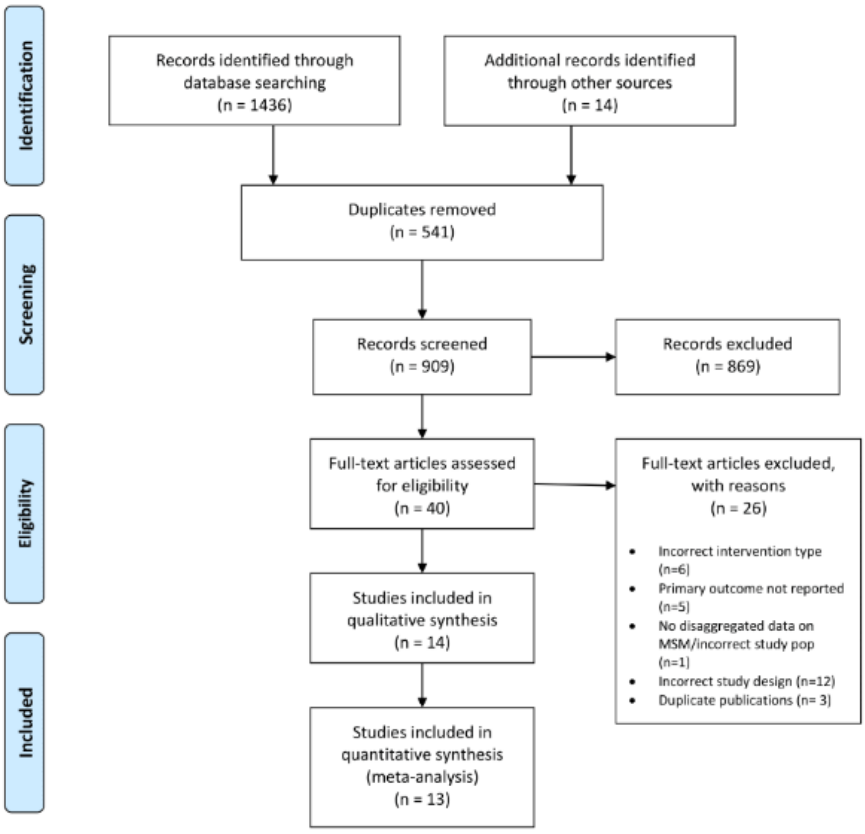
XSL•FO
RenderX

Results

Search Results

The systematic literature search resulted in 1436 identified records, including 8 conference abstracts and 6 papers identified in the references of the included studies. 37.30% (541/1436) were removed as duplicates. The remaining 909 records were reviewed at the abstract level, 4.4% (40/909) of which were retained for full-text review. Thirteen papers were included in the final analysis [46-58] (Figure 1).

Figure 1. The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) flow diagram depicting study screening and selection.



Study and Sample Characteristics

Table 1 presents the characteristics of the included studies. All included studies specifically targeted men who identified as gay or self-reported anal sex with a male partner. Two studies included TW [53,54]. Study participants were typically aged in their midtwenties, and the overall level of educational attainment varied substantially across studies. On the basis of the World Bank classifications [59], the majority of studies took place in high-income countries (Hong Kong, n=1; Taiwan, n=1; and United States, n=6), 4 occurred in upper-middle-income countries (China, n=1 and Peru, n=3), and 1 in a low-middle-income country (India). The studies collectively included 8875 participants, 64 (0.70%) of whom were reported as TW, with a mean sample size of 682 participants (SD 808; range 56-3092 participants). Two studies excluded participants with a history of HIV testing, and two studies did not report on participants' previous testing history. Among the 9 remaining studies, the mean proportion of participants reporting a history of HIV testing was 44.1% (SD 19.0; range 21.3%-73.8%). The recall period for these studies varied, with one study reporting lifetime HIV testing history and the remaining studies using recall periods that ranged from 3 months to 3 years. A total of 10 studies used an RCT design [46-49,51,54-58]. The specific study design used by the remaining 3 studies was a prospective cross-sectional study with nonequivalent control group [50], a prospective single-group cross-sectional study [52], and a prospective cross-sectional matched-pair randomized trial design [53]. Due to the lack of individual-level randomization and nonequivalence of control groups, we calculated relative risk for these 3 studies using reported HIV testing uptake at preintervention baseline and postintervention endpoints in the intervention arm only. One RCT described two intervention arms and did not present disaggregated findings; therefore, this study was treated as a single cohort, and relative risk was calculated based on HIV testing uptake at baseline and endpoint [51].

Aims of Included Studies

Promoting HIV testing uptake was identified as the primary aim of all 13 studies. Five studies had multiple primary aims: mutual disclosure of HIV status with sexual partners (ie, asking and disclosing HIV serostatus [49]), STI testing [46], intention to test [47,51], and reduction of unprotected anal intercourse [49,50] (data not reported).

XSL•FO
RenderX

Table 1. Overview of included studies.

Reference	Description of study participants	Location (country of income classification[a])	Total sample size	TW[b] participants, n (%)	Percentage of participants reporting previous HIV testing preintervention (time frame for testing history)	Study design
Bauermeister et al (2015) [46]	Inclusion criteria: Young MSM[c] aged 15-24 years, self-identified cis male, reported sex with a male partner in the past 6 months; sample characteristics: mean age 21 years, 92.3% educated to high school or GED[e] level	Michigan, United States (high)	130	0 (0)	73.8 (lifetime)	RCT[d]
Blas et al (2010) [47]	Inclusion criteria: MSM aged 18 years and above, reporting lifetime sex with another man, and not reporting testing within the past 12 months; sample characteristics: mean age 26.1 years (range 18-61), 42% educated to university or technical graduate level	Lima, Peru (upper middle)	459	0 (0)	21.3 (more than 1 year ago)	RCT
Blas et al (2014) [48]	NR[f]	Peru (upper middle)	400	0 (0)	NR	RCT
Hirshfield et al (2012) [49]	Inclusion criteria: male aged 18 years or above, reporting oral or anal sex with a current male partner and oral, anal, or vaginal sex with at least one new partner (male or female) in the past 6 months; sample characteristics: median age 39 years (range 18-81), 55% educated to college degree or higher level	United States (high)	3092	0 (0)	69.0 (past 3 years)	RCT
Ko et al (2013) [50]	Inclusion criteria: MSM aged 18 years and above, reported sex with another man in the past 12 months; sample characteristics: mean age 24.8 years, 63% educated to college level	Taiwan (high)	1037	0 (0)	29.4 (past 6 months)	Prospective cross-sectional study with nonequivalent control
Patel (2016) [51]	Inclusion criteria: MSM aged 18 years or older living in Mumbai	Mumbai, India (low middle)	244	0 (0)	61.5 (past 6 months)	RCT
Rhodes et al (2011) [52]	Inclusion criteria: male, registered user of chat room servicing MSM in North Carolina, United States; mean age 37 years; gay-identifying; sample characteristics: mean age 37.1 years (range 18-71 years; education not reported)	North Carolina, United States (high)	346	0 (0)	44.5 (past 3 years)	Prospective cross-sectional study
Rhodes et al (2016) [53]	Inclusion criteria: male, social media user, mean age 40 years, gay-identifying; sample characteristics: mean age 40.9 years (range 18-74 years; education not reported)	United States (high)	1292	28 (2.17)	36.6 (past 12 months)	Prospective cross-sectional matched-pair randomized trial design
Tang et al (2016) [54]	Inclusion criteria: male aged 16 years or above, reporting lifetime anal sex with another man and no HIV testing history; sample characteristics: 37% aged between 21 and 25 years, 65% college educated	China (upper middle)	721	36 (5.0)	N/A[g]	RCT
Washington et al (2017) [55]	Inclusion criteria: Black or African American male aged 18-30 years reporting sex with another man during the past 3 months and not tested within the past 6 months; sample characteristics: mean age 23.1 years, 52% educated to high school or GED level	Los Angeles, United States (high)	56	0 (0)	46.5 (past 12 months)	RCT

XSL•FO
RenderX

Reference	Description of study participants	Location (country of income classification[a])	Total sample size	TW[b] participants, n (%)	Percentage of participants reporting previous HIV testing preintervention (time frame for testing history)	Study design
Wang et al (2018) [56]	Inclusion criteria: Chinese-speaking males in Hong Kong aged 18 years or over reporting anal sex with 1 or more male partner in the past 6 months with access to online live chat apps (Line, WhatsApp, and Skype); sample characteristics: 63.1% aged between 18 and 31 years, 81.4% educated to a university level	Hong Kong, China (high)	430	0 (0)	N/A	RCT
Young et al (2013) [57]	Inclusion criteria: African American or Latino male aged 18 years or above, registered Facebook users, reporting sex with another man in the past 12 months; sample characteristics: mean age 31.8 years (SD 10.2 years), 36.4% educated to high school level	Los Angeles, United States (high)	112	0 (0)	NR	RCT
Young et al (2015) [58]	Inclusion criteria: MSM aged 18 years and over, reporting sex with another man during the past 12 months; sample characteristics: 28.9 mean age (SD 7.9 years), 37.8% educated to vocational school level	Lima, Peru (upper middle)	556	0 (0)	33.4 (past 3 months)	RCT

[a]Classification based on World Bank countries and lending groups [59].

[b]TW: transgender women.

[c]MSM: men who have sex with men.

[d]RCT: randomized controlled trial.

[e]GED: general education development.

[f]NR: not reported.

[g]N/A: not applicable.

Intervention Characteristics of the Included Studies

Table 2 presents the characteristics of the digital interventions among the included studies. Studies utilized a variety of digital communication platforms to deliver interventions. Six used only social media platforms, 3 used only online videos, and 1 delivered a tailored online HIV/STI testing intervention through a customized website. The remaining 3 studies used multiple platforms: 1 used online videos in conjunction with motivational messages sent via email or instant messaging, 1 used social media and live chat apps, and 1 used online videos in conjunction with live chat apps.

Of the 7 studies that used social media, 5 delivered interventions using Facebook [50,51,55,57,58]. All of these studies used closed or private Facebook groups to promote HIV testing, using a range of accompanying features and modalities using internet popular opinion leaders to disseminate HIV-related information and engage in conversations about HIV testing, prevention, and risk behavior [50]; using trained peer educators to deliver HIV prevention and testing information and promote HIV testing uptake [55,58], including through the provision of HIV self-tests [57]; delivering weekly videos promoting HIV testing alongside moderated group discussion [47]; and sending HIV prevention and HIV testing promotion messages (in conjunction with messages sent by WhatsApp and email) [51].

Two social-media-based interventions were delivered through sites specifically targeting the gay community [52,53]. Both reported on the Cyber-Based Education and Referral/testing (CyBER/testing) intervention, which utilized trained peers to promote HIV testing through existing social and sexual networking sites popular among the gay community. CyBER/testing was implemented through an existing chat room used by MSM [52] and later through four geographically focused social media sites used by MSM and TW [53], both in the United States.

Only one intervention was delivered through a custom-built website (Get Connected!), which aimed to promote and connect HIV and STI testing to young MSM [46]. The website delivered customized content to participants based on sociodemographic, sexual identity, and behavior and previous engagement with HIV testing data provided during a baseline assessment, so that messaging and content were personalized to mirror participant profiles, and experiences with testing, including motivations and perceived barriers.

XSL•FO
RenderX

Four studies used online videos to promote HIV testing [47-49,54]. One used single-session online videos focusing on HIV prevention to motivate HIV testing among MSM in the United States, in which participants received either dramatic- or documentary-style videos, or both [49]. A 2010 intervention delivered one 5 min video promoting HIV testing customized based on self-reported sexual identity (gay or nongay identified) to MSM in Peru [47]. In 2014, this intervention was repeated with the addition of motivational messages sent by email or instant messaging to encourage HIV testing to MSM in Peru [48]. Finally, online videos developed with crowdsourced content depicted two Chinese men falling in love and getting tested together to target MSM and TW naïve testers in China [54]. One study used online videos in combination with live chat apps to deliver a home-based HIV self-testing service to MSM in Hong Kong [56]. In this study, all participants were exposed to an online video promoting HIV testing, and then, the participants in the intervention group were offered a home-based self-testing kit and online, real-time HIV pre- and posttest counseling and instruction through Line, WhatsApp, or Skype.

Regarding the length of intervention and the frequency of exposure, interventions that used online videos were typically shorter in duration and involved a single exposure to a video [47,54,56]. Two studies sent multiple videos [49,55] and one combined videos and motivational messages via text, email, or instant messaging but did not report on the length or number of videos and frequency of motivational messages [48]. The customized website-based intervention [46] also provided a single exposure to the website content, but data on the length of time spent on the website were not reported.

Most interventions (n=8) were categorized as interactive for their ability to allow user engagement, for example, through social media platforms, responses on messaging platforms or through email, or via a website. Interventions involving online peer educators interacting with participants through social media platforms were typically longer in duration, ranging from 12 weeks [51,57,58] to 12 months [53]. The frequency of exposure to online peer educators was user-determined, that is, participants were free to choose how often, if at all, they interacted with online peer educators. The exception involved participants receiving twice weekly messages through Facebook, email, or instant messaging, over a 12-week period [51].

Seven studies directly facilitated HIV testing, through the provision of HIV home or self-testing kits [56,57]; referrals to specific, local HIV testing clinics [47,48,58]; or providing location details of free, local HIV testing sites [51,55]. The remaining six studies provided general promotion of HIV testing only.

Five studies specifically mentioned the involvement of intended end users in the implementation design process [46,47,52,55,56]. Theoretical underpinning to intervention development was described by 9 studies [46,47,49,51-53,55,56,58] (Table 2).

XSL•FO
RenderX

Table 2. Characteristics of digital interventions of included studies.

Reference	Platform	Description	Interactive	Intervention length (frequency of exposure)	HIV testing provided or facilitated	Measurement of HIV testing uptake used	Length of follow-up period for outcome	Comparator	Involvement of end users in intervention design	Theoretical framework
Bauermeister et al (2015) [46]	Website	Interactive, customized website (Get Connected!) that delivered HIV/STI[a] testing and prevention content tailored to specific participant profiles of based on psychosocial data and previous engagement with HIV testing	Yes	One time	No	Self-reported	30 days	Control group: test-locator website	Yes	Self-determination theory principles and integrated behavioral model
Blas et al (2010) [47]	Online video	One 5-min video delivered through existing gay and commercial websites promoting HIV testing customized based on self-identification of participant as either gay or nongay	No	One time (5 min)	Yes—facilitated (referral)	Attendance based	125 days[b]	Control group: standard public health text	Yes	Health belief model
Blas et al (2014) [48]	Multiple: online videos, email/instant messaging	Motivational videos and messages about HIV testing sent through email and instant messaging, respectively	No	NR[c]	Yes—facilitated (referral)	Attendance based	184 days[b]	Control group: health promotion message with invitation for free HIV testing	No or not reported	None reported
Hirshfield et al (2012) [49]	Online video	HIV prevention videos in either dramatic or documentary style (or both), accessed via banner ads on gay-oriented sexual networking sites, and designed to promote critical thinking about HIV disclosure, testing, and condom use	No	One time (9 and 5 min)	No	Self-reported	60 days	Control group: no content	No or not reported	Social learning theory
Ko et al (2013) [50]	Social media (Facebook)	Trained internet popular opinion leaders promoting HIV testing and prevention to members of a closed Facebook group	Yes	6 months (user-dependent)	No	Self-reported	6 months	Baseline	No or not reported	None reported

XSL•FO
RenderX

Reference	Platform	Description	Interactive	Intervention length (frequency of exposure)	HIV testing provided or facilitated	Measurement of HIV testing uptake used	Length of follow-up period for outcome	Comparator	Involvement of end users in intervention design	Theoretical framework
Patel (2016) [51]	Multiple: social media (Facebook); online live chat apps (WhatsApp); email	16 health promotion messages promoting HIV testing framed in either approach or avoidance style of messaging sent by trained peers via their preferred modality (private Facebook group, individual WhatsApp messaging, or email)	No	12 weeks (twice weekly)	Yes—facilitated (test locator)	Self-reported	12 weeks	Baseline	No or not reported	Information motivation behavioral skills model
Rhodes et al (2011) [52]	Social media (MSM[d]-specific sites)	Trained peer posting regular triggers about HIV and HIV testing in existing chat room used by gay and other MSM and engaging in direct communication about testing services, processes, and locations with chat room users	Yes	6 months (daily)	No	Self-reported	6 months	Baseline	Yes	Natural helping
Rhodes et al (2016) [53]	Social media (MSM-specific sites)	Trained peer posting regular triggers in four existing social media sites used by gay and other MSM about HIV and HIV testing and engaging in direct communication with users about testing services, processes, and locations	Yes	12 months (daily)	No	Self-reported	12 months	Baseline	No or not reported	Empowerment education, social cognitive theory, and natural helping
Tang et al (2016) [54]	Online video	Online video promoting HIV testing based on a crowdsourced design accessed via banner ads placed on gay-oriented social networking platforms	No	4 weeks (one time)	No	Self-reported	3 weeks	Control group: noncrowd sourced online video (standard public health text)	Yes	None reported
Washington et al (2017) [55]	Social media (Facebook)	Five, 1-min long videos promoting HIV testing sent through a private Facebook group to black or African American MSM, with moderated group discussion	Yes	6 weeks (weekly)	Yes—facilitated (test locator)	Self-reported	6 weeks	Control group: closed Facebook group receiving generic health information	Yes	Integrative model of behavior change

XSL•FO
RenderX

Reference	Platform	Description	Interactive	Intervention length (frequency of exposure)	HIV testing provided or facilitated	Measurement of HIV testing uptake used	Length of follow-up period for outcome	Comparator	Involvement of end users in intervention design	Theoretical framework
Wang et al (2018) [56]	Multiple: online videos; online live chat apps (Line, WhatsApp, and Skype)	Home-based self-testing service comprising online promotional video about HIV testing, plus additional videos on home-based HIV self-testing and offer of free HIV self-testing kit and online real-time instructions and pre- and posttest counseling provided via live chat apps	Yes	6 months (one time)	Yes—provided (HIV self-testing)	Self-reported or observed uptake of self-testing	6 months	Control group: online video about (general) HIV testing only	Yes	Health belief model
Young et al (2013) [57]	Social media (Facebook)	Trained peer educators providing HIV prevention and testing messages, including 4 weekly reminders about availability of HIV home testing, to participants of a closed Facebook group	Yes	12 weeks (user -dependent)	Yes—provided (HIV self-testing)	Requested and returned home-based HIV testing kit and followed-up results	12 weeks	Control group: closed Facebook group receiving per-delivered generic health information	No or not reported	None reported
Young et al (2015) [58]	Social media (Facebook)	Trained peer educators providing HIV prevention and testing messages, including 4 weekly reminders about availability of HIV home testing, to participants of a closed Facebook group	Yes	12 weeks (user-dependent)	Yes—facilitated (referral)	Attendance based	12 weeks	Control group: closed Facebook group providing HIV testing information without peer leaders	No or not reported	Diffusions of innovation theory and social normative theory

[a]STI: sexually transmitted infection.

[b]Reported as the average follow-up time.

[c]NR: not reported.

[d]MSM: men who have sex with men.

Outcomes of the Included Studies

Table 3 presents the reported outcomes of studies using the RCT study design. Most RCT studies measured testing uptake through self-reports [46,49,54,55]. The length of follow-up over which testing was measured varied greatly from 3 weeks [54] to 6 months [56].

On the basis of the calculated RRs, 3 of 9 RCTs demonstrated a significant improvement in HIV testing uptake [55,56,58], whereas 4 demonstrated nonsignificant improvements [46,47,49,54]. The two remaining RCTs did not demonstrate any impact; in one RCT, testing uptake was extremely low in both intervention and control arms [48], and in another, no members of the control group tested for HIV during follow-up [57].

XSL•FO
RenderX

Table 3. Reported HIV testing outcomes in included randomized controlled trial studies.

Reference	Control group		Intervention group		Risk ratio (95% CI)
	Total number of participants	Participants tested, n (%)	Total number of participants	Participants tested, n (%)	
Bauermeister et al (2015) [46]	36	4 (11)	68	18 (26)	2.4 (0.9-6.5)
Blas et al (2010) [47]	220	10 (4.5)	239	19 (7.9)	1.7 (0.8-3.7)
Blas et al (2014) [48]	200	3 (1.5)	200	2 (1.0)	0.7 (0.1-3.9)
Hirshfield et al (2012) [49]	240	48 (20.0)	676	142 (21.0)	1.1 (0.8-1.4)
Tang et al (2016) [54]	317	111 (35.0)	307	114 (37.1)	1.1 (0.9-1.3)
Wang et al (2018) [56]	215	109 (50.7)	215	193 (89.8)	1.8 (1.5-2.0)
Washington et al (2017) [55]	22	8 (36)	20	16 (80)	2.2 (1.2-4.0)
Young et al (2013) [57]	55	0 (0)	57	8 (14)	N/A[a]
Young et al (2015) [58]	246	16 (6.5)	252	43 (17.1)	2.6 (1.5-4.5)

[a]N/A: not applicable.

Table 4 presents the outcomes from nonrandomized studies. All of the 4 nonrandomized studies relied on self-reported HIV testing uptake. The length of follow-up ranged from 12 weeks [51] to 12 months [53], and all 4 studies demonstrated significant improvements in testing uptake based on calculated RRs (Table 4).

Across all 13 studies included in this review, 21.75% participants (1930/8875) received an HIV test during a cumulative 3.6 years of study follow-up. Three studies reported on HIV diagnoses (all RCTs) [47,54,56], with 75 new HIV infections detected across these studies (56% in the intervention arm; data not reported).

Table 4. Reported HIV testing outcomes in included quasi-experimental studies.

Reference	Baseline		End line		Risk ratio (95% CI)
	Total number of participants	Participants tested, n (%)	Total number of participants	Participants tested, n (%)	
Ko et al (2013) [50]	501	150 (29.9)	499	219 (43.9)	1.5 (1.2-1.7)
Patel (2016) [51]	130	42 (32.3)	130	57 (43.8)	1.4 (1.0-1.9)
Rhodes et al (2011) [52]	346	154 (44.5)	315	187 (59.4)	1.3 (1.4-1.5)
Rhodes et al (2016) [53]	353	122 (34.6)	399	216 (54.1)	1.6 (1.3-1.9)

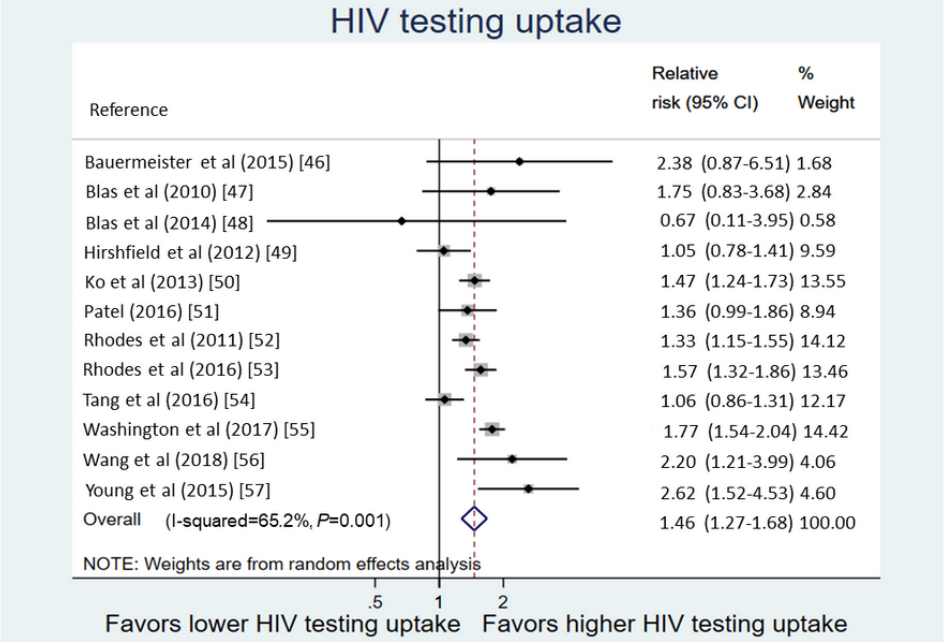
Meta-Analysis

Primary Outcome

The pooled RR across 12 studies (RR could not be calculated for one study because no tests were recorded in the control arm) [57] indicated a significant increase in the uptake of HIV testing following exposure to digital interventions (RR 1.5; 95% CI 1.3-1.7). Statistical heterogeneity was high (χ^2_{11}=31.7; I^2=65.2%; Figure 2).

XSL•FO
RenderX

Figure 2. Forest plot of relative risk of HIV testing uptake among MSM and TW following digital intervention exposure.



Subanalysis of the Study Characteristics

A positive impact was seen on the HIV testing uptake across all intervention type subanalyses. The highest pooled RR was seen for interventions that were delivered through the mainstream social media–based platforms (RR 1.7; 95% CI 1.3-2.1), interventions that included direct facilitation of HIV testing (RR 1.6; 95% CI 1.4-1.9), interventions that were interactive (RR 1.6; 95% CI 1.4-1.8), and interventions that involved end users in the design process (RR 1.6; 95% CI 1.3-2.0; Table 5).

XSL•FO
RenderX

Table 5. Subanalyses by selected study and intervention characteristics.

Intervention characteristic	k[a]	Risk ratio (95% CI)	χ^2 (*df*)	I^2
Overall effect size	12	1.5 (1.3-1.7)	31.7 (11)	65.2
Randomized controlled trials only	8	1.6 (1.2-2.1)	29.1 (7)	76.0
Quasi-experimental	4	1.4 (1.3-1.6)	2.1 (3)	0
Direct facilitation of HIV testing				
Yes	7	1.6 (1.4-1.9)	13.8 (6)	65.2
No	5	1.3 (1.0-1.6)	9.4 (4)	58.1
Interactive intervention				
Yes	7	1.6 (1.4-1.8)	15.0 (6)	46.7
No	5	1.1 (1.0-1.3)	1.7 (4)	65.2
User involvement in design				
Yes	5	1.6 (1.3-2.0)	9.4 (4)	57.6
No or not reported	6	1.4 (1.1-1.6)	18.0 (5)	66.7
Theoretical basis to intervention				
Yes	9	1.6 (1.3-1.8)	20.5 (8)	61.0
No or not reported	3	1.2 (0.9-1.7)	6.2 (2)	67.5
Social media–based intervention				
Yes—general	4	1.7 (1.3-2.1)	6.0 (3)	49.8
Yes—gay oriented	2	1.4 (1.2-1.7)	2.0 (1)	48.7
No	6	1.4 (1.0-1.9)	23.1 (5)	78.4
Single-dose intervention				
Yes	5	1.4 (1.0-1.9)	22.4 (4)	82.1
No	6	1.5 (1.3-1.7)	8.5 (5)	41.2

[a]k: number of studies included in the subcategory.

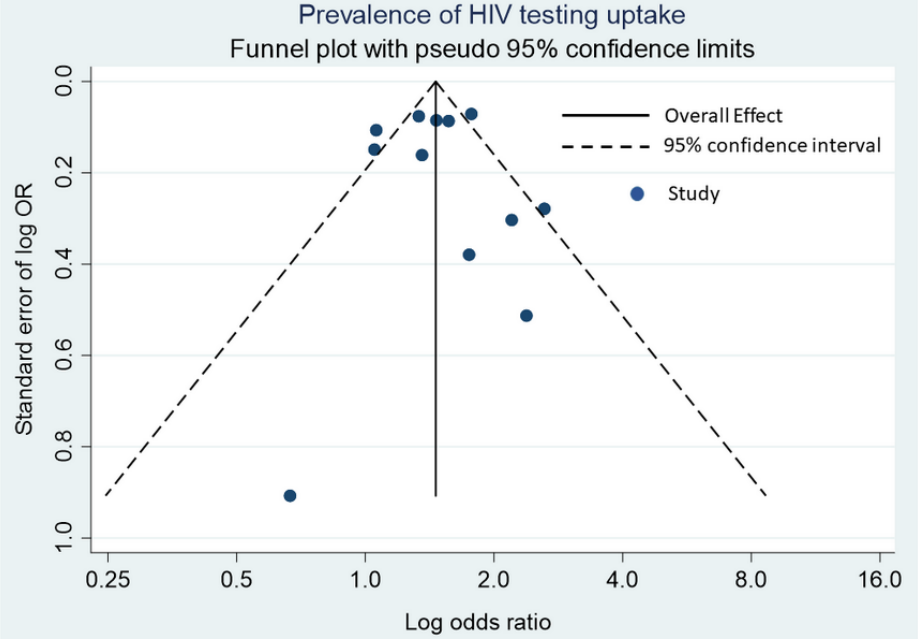
Study Quality

The majority of studies (n=8) were classified as *moderate* quality [46,47,52-56,58], 4 were classified as *weak* [48-51], and 1 was classified as *strong* [57]. The most common limitation across studies was insufficient description of blinding procedures (rated as *weak* in 9 studies), whereas controlling for confounding in either study design or analyses was a common strength (rated as *strong* in 9 studies; Multimedia Appendix 2).

The observed asymmetry in the study funnel plot (Figure 3) may be explained by the heterogeneity between studies, given the variability in the intervention design. The attribution of asymmetry to heterogeneity is also supported by the high level of variance between studies (I^2=65.2%).

XSL•FO
RenderX

Figure 3. Funnel plot for estimating publication bias and precision of estimate.



Discussion

In this systematic review, exposure to digital communication interventions was associated with greater HIV testing uptake among MSM and TW compared with those unexposed to digital communication interventions. Our findings provide broad support for the integration of technology into existing tools and approaches to HIV prevention among MSM and TW. We extend the current state of evidence regarding the impact of digital interventions on HIV testing by examining the role of key intervention features on effectiveness. Our findings identified key features that were associated with greater uptake for HIV testing for MSM and TW—specifically, interventions that facilitate HIV testing through provision or direct referral to HIV testing services, the use of mainstream social media platforms to engage the target population, interactivity, and involvement of end users in design processes.

Global guidance has recommended the integration of digital technology as a strategy to enhance the reach and effectiveness of HIV prevention efforts among MSM and TW [36,60]. Numerous acceptability studies have identified that MSM would be willing to use phone- and web-based technologies for HIV prevention [39,40,61,62]. Recent systematic reviews have explored the potential of digital technology to advance various HIV priorities among MSM and other key populations [63,64]. One systematic review to date has attempted to quantify the role of digital communication technologies in increasing engagement among key populations across the care continuum through meta-analysis [65]; this study specifically looked at randomized and observational studies describing the impact of social media–based interventions on HIV testing uptake among key populations. All 9 studies included in this meta-analysis targeted MSM and were collectively associated with an approximately 50% increase in HIV testing uptake. Compared with this review, our study adopted a broader definition of digital interventions that encompassed websites, online videos, instant messaging, and live chat apps, resulting in the inclusion of 6 additional studies. Overall, we found that exposure to digital interventions was associated with an approximately 50% increase in HIV testing uptake, in line with the previous study's estimates, supporting the conclusion that digital communication technologies are effective in promoting HIV testing among MSM and TW. Interventions included in this review invariably took advantage of various online engagement tools such as videos, digital messages tailored to participant profiles, and online social networking and compared the outcomes with more generic and less interactive online messaging. The impact of digital communication technology to increase testing uptake may be even greater when compared with more traditional health promotion mediums such as messaging through public media (such as posters and billboards). In addition, our study identified specific features of digital interventions associated with greater impact on HIV testing uptake among MSM and TW.

First, interventions that went beyond creating general demand for testing, either through health promotion or providing educational content, and instead directly facilitated HIV testing through service referrals or the provision of self-tests demonstrated greater impact on HIV testing uptake compared with the overall estimate. Direct facilitation of HIV testing may potentially address some of the more structural barriers MSM and TW face to HIV testing [17,66,67]. In particular, HIV self-testing has emerged as a strategy to mitigate barriers related to HIV service access [68,69]; however, some MSM and TW populations have expressed concerns about the limited availability of support during the testing process [70,71]. The study by Wang et al [56] included in this review suggests that digital interventions can play a role in promoting HIV self-testing by mitigating user concerns through the provision of real-time, online counseling. It should also be noted that the majority of studies included in this review recruited participants with generally high levels of previous HIV testing behaviors, which may be critical to the success of this approach. Past research has shown that a history of testing is a strong predictor of future testing behaviors among MSM [72-74], suggesting

XSL•FO
RenderX

that the direct facilitation of testing through digital communication interventions may work best for those who are experienced in HIV testing and may enable participants to access more frequent testing.

Second, interactive interventions—those that allowed participants to engage directly with online content or other users—demonstrated greater impact on HIV testing uptake compared with the overall pooled estimate. Interactivity in digital interventions has been associated with achieving a greater impact on behavior change across a range of health areas [75-77] and has been identified as a desired feature of digital HIV prevention interventions among MSM and TW [43,78]. Third, our subanalysis also identified that interventions that used mainstream social media platforms, such as Facebook, were also associated with greater uptake of HIV testing. However, all interventions that utilized social media platforms were also categorized as interactive, making it difficult to isolate the source of the enhanced effect. Social networking–based interventions are commonly used for sexual health promotion [79], and the use of existing social media platforms has been identified as a way to enhance retention among young MSM and TW in online HIV prevention activities [43]. Using existing and well-utilized social media platforms may also enhance the reach of digital HIV prevention interventions [80] compared with those that are delivered through new or separate platforms. Interestingly, interventions that used social media sites specifically for gay and other MSM were less effective in increasing HIV testing uptake than interventions delivered through general social media. Others have noted the reluctance of users of gay social networking sites to receive health promotion messages, which are often seen as an intrusion or surveillance and may limit user engagement [81].

Fourth, evidence of enhanced impact was found among digital interventions that reported involvement of end users in the design process. This involvement is a key component of user-centered design, an approach that prioritizes user needs and experiences to maximize functionality and increase engagement and relevance to the target population [82-84]. User involvement is particularly important when developing digital interventions tailored to specific target populations to ensure that such interventions appropriately reflect group priorities, preferences, and culture [83]. Although the literature confirms the value of user-centered design, our review only assessed studies on whether any involvement of end users in the design process was reported; however, it is probable that the quality and depth of this involvement may be a stronger determinant of overall effectiveness. In addition, grounding in theoretical frameworks may also be another indicator of effectiveness, as suggested by the greater uptake of HIV testing reported by theoretically based interventions included in this review. Although the majority of studies in this review reported a theoretical basis for their intervention, the limited number of theory-based, HIV-focused digital interventions have been noted by others [85,86]; this may be attributed to the speed of development and proliferation of digital approaches to improving HIV outcomes. The findings presented here suggest that theoretical grounding is an important component of effective interventions and should be prioritized in future development.

The findings of this review should be considered with the following limitations. First, due to the restrictions we placed on study design, the interventions included in this review reflect only those conducted as research projects and may not reflect the real-world application of digital technology, including interventions that were not formally evaluated or represented in the published literature. Second, despite our finding that the digital interventions included in this review had a positive overall effect on HIV testing uptake among MSM and TW, the findings do not necessarily reflect the actual quality of the content delivered or levels of end-user acceptability, which are likely to interact in important ways with intervention impact. Third, TW participants were underrepresented in the included studies. TW may use social and sexual networking apps less frequently than MSM, which may reflect the limited number of social and sexual networking sites specifically catering to TW relative to MSM [87]. However, factors such as TW's reported reliance on online sources of sexual health information [88], perceived acceptability of digital approaches to HIV prevention [43], and examples of real-world applications of digital communication technologies to HIV prevention among TW [87] suggest that TW also stand to benefit from digital approaches to HIV prevention and their inclusion in future trials should be prioritized. Finally, the majority of studies were conducted in high- or middle-income settings. Although the use of digital technology to advance HIV prevention priorities in low-resource settings has been both recommended and applied [64,89], further research is warranted to assess the impact of digital interventions of HIV testing among MSM and TW in these settings.

HIV testing is a key focus of global HIV prevention efforts among MSM and TW, yet multiple barriers continue to prevent levels and frequency of testing required to facilitate the early detection of undiagnosed HIV and initiation of treatment. Digital communication technologies are now an accepted medium for HIV prevention efforts; this review provides further evidence of the role of such technologies in increasing HIV testing uptake among MSM and TW. The inclusion of intervention features such as direct facilitation of HIV testing, involvement of end users in the design process, interactivity, and delivery through the existing mainstream social media platforms may enhance the overall impact and maximize the contribution of digital communication technologies to advancing HIV prevention priorities among MSM and TW.

Acknowledgments

This study formed part of the PhD of VV, who was funded through an National Health and Medical Research Council post graduate scholarship.

XSL•FO
RenderX

Authors' Contributions

VV, MS, ML, and MP were responsible for study design; VV and KR conducted the literature review and data extraction; and VV analyzed data and developed the manuscript. KR, ML, AP, CH, and MS provided input into manuscript development. All authors have read and approved the final manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ovid Medical Literature Analysis and Retrieval System Online, or MEDLARS Online search strategy.
[DOCX File , 13 KB - jmir_v22i7e14230_app1.docx]

Multimedia Appendix 2
Study quality appraisal.
[DOCX File , 32 KB - jmir_v22i7e14230_app2.docx]

References

1. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, Wirtz AL, et al. The increase in global HIV epidemics in MSM. AIDS 2013 Nov 13;27(17):2665-2678. [doi: 10.1097/01.aids.0000432449.30239.fe] [Medline: 23842129]
2. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
3. Baral S, Sifakis F, Cleghorn F, Beyrer C. Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000-2006: a systematic review. PLoS Med 2007 Dec;4(12):e339 [FREE Full text] [doi: 10.1371/journal.pmed.0040339] [Medline: 18052602]
4. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis 2013 Mar;13(3):214-222. [doi: 10.1016/S1473-3099(12)70315-8] [Medline: 23260128]
5. Global AIDS Update. UNAIDS. 2016. URL: https://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf [accessed 2020-06-02]
6. Stahlman S, Lyons C, Sullivan PS, Mayer KH, Hosein S, Beyrer C, et al. HIV incidence among gay men and other men who have sex with men in 2020: where is the epidemic heading? Sex Health 2017 Feb;14(1):5-17. [doi: 10.1071/SH16070] [Medline: 27491699]
7. Ross AG, Ditangco RA, Belimac JG, Olveda RM, Mercado ES, Rogers GD, et al. HIV epidemic in men who have sex with men in Philippines. Lancet Infect Dis 2013 Jun;13(6):472-473. [doi: 10.1016/S1473-3099(13)70129-4] [Medline: 23718916]
8. Sullivan PS, Hamouda O, Delpech V, Geduld JE, Prejean J, Semaille C, Annecy MSM Epidemiology Study Group. Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996-2005. Ann Epidemiol 2009 Jun;19(6):423-431. [doi: 10.1016/j.annepidem.2009.03.004] [Medline: 19460672]
9. Hu H, Liu X, Zhang Z, Xu X, Shi L, Fu G, et al. Increasing HIV incidence among men who have sex with men in Jiangsu province, China: results from five consecutive surveys, 2011-2015. Int J Environ Res Public Health 2016 Aug 6;13(8):795 [FREE Full text] [doi: 10.3390/ijerph13080795] [Medline: 27509513]
10. Poon CM, Wong NS, Kwan TH, Wong HT, Chan KC, Lee SS. Changes of sexual risk behaviors and sexual connections among HIV-positive men who have sex with men along their HIV care continuum. PLoS One 2018;13(12):e0209008 [FREE Full text] [doi: 10.1371/journal.pone.0209008] [Medline: 30540851]
11. Khosropour CM, Dombrowski JC, Kerani RP, Katz DA, Barbee LA, Golden MR. Changes in condomless sex and serosorting among men who have sex with men after HIV diagnosis. J Acquir Immune Defic Syndr 2016 Dec 1;73(4):475-481. [doi: 10.1097/QAI.0000000000001128] [Medline: 27792685]
12. World Health Organization. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach. Geneva, Switzerland: World Health Organization; 2016.
13. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
14. INSIGHT START Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med 2015 Aug 27;373(9):795-807 [FREE Full text] [doi: 10.1056/NEJMoa1506816] [Medline: 26192873]
15. The Lancet HIV. U=U taking off in 2017. Lancet HIV 2017 Nov;4(11):e475. [doi: 10.1016/S2352-3018(17)30183-2] [Medline: 29096785]

XSL•FO
RenderX

16. 90-90-90: an Ambitious Treatment Target to Help End the Aids Epidemic. UNAIDS. 2014. URL: https://www.unaids.org/en/resources/909090 [accessed 2020-06-02]
17. Arreola S, Santos G, Beck J, Sundararaj M, Wilson PA, Hebert P, et al. Sexual stigma, criminalization, investment, and access to HIV services among men who have sex with men worldwide. AIDS Behav 2015 Feb;19(2):227-234. [doi: 10.1007/s10461-014-0869-x] [Medline: 25086670]
18. Schwartz SR, Nowak RG, Orazulike I, Keshinro B, Ake J, Kennedy S, TRUST Study Group. The immediate eff ect of the same-sex marriage prohibition act on stigma, discrimination, and engagement on HIV prevention and treatment services in men who have sex with men in Nigeria: analysis of prospective data from the TRUST cohort. Lancet HIV 2015 Jul;2(7):e299-e306 [FREE Full text] [doi: 10.1016/S2352-3018(15)00078-8] [Medline: 26125047]
19. Prevention Gap Report. UNAIDS. 2016. URL: https://www.unaids.org/en/resources/documents/2016/prevention-gap [accessed 2020-06-02]
20. Nelson KM, Pantalone DW, Gamarel KE, Carey MP, Simoni JM. Correlates of never testing for HIV among sexually active internet-recruited gay, bisexual, and other men who have sex with men in the United States. AIDS Patient Care STDS 2018 Jan;32(1):9-15 [FREE Full text] [doi: 10.1089/apc.2017.0244] [Medline: 29232170]
21. Marcus U, Gassowski M, Drewes J. HIV risk perception and testing behaviours among men having sex with men (MSM) reporting potential transmission risks in the previous 12 months from a large online sample of MSM living in Germany. BMC Public Health 2016 Oct 22;16(1):1111 [FREE Full text] [doi: 10.1186/s12889-016-3759-5] [Medline: 27770775]
22. Bolsewicz K, Vallely A, Debattista J, Whittaker A, Fitzgerald L. Factors impacting HIV testing: a review-perspectives from Australia, Canada, and the UK. AIDS Care 2015;27(5):570-580. [doi: 10.1080/09540121.2014.986050] [Medline: 25483628]
23. Berg RC. Predictors of never testing for HIV among a national online sample of men who have sex with men in Norway. Scand J Public Health 2013 Jun;41(4):398-404. [doi: 10.1177/1403494813483216] [Medline: 23567644]
24. Conway DP, Holt M, Couldwell DL, Smith DE, Davies SC, McNulty A, Sydney Rapid HIV Test Study. Barriers to HIV testing and characteristics associated with never testing among gay and bisexual men attending sexual health clinics in Sydney. J Int AIDS Soc 2015;18:20221 [FREE Full text] [doi: 10.7448/IAS.18.1.20221] [Medline: 26318960]
25. Wirtz AL, Walker DG, Bollinger L, Sifakis F, Baral S, Johns B, et al. Modelling the impact of HIV prevention and treatment for men who have sex with men on HIV epidemic trajectories in low- and middle-income countries. Int J STD AIDS 2013 Jan;24(1):18-30. [doi: 10.1177/0956462412472291] [Medline: 23512511]
26. Race K. Speculative pragmatism and intimate arrangements: online hook-up devices in gay life. Cult Health Sex 2015;17(4):496-511. [doi: 10.1080/13691058.2014.930181] [Medline: 24992161]
27. Sun CJ, Reboussin B, Mann L, Garcia M, Rhodes SD. The HIV risk profiles of Latino sexual minorities and transgender persons who use websites or apps designed for social and sexual networking. Health Educ Behav 2016 Feb;43(1):86-93 [FREE Full text] [doi: 10.1177/1090198115596735] [Medline: 26272786]
28. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
29. Garofalo R, Herrick A, Mustanski BS, Donenberg GR. Tip of the iceberg: young men who have sex with men, the internet, and HIV risk. Am J Public Health 2007 Jun;97(6):1113-1117. [doi: 10.2105/AJPH.2005.075630] [Medline: 17463378]
30. Phillips G, Magnus M, Kuo I, Rawls A, Peterson J, Jia Y, et al. Use of geosocial networking (GSN) mobile phone applications to find men for sex by men who have sex with men (MSM) in Washington, DC. AIDS Behav 2014 Sep;18(9):1630-1637. [doi: 10.1007/s10461-014-0760-9] [Medline: 24682866]
31. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
32. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
33. Cao B, Liu C, Stein G, Tang W, Best J, Zhang Y, et al. Faster and riskier? Online context of sex seeking among men who have sex with men in China. Sex Transm Dis 2017 Apr;44(4):239-244 [FREE Full text] [doi: 10.1097/OLQ.0000000000000575] [Medline: 28282651]
34. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, et al. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional sample of men who have sex with men in Los Angeles County. Sex Transm Infect 2014 Nov;90(7):567-572 [FREE Full text] [doi: 10.1136/sextrans-2013-051494] [Medline: 24926041]
35. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
36. Implementing Comprehensive HIV and STI Programmes with Men Who Have Sex with Men. UNFPA - United Nations Population Fund. 2015. URL: https://www.unfpa.org/publications/implementing-comprehensive-hiv-and-sti-programmes-men-who-have-sex-men [accessed 2020-06-02]

XSL•FO
RenderX

37. Lelutiu-Weinberger C, Pachankis JE, Gamarel KE, Surace A, Golub SA, Parsons JT. Feasibility, acceptability, and preliminary efficacy of a live-chat social media intervention to reduce HIV risk among young men who have sex with men. AIDS Behav 2015 Jul;19(7):1214-1227 [FREE Full text] [doi: 10.1007/s10461-014-0911-z] [Medline: 25256808]
38. Pachankis JE, Lelutiu-Weinberger C, Golub SA, Parsons JT. Developing an online health intervention for young gay and bisexual men. AIDS Behav 2013 Nov;17(9):2986-2998 [FREE Full text] [doi: 10.1007/s10461-013-0499-8] [Medline: 23673791]
39. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
40. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
41. Bowen AM, Williams ML, Daniel CM, Clayton S. Internet based HIV prevention research targeting rural MSM: feasibility, acceptability, and preliminary efficacy. J Behav Med 2008 Dec;31(6):463-477 [FREE Full text] [doi: 10.1007/s10865-008-9171-6] [Medline: 18770021]
42. Young SD, Jaganath D. Online social networking for HIV education and prevention: a mixed-methods analysis. Sex Transm Dis 2013 Feb;40(2):162-167 [FREE Full text] [doi: 10.1097/OLQ.0b013e318278bd12] [Medline: 23324979]
43. Rawat S, Wilkerson JM, Lawler SM, Patankar P, Rosser BS, Shukla K, et al. Recommendations for the development of a mobile HIV prevention intervention for men who have sex with men and hijras in Mumbai: qualitative study. JMIR Public Health Surveill 2018 May 3;4(2):e46 [FREE Full text] [doi: 10.2196/publichealth.9088] [Medline: 29724705]
44. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
45. Quality Assessment Tool for Quantitative Studies. Effective Public Healthcare Panacea Project. 1998. URL: https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/ [accessed 2017-03-02]
46. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
47. Blas MM, Alva IE, Carcamo CP, Cabello R, Goodreau SM, Kimball AM, et al. Effect of an online video-based intervention to increase HIV testing in men who have sex with men in Peru. PLoS One 2010 May 3;5(5):e10448 [FREE Full text] [doi: 10.1371/journal.pone.0010448] [Medline: 20454667]
48. Blas M, Menacho L, Alva I. Randomized Controlled Trial to Evaluate the Effect of a Novel Web-based Intervention to Increase HIV Testing in Men Who Have Sex With Men in Lima-Peru. ICH GCP: Good Clinical Practice. 2014. URL: https://ichgcp.net/clinical-trials-registry/NCT01760057 [accessed 2020-06-02]
49. Hirshfield S, Chiasson MA, Joseph H, Scheinmann R, Johnson WD, Remien RH, et al. An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. PLoS One 2012;7(10):e46252 [FREE Full text] [doi: 10.1371/journal.pone.0046252] [Medline: 23071551]
50. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of internet popular opinion leaders (iPOL) among internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
51. Patel V, Rawat S, Dange A, Lelutiu-Weinberger C, Golub SA. An internet-based, peer-delivered messaging intervention for hiv testing and condom use among men who have sex with men in India (CHALO!): pilot randomized comparative trial. JMIR Public Health Surveill 2020 Apr 16;6(2):e16494 [FREE Full text] [doi: 10.2196/16494] [Medline: 32297875]
52. Rhodes SD, Vissman AT, Stowers J, Miller C, McCoy TP, Hergenrather KC, et al. A CBPR partnership increases HIV testing among men who have sex with men (MSM): outcome findings from a pilot test of the CyBER/testing internet intervention. Health Educ Behav 2011 Jun;38(3):311-320 [FREE Full text] [doi: 10.1177/1090198110379572] [Medline: 21393625]
53. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using social media to increase HIV testing among gay and bisexual men, other men who have sex with men, and transgender persons: outcomes from a randomized community trial. Clin Infect Dis 2016 Jun 1;62(11):1450-1453 [FREE Full text] [doi: 10.1093/cid/ciw127] [Medline: 26980878]
54. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV test promotion videos: a noninferiority randomized controlled trial in China. Clin Infect Dis 2016 Jun 1;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]
55. Washington TA, Applewhite S, Glenn W. Using Facebook as a platform to direct young black men who have sex with men to a video-based HIV testing intervention: a feasibility study. Urban Soc Work 2017;1(1):36-52 [FREE Full text] [doi: 10.1891/2474-8684.1.1.36] [Medline: 29276800]

XSL•FO
RenderX

56. Wang Z, Lau JT, Ip M, Ho SP, Mo PK, Latkin C, et al. A randomized controlled trial evaluating efficacy of promoting a home-based HIV self-testing with online counseling on increasing HIV testing among men who have sex with men. AIDS Behav 2018 Jan;22(1):190-201. [doi: 10.1007/s10461-017-1887-2] [Medline: 28831616]
57. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
58. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
59. World Bank Country and Lending Groups. The World Bank. 2018. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-06-02]
60. World Health Organization. Consolidated Guidelines on HIV Prevention, Diagnosis, Treatment and Care for Key Populations. Geneva, Switzerland: World Health Organization; 2014.
61. Khosropour CM, Lake JG, Sullivan PS. Are MSM willing to SMS for HIV prevention? J Health Commun 2014;19(1):57-66. [doi: 10.1080/10810730.2013.798373] [Medline: 23905653]
62. Czarny HN, Broaddus MR. Acceptability of HIV prevention information delivered through established geosocial networking mobile applications to men who have sex with men. AIDS Behav 2017 Nov;21(11):3122-3128 [FREE Full text] [doi: 10.1007/s10461-017-1743-4] [Medline: 28260137]
63. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
64. Purnomo J, Coote K, Mao L, Fan L, Gold J, Ahmad R, et al. Using ehealth to engage and retain priority populations in the HIV treatment and care cascade in the Asia-Pacific region: a systematic review of literature. BMC Infect Dis 2018 Feb 17;18(1):82 [FREE Full text] [doi: 10.1186/s12879-018-2972-5] [Medline: 29454322]
65. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
66. Ayala G, Makofane K, Santos G, Beck J, Do TD, Hebert P, et al. Access to basic HIV-related services and PReP acceptability among men who have sex with men worldwide: barriers, facilitators, and implications for combination prevention. J Sex Transm Dis 2013;2013:953123 [FREE Full text] [doi: 10.1155/2013/953123] [Medline: 26316968]
67. Qiao S, Zhou G, Li X. Disclosure of same-sex behaviors to health-care providers and uptake of HIV testing for men who have sex with men: a systematic review. Am J Mens Health 2018 Sep;12(5):1197-1214 [FREE Full text] [doi: 10.1177/1557988318784149] [Medline: 29947563]
68. Johnson C, Baggaley R, Forsythe S, van Rooyen H, Ford N, Mavedzenge SN, et al. Realizing the potential for HIV self-testing. AIDS Behav 2014 Jul;18(Suppl 4):S391-S395. [doi: 10.1007/s10461-014-0832-x] [Medline: 24986599]
69. Figueroa C, Johnson C, Verster A, Baggaley R. Attitudes and acceptability on HIV self-testing among key populations: a literature review. AIDS Behav 2015 Nov;19(11):1949-1965 [FREE Full text] [doi: 10.1007/s10461-015-1097-8] [Medline: 26054390]
70. Pal K, Ngin C, Tuot S, Chhoun P, Ly C, Chhim S, et al. Acceptability study on HIV self-testing among transgender women, men who have sex with men, and female entertainment workers in Cambodia: a qualitative analysis. PLoS One 2016;11(11):e0166129 [FREE Full text] [doi: 10.1371/journal.pone.0166129] [Medline: 27829064]
71. Wirtz AL, Clouse E, Veronese V, Thu KH, Naing S, Baral SD, et al. New HIV testing technologies in the context of a concentrated epidemic and evolving HIV prevention: qualitative research on HIV self-testing among men who have sex with men and transgender women in Yangon, Myanmar. J Int AIDS Soc 2017 Apr 25;20(1):21796 [FREE Full text] [doi: 10.7448/IAS.20.01.21796] [Medline: 28453242]
72. Ryan KE, Wilkinson AL, Leitinger D, El-Hayek C, Ryan C, Pedrana A, et al. Characteristics of gay, bisexual and other men who have sex with men testing and retesting at Australia's first shop-front rapid point-of-care HIV testing service. Sex Health 2016 Nov;13(6):560-567. [doi: 10.1071/SH16027] [Medline: 27509318]
73. Wilkinson AL, Pedrana AE, El-Hayek C, Vella AM, Asselin J, Batrouney C, et al. The impact of a social marketing campaign on HIV and sexually transmissible infection testing among men who have sex with men in Australia. Sex Transm Dis 2016 Jan;43(1):49-56. [doi: 10.1097/OLQ.0000000000000380] [Medline: 26650997]
74. Katz DA, Dombrowski JC, Swanson F, Buskin SE, Golden MR, Stekler JD. HIV intertest interval among MSM in King County, Washington. Sex Transm Infect 2013 Feb;89(1):32-37 [FREE Full text] [doi: 10.1136/sextrans-2011-050470] [Medline: 22563016]
75. Strecher V, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 4;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]

XSL•FO
RenderX

76. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
77. Bailey J, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010 Sep 8(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
78. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
79. Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, et al. A systematic examination of the use of online social networking sites for sexual health promotion. BMC Public Health 2011 Jul 21;11:583 [FREE Full text] [doi: 10.1186/1471-2458-11-583] [Medline: 21777470]
80. Jones J, Salazar LF. A review of HIV prevention studies that use social networking sites: implications for recruitment, health promotion campaigns, and efficacy trials. AIDS Behav 2016 Nov;20(11):2772-2781. [doi: 10.1007/s10461-016-1342-9] [Medline: 26902295]
81. Mowlabocus S, Haslop C, Dasgupta RK. From scene to screen: the challenges and opportunities of commercial digital platforms for HIV community outreach. Social Media Soc 2016 Oct 20;2(4):205630511667288. [doi: 10.1177/2056305116672886]
82. Norman D. The Design of Everyday Things. New York, USA: Basic Books; 1988.
83. Thomas K, Linderoth C, Bendtsen M, Bendtsen P, Müssener U. Text message-based intervention targeting alcohol consumption among university students: findings from a formative development study. JMIR Mhealth Uhealth 2016 Oct 20;4(4):e119 [FREE Full text] [doi: 10.2196/mhealth.5863] [Medline: 27765732]
84. Duff O, Walsh D, Malone S, McDermott L, Furlong B, O'Connor N, et al. MedFit app, a behavior-changing, theoretically informed mobile app for patient self-management of cardiovascular disease: user-centered development. JMIR Form Res 2018 Apr 27;2(1):e8 [FREE Full text] [doi: 10.2196/formative.9550] [Medline: 30684426]
85. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A rapid review of ehealth interventions addressing the continuum of HIV care (2007-2017). AIDS Behav 2018 Jan;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
86. Simoni JM, Ronen K, Aunon FM. Health behavior theory to enhance ehealth intervention research in HIV: rationale and review. Curr HIV/AIDS Rep 2018 Dec;15(6):423-430 [FREE Full text] [doi: 10.1007/s11904-018-0418-8] [Medline: 30511186]
87. Allison SM, Adams D, Klindera KC, Poteat T, Wolf RC. Innovative uses of communication technology for HIV programming for men who have sex with men and transgender persons. J Int AIDS Soc 2014;17:19041 [FREE Full text] [doi: 10.7448/IAS.17.1.19041] [Medline: 25280864]
88. Magee JC, Bigelow L, Dehaan S, Mustanski BS. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. Health Educ Behav 2012 Jun;39(3):276-289. [doi: 10.1177/1090198111401384] [Medline: 21490310]
89. Expanding Access to HIV Services for Men Who Have Sex With Men Through Mobile Phone and Web-based Platforms in Myanmar. MEDBOX. 2017. URL: https://www.medbox.org/document/expanding-access-to-hiv-services-for-men-who-have-sex-with-men-through-mobile-phone-and-web-based-platforms-in-myanmar#GO [accessed 2020-06-02]

Abbreviations

MeSH: Medical Subject Headings
MSM: men who have sex with men
RCTs: randomized controlled trials
RR: risk ratio
STI: sexually transmitted infection
TW: transgender women

XSL•FO
RenderX

Edited by G Eysenbach; submitted 02.04.19; peer-reviewed by T Torres, J Xu, Z Niu, J Liu, A Daros, R Mpofu; comments to author 03.10.19; revised version received 20.02.20; accepted 10.04.20; published 28.07.20.

Please cite as:
Veronese V, Ryan KE, Hughes C, Lim MSC, Pedrana A, Stoové M
Using Digital Communication Technology to Increase HIV Testing Among Men Who Have Sex With Men and Transgender Women: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(7):e14230
URL: https://www.jmir.org/2020/7/e14230
doi:10.2196/14230
PMID:32720902

©Vanessa Veronese, Kathleen Elizabeth Ryan, Chad Hughes, Megan SC Lim, Alisa Pedrana, Mark Stoové. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Information Needs of Breast Cancer Patients: Theory-Generating Meta-Synthesis

Hongru Lu[1], MSc; Juan Xie[1], BSc; Lynette Hammond Gerido[2], MPH; Ying Cheng[1], PhD; Ya Chen[1], PhD; Lizhu Sun[3], MSc

[1]School of Information Management, Nanjing University, Nanjing, China

[2]School of Information, Florida State University, Tallahassee, FL, United States

[3]Department of Oncology, The Affiliated Shuyang Hospital of Xuzhou Medical University, Shuyang, China

Corresponding Author:
Ying Cheng, PhD
School of Information Management
Nanjing University
163 Xianlin Avenue
Nanjing
China
Phone: 86 138 5183 8364
Email: chengy@nju.edu.cn

Abstract

Background: Breast cancer has become one of the most frequently diagnosed carcinomas and the leading cause of cancer deaths. The substantial growth in the number of breast cancer patients has put great pressure on health services. Meanwhile, the information patients need has increased and become more complicated. Therefore, a comprehensive and in-depth understanding of their information needs is urgently needed to improve the quality of health care. However, previous studies related to the information needs of breast cancer patients have focused on different perspectives and have only contributed to individual results. A systematic review and synthesis of breast cancer patients' information needs is critical.

Objective: This paper aims to systematically identify, evaluate, and synthesize existing primary qualitative research on the information needs of breast cancer patients.

Methods: Web of Science, EBSCO, Scopus, ProQuest, PubMed, PsycINFO, The Cochrane Library, the Cumulative Index to Nursing and Allied Health Literature were searched on February 12 and July 9, 2019, to collect relevant studies. A Google Scholar search, interpersonal network recommendations, and reference chaining were also conducted. Eligible studies included qualitative or mixed-methods studies focusing on the information needs (across the cancer continuum) of breast cancer patients or their social networks. Subsequently, a Critical Appraisals Skills Programme checklist was used to assess the quality of included research. The results, findings, and discussions were extracted. Data analysis was guided by the theory-generating meta-synthesis and grounded theory approach.

Results: Three themes, 19 categories, and 55 concepts emerged: (1) incentives (physical abnormality, inquiry from others, subjective norm, and problems during appointments); (2) types of information needs (prevention, etiology, diagnosis, clinical manifestation, treatment, prognosis, impact and resumption of normal life, scientific research, and social assistance); (3) moderating variables (attitudes, health literacy, demographic characteristics, disease status, as well as political and cultural environment). The studies revealed that the information needs of breast cancer patients were triggered by different incentives. Subsequently, the patients sought a variety of information among different stages of the cancer journey. Five types of variables were also found to moderate the formation of information needs.

Conclusions: This study contributes to a thorough model of information needs among breast cancer patients and provides practical suggestions for health and information professionals.

(J Med Internet Res 2020;22(7):e17907) doi:10.2196/17907

KEYWORDS

breast cancer patients; information needs; incentives; moderating variables; meta-synthesis

XSL•FO
RenderX

Introduction

Breast cancer is one of the most commonly diagnosed malignant tumors worldwide [1]. In 2018, around 2.1 million new female breast cancer cases were diagnosed worldwide, accounting for almost 25% of the cancer cases among women [1]. In the United States, an estimated 42,260 breast cancer deaths were predicted in 2019 [2]. The dramatic growth in the number of breast cancer patients has led to an increasing need for information to manage symptoms [3], make decisions, control lives, and prepare for the future [4]. However, information needs have not been thoroughly examined. For example, a lack of targeted education materials has been reported in clinical practice [3]. Several large-scale patient investigations have identified inadequate and inaccessible patient information [5]. Therefore, it is crucial and challenging for health service providers to systematically investigate patients' information needs and offer relevant information.

We identify two existing reviews on breast cancer patients' information needs. A review of relevant literature published between 1988 and 1998 classified individual studies to examine information needs and source preferences of breast cancer patients and their family members at different points in their cancer journey [6]. They found that essential information needs changed over time, and family members also needed information to support the patients. A recent protocol for a scoping review has been put forward that aims to summarize studies exploring information needs, source preferences, and engagement behaviors of women with a specific breast cancer type (metastatic) [5]. However, it does not address the specific information needs of general breast cancer patients, and no further findings have been published. Therefore, there is no updated review and synthesis of breast cancer patients' information needs.

There is a large body of qualitative studies that have mainly collected data through interviewing and nonparticipatory observation on this topic. These studies contribute to an insightful understanding of patients' information needs and characteristics [7], but they have mostly focused on particular topics with limited samples and produced mixed results. For example, different results have identified the information needs of patients in specific age groups [8,9]. Therefore, a qualitative meta-synthesis will help to address the mixed results and establish a general model of the information needs of breast cancer patients.

To narrow these gaps, this study aims to identify, evaluate, and synthesize existing primary qualitative research on the information needs of breast cancer patients and generate an integrated model to articulate their information needs (ie, incentives, types and moderating variables) across the cancer care continuum. The findings can inform health professionals and information service providers to help breast cancer patients receive appropriate information and be well equipped to cope with the disease.

Methods

This study follows the processes of theory-generating meta-synthesis [10] and grounded theory [11].

Search Strategy

The search strategy was first developed for Web of Science (Clarivate Analytics) and then adjusted to search EBSCO, Scopus, ProQuest, PubMed, PsycINFO, The Cochrane Library, and the Cumulative Index to Nursing and Allied Health Literature. These databases were chosen based on related studies [12,13]. In addition, a Google Scholar search, interpersonal network recommendations, and reference chaining of the included articles were applied. The literature search was conducted on February 12, 2019, and updated on July 9, 2019. Duplicate articles were removed.

Search terms were chosen from two categories: breast cancer patients and information needs. The search terms were chosen according to MeSH vocabulary [14] and related research on information behavior [15-20] (see Multimedia Appendix 1 for the detailed search strategies). "Behavio*," "seek*," "source*," and other words were added so all related studies could be identified since findings on information needs may have also been covered in information behavior research (eg, what information did the patients search?). Limits on research methods were not initially placed on the search terms since it was difficult to accurately identify all qualitative studies when searching for the literature on this topic in the databases [21]. Therefore, qualitative studies were selected in the screening process.

Selection Criteria

Articles were selected according to inclusion and exclusion criteria [12,13] (Textbox 1).

XSL•FO
RenderX

Textbox 1. Selection criteria for the study.

Inclusion criteria:

- Empirical studies using qualitative or mixed research methods
- Studies focusing on breast cancer patients or support groups encompassing breast cancer patients (can include other types of patients)
- Studies related to information needs

Exclusion criteria:

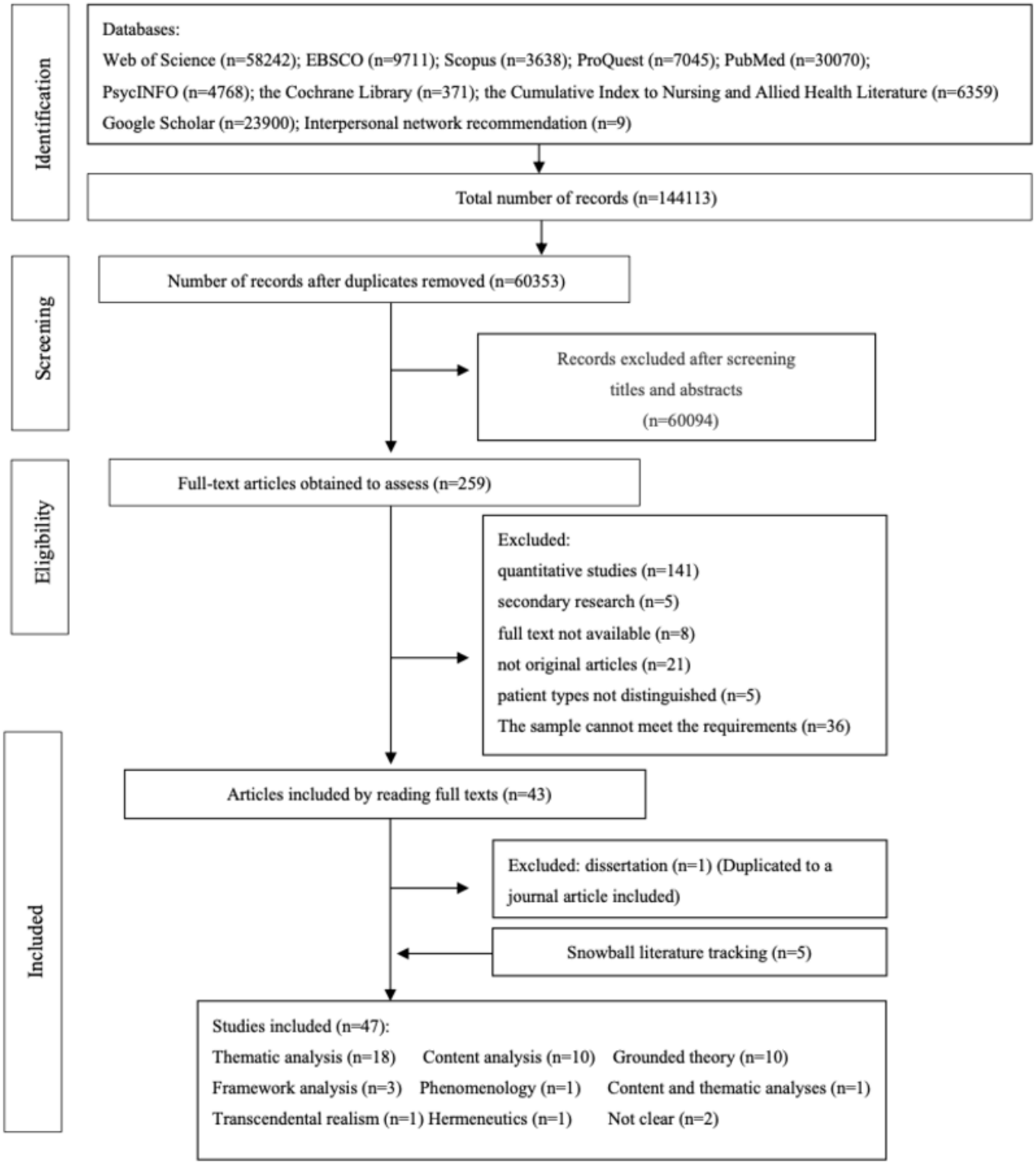
- Reviews or patents
- Mixed-method studies from which no qualitative findings can be extracted
- Studies focusing on participants with the risk of breast cancer (family history)
- Studies concerning breast cancer patients' relatives, friends, or spouses
- Studies concentrating on doctors of breast cancer patients
- Studies related to breast cancer patients and other kinds of participants that fail to distinguish breast cancer patients from other groups

Screening

The first author and three members of the research team (Juan Xie, Ying Cheng, and Ya Chen) browsed the titles and abstracts to identify possible studies. The full texts of these studies were then retrieved for further screening based on the inclusion and exclusion criteria. The screening was conducted separately by the four authors, and disagreements were addressed through discussion. An additional 5 articles were found through a snowballing literature track. Finally, 47 full-text articles were included after removing one duplicate. The literature search process is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart of searched, excluded, and included items.



Quality Appraisal

The quality of the included studies was appraised using the Critical Appraisals Skills Programme [22] (see Multimedia Appendix 2 for scores). As shown, all 47 studies received a score indicating high quality (ie, over 8). Therefore, no paper was excluded through this appraisal. A total of 38 articles received a score of 10, which indicated that they fully reported their research objectives, qualitative research methods, research design, recruitment strategy, data collection methods, relationship between researchers and participants, ethical issues, data analysis, statement of findings, and value of the research [22]. Nine articles scored 9 indicating that they did not clearly describe the research objectives, method of data analysis, or ethics statement.

Data Extraction and Synthesis

All six researchers independently read and extracted the results, findings, and discussions of the included qualitative studies. To extract as many relevant findings as possible, researchers discussed, reached consensus, and adopted the same codes to analyze the articles.

The extracted data were iteratively synthesized in three stages. First, four authors (HL, JX, Y Cheng, and LS) identified the concepts that best fit the extracted raw texts. Categories were

XSL•FO
RenderX

then formed by gathering concepts with the same properties and dimensions through continuous comparison. Second, the relationships among categories were also examined. Themes emerged after an in-depth analysis and classification of categories. Third, concerning the evidence found in the extracted data, all themes were integrated according to their logical relationship. Finally, a model clarifying the information needs of breast cancer patients was built.

Results

Overview of the Literature

In all, 47 journal articles were included. The characteristics of the included studies are shown in Multimedia Appendix 2. The data collection methods used in these studies were semistructured interviews, focus groups, observations, an online survey, and/or an open-ended questionnaire. Most studies applied thematic analysis, content analysis, and grounded theory to analyze the data. The recruitment strategies were generally convenience sampling and purposive sampling. These studies were conducted in the United States (13), Australia (12), United Kingdom (7), Canada (10), Turkey (1), Japan (1), Iran (1), Poland (1), and Switzerland (1). As for the populations, 26 studies reported the race of the participants, among which 18 studies focused mainly on white participants (over 60%), and 6 studies recruited all Asian participants (eg, Japanese and Chinese-Australian). Participants of one study were all African American. In addition, one study recruited mixed participants such as Asian, white, and Filipino. In terms of the types of information patients sought, the studies revealed that participants wanted information from various sources, encompassing information from health care providers (47); peers/support teams (24); the internet (18); books (14); cancer organizations (13); families (13); friends/coworkers (13); brochures and pamphlets (12); magazines (9); audio/videos (7); complementary, alternative, or unconventional practitioners (6); newspapers (4); libraries (3); television (3); telephone hotlines (3); medical records (2); and broadcasts (1). In addition, 10 studies used theoretical frameworks, such as planned behavior theory and social learning theory, in their research.

Synthesized Findings

Summary

Three themes emerged to articulate the developing process of information needs of breast cancer patients, with 19 categories and 55 concepts underpinning the themes (Table 1). This study also explored the logical relations between themes and built an integration model (Figure 2) guided by incentive theory [23]. According to the theory, individual activities are not only derived from internal motivation but are also stimulated by external factors, including positive and negative incentives [23]. The process of diagnosis and treatment in breast cancer patients' journeys is sequential [24] and often accompanied by various diagnostic approaches and treatment regimens in a clinical context [24]. Therefore, diagnosis and treatment were synthesized as one stage in this study (phase 2). Consequently, we identified another two phases. Phase 1 is the stage before diagnosis: before going to the hospital for medical advice or examination in allusion to physical abnormalities. Phase 3 is conceptualized as survivorship. Incentives and content of patients' information needs vary in each of the phases. In particular, 7 studies reported the content and incentives of information needs before diagnosis (phase 1), while 44 studies focused on diagnosis and treatment (phase 2) and 29 studies presented information needs during survivorship (phase 3). Variables moderating the formation of information needs are also shown in the model.

XSL•FO
RenderX

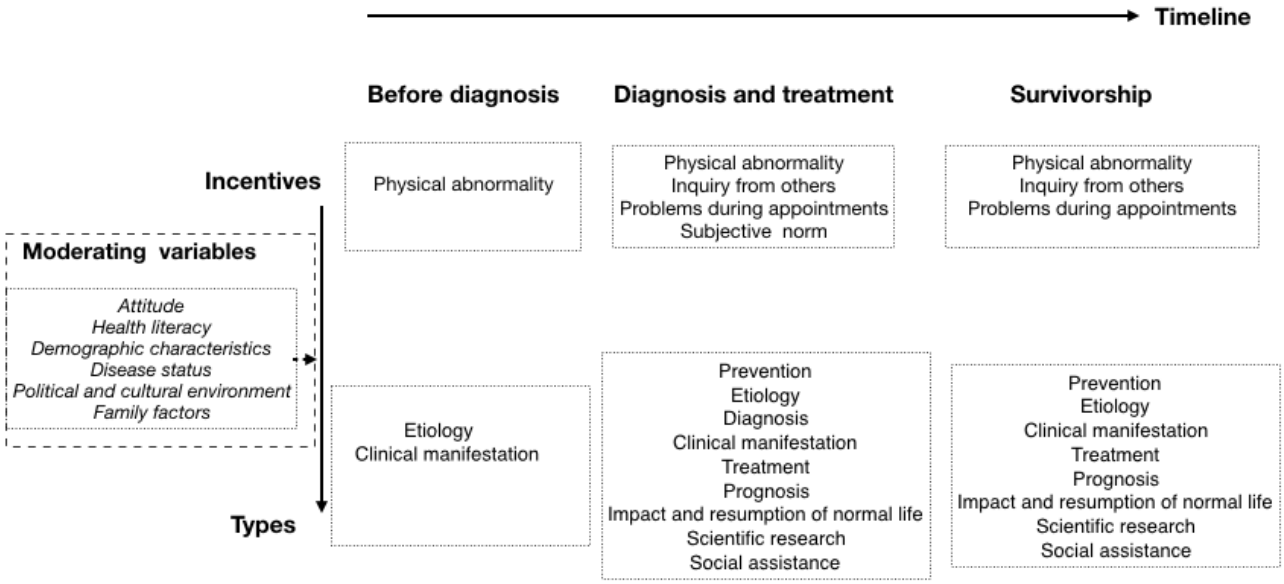
Table 1. Synthesized results of the studies.

Themes, categories, and concepts		Examples of quotations	References
Incentives			
Physical abnormality	Appearance of symptoms of the primary tumor; appearance of symptoms of side effects	"Information needs also became more prevalent when patients experienced side effects..." [25].	[9,25-38]
Inquiry from others	Inquiry from doctors; inquiry from patients around them; inquiry from general people around them	"I was being asked so many questions from those around me, and I wished if I have asked the physicians these questions and knew its answers" [36].	[26,31,32,36,39]
Subjective norm	Advice from support group member	"Leigh followed the advice of another medically savvy support group member and requested copies of 'everything' regarding her diagnosis" [40].	[40]
Problems during appointments	Time span; perfunctory doctors; changes in health care staff; inconsistent information	"When this time frame was more than 4 weeks, patients found it hard to remember everything that had been discussed. They therefore had additional information needs that required attention during their planning appointment" [25].	[25,30,37,41]
Types of information needs			
Prevention	Effectiveness of breast self-examination; prevention for family members	"Themes related to important content issues include:...prevention for daughters" [30].	[30,40,42,43]
Etiology	Internal factors; external factors	"The exchange of misinformation also led many of the women interviewed...to hold misconceptions about breast cancer including misconceptions about risk factors...of the disease" [36].	[36,40,42,44]
Diagnosis	Specific biopsy procedures; pathologic results; precision and applicability of examination tools; explanations of technical terms related to diagnosis; clinical stage; waiting time of diagnostic tests and its impact on prognosis	"Faced with having to go for a mammogram, the women were concerned that this imaging tool lacked the precision needed to detect a tumor through dense breast tissue...the mammograms don't work with young women" [9].	[9,36,40,43,45-47]
Clinical manifestation	Symptoms of the primary tumor; symptoms of side effects; meaning of corresponding symptoms	"Participants reported a preference for a list of signs and symptoms of breast cancer recurrence as a means to reduce unnecessary anxiety" [48].	[30,33,36,37,43-45,48-50]
Treatment	Treatment options; side effects of treatment; management of side effects; treatment preparation; treatment procedures; treatment evaluation summaries; prevention of recurrence	"A small number of women also reported asking the health care provider which treatment option they would recommend" [28].	[8,9,25,26,28-66]
Prognosis	Survival rates; risk of complications; survival statistics of treatment regimens; risk of recurrence	"The search for survival statistics proved fruitless, though, since such information cannot simply be applied to an individual case to determine prognosis" [26].	[26,27,30-32,42,46,47,51,54,55,59,66]
Impact and resumption of normal life	Ways to communicate the diagnosis results with family members; impact of treatment on quality of life; strategies to improve quality of life	"Patients had information needs relating to...whether they can return to work and other health services" [25].	[8,9,25-30,32-36,38,39,41-44,47,48,51-53,56,60-64,67-69]
Scientific research	Recent research findings; clinical trials	"...every week there's another breakthrough. And you go to your doctor with the clipping" [30].	[30,40,44,47,65]
Social assistance	Insurance; financial support	"Women's careers were often affected by the illness yet information about financial support was extremely difficult to come by" [52].	[9,26,30,43,46,52,60]
Moderating variables			
Attitudes	Affect; behavior; cognition	"Some women talked of being shocked, frightened, and worried when discovering a breast lump and immediately sought medical advice. Other women appeared less concerned and mentioned breast symptoms only when attending their general practitioner for other reasons" [28].	[8,9,25,26,28,33,35,36,46,48,57]

XSL•FO
RenderX

Themes, categories, and concepts		Examples of quotations	References
Health literacy	Health beliefs; health knowledge; health styles	"The possibility of symptoms being breast cancer was the first thought of some women and most were fatalistic about this. 'I just thought I have cancer and I wasn't bothered about it because let sleeping dogs lie. The less you know the less you bother about it'" [28].	[9,26,28,33,34,40]
Demographic characteristics	Age; education level; economic status	"Ongoing information needs...younger women also discussed the use of complementary therapies...more than their older counterparts" [44].	[8,25,28,40,42,44,57,64]
Disease status	Comorbidity; clinical stage	"Experiences of some participants with special disease showed they needed information regarding secondary prevention" [42].	[41,42,60]
Political and cultural environment	Cultural background; health care policy	"Given that food therapy plays a significant role in Chinese culture, many participants expressed a strong desire for information on diet" [49].	[30,32,33,35,38,49]
Family factors	Age of children	"At diagnosis, women wanted age-appropriate information about how to communicate with their children about cancer" [9].	[9]

Figure 2. Model of information needs of breast cancer patients.



Theme 1. Incentives

This theme illustrates how breast cancer patients formed a cognitive state of information gaps, which led to their information needs. These needs were triggered by various factors.

Physical Abnormalities

Breast cancer patients might experience physical abnormalities anytime during the cancer continuum. Their information needs were often triggered by the appearance of symptoms of the primary tumor (eg, lumps, pain, or discharge) [9,26,28,31-33,36] because patients' panic and uncertainty increased before diagnosis. Nine studies also reported side effects caused by the treatment (eg, irregular menstruation, arm complications, fatigue, sexual dysfunction, lymphedema, and pain) [25,27,29,30,32,34,35,37,38] that would last from the treatment to the survivorship stages and could lead to specific information needs.

Inquiry From Others

Three studies suggested that doctors provided patients with several surgery choices and asked patients to quickly make treatment decisions (in phase 2), resulting in information needs about treatment [26,32,39]. During treatment (in phase 2), a patient wished that she had asked doctors more questions because she could not answer the queries of patients around her, indicating that other patients' questions helped form information needs that some patients themselves were not aware of [36]. In the survivorship stage, a patient was asked questions by general people around her, which reminded her of the risk of a recurrence, leading her to go back to the clinic for a follow-up appointment [31].

Subjective Norm

Patients' information needs were also triggered by subjective norms ("the belief that an important person or group of people will approve and support a particular behavior" [70]). For example, in phase 2, one patient reportedly requested copies of

XSL•FO
RenderX

all her diagnostic results from the hospital [40], following the advice from a support group member.

Problems During Appointments

Similar to the inquiries from others, problems could occur during appointments. For example, during the process of treatment decision making, the long waiting period between diagnosis and making a final treatment decision made it difficult for patients to remember all the information discussed, thus generating more information needs [25]. In addition, the doctors sometimes did not provide patients with enough information about breast cancer treatment options and simply made the treatment decision without discussion with the patient, which left the patient with more uncertainty [41]. The patients also had many questions (eg, who and how often should they see the doctor) as they transitioned from the cancer care group to the primary care physician shortly after treatment (in phase 3) [30,37]. Patients were also confused when health care professionals provided them with inconsistent information (eg, tamoxifen discontinuation) in the survivorship stage [30], triggering the need for more exact information. In a follow-up appointment (in phase 3), it was reported that the doctor's indifferent attitude also led to patients' anxiety and uncertainty when doing a routine examination [30].

Theme 2. Information Needs

This theme refers to specific information patients needed. The included studies suggested that breast cancer patients generally had 9 types of information needs at different stages of cancer care including information on prevention, etiology, diagnosis, clinical manifestation, treatment, prognosis, impact on normal life and coping, research progress, and social assistance. The types of information needs can be divided into several concepts.

Prevention

Three studies reported that patients were concerned about information on prevention, including effectiveness of breast self-examination [40] in phase 2. Patients reported that they also needed information about prevention for family members [30,42,43] during the survivorship stage.

Etiology

Etiology information was needed throughout the whole cancer journey. Four studies reflected on this type of information need. For example, patients wanted to know about internal and external factors (risk factors) before diagnosis in phase 1 [36] and phase 2 [40]. Patients also needed information on internal (genetic) [42,44] and external factors (possible carcinogens and environmental triggers) [44] in the survivorship stage in phase 3.

Diagnosis

In the diagnosis and treatment phase (in phase 2), patients were concerned about specific biopsy procedures (arrangement, temporal sequence, and actual procedure) [45], pathologic results (results of the biopsy) [45], and precision and applicability of the examination tools [9]. They also wanted an explanation of technical terms related to the diagnosis (used to describe tumor types) [46], pathologic results (pathological stage or the type of breast cancer) [36,43,46], clinical stage [36,40], waiting time of diagnostic tests, and the corresponding impact on prognosis [47].

Clinical Manifestation

Patients wanted to know about the signs of breast cancer (symptoms of the primary tumor) [36] before diagnosis (phase 1). Information about the meaning of the corresponding symptoms was needed by patients who were undergoing a breast biopsy (in phase 2) [45]. They also needed to be informed of the symptoms of the side effects caused by the treatment [30,37] during phases 2 and 3. Eight studies also reported that patients were concerned about symptoms of a recurrence (symptoms of the primary tumor) [30,33,37,43,44,48-50] during phase 3.

Treatment

Patients needing treatment information during the cancer continuum was commonly reported across studies (n=44): 23 studies found that patients needed treatment options while waiting for diagnosis results [45], shortly after receiving diagnosis results [36,40,51], and during the treatment decision process [8,9,26,28,31,32,34,39,40,42,46-48,52-59] such as popularity, advantages and disadvantages, risks, applicability, recovery time, comparison, cost, and optimal time of different treatment options in phase 2. Twelve articles also mentioned that patients wanted to be educated about the side effects of treatment [8,9,25,28,31,36,40,46,56,60] when considering the treatment decision. The information needs of patients during treatment preparation included details about treatment preparation [33,42,48,52] and treatment procedures [9,25,28,30,42,46,61,62].

During treatment, patients' information needs focused on the side effects of treatment [25,47] and management of the side effects [8,25,38,40,46,49,50,53,60,63,64]. The studies also reported that patients needed details on the treatment options (alternative and complementary medicines and treatment) [25,65] and treatment procedures (eg, location of the pharmacy) during treatment [28,40,42,46,53,57,61]. In the survivorship stage, patients also paid attention to information on side effects [8,29,30,32,33,36-38,41,42,48] and management of side effects [25,30,35,37,43,48,49]. In phase 3, patients needed information about treatment options (eg, complementary and alternative therapies) [42,44,64,65] and treatment evaluation summaries [43,48]. In addition, patients wanted information about preventing a recurrence (eg, prevention-related policies, signs and symptoms of a recurrence, natural remedies to prevent recurrence, preventive health actions to minimize the risk of a recurrence) [30,33,37,41-43,49,52,61,66] in the survivorship stage.

Prognosis

Patients' need for prognosis information included general survival rates [26,47,51,54] and the risk of a recurrence [27] in phase 2. Studies also found that patients desired information about the risk of complications [42], risk of a recurrence [31,46], and survival statistics of treatment regimens [55,59] when they made treatment decisions. In particular, patients needed to know the risk of a recurrence in phases 2 [47] and 3 [30,32,66].

XSL•FO
RenderX

Impact and Resumption of Normal Life

Patients wanted to know how to communicate about the cancer with their family members in phase 2 [9,26,52,67,68]. Patients also needed information about the impact of treatment on their own body image [28,32,33,35,42,62,69], fertility (or menopause) [8,35], work [25,26], and psychology [8,26,64] in phase 2. In the survivorship stage, patients were concerned about the impact of treatment on work [44,52], intimate relationships [44,69], fertility (or menopause) [9,30,44], psychology [42,44,69], and body image [9,27,34,35,69]. In 26 studies, strategies to improve patients' quality of life (eg, fertility preservation options, sex advice, instructions about diet and exercise, stress management, reconstruction) were important for patients in both phase 2 [9,26,27,35,38,39,47,51,53,56,60-62] and phase 3 [29,30,32,36,41-44,48,63,64,69].

Scientific Research

Some patients needed information about recent breast cancer research findings [40] and clinical trials [47,65] in phase 2. In addition, medical breakthroughs (recent research findings), corresponding clinical trials [30], new treatment, and research developments [44] were requested in phase 3.

Social Assistance

Patients generally needed information about insurance (eg, health insurance) in phases 2 [46] and 3 [30]. Five studies reported that patients sought information on financial support (eg, employment benefits, where to get help) during the survivorship stage [9,26,43,52,60].

Theme 3. Moderating Variables

A prominent theme reported in the included studies was variables that moderated the development of information needs of breast cancer patients. Differences among particular subgroups and their forms of information needs are shown in Table 2. No major inconsistencies were detected in the studies.

XSL•FO
RenderX

Table 2. Synthesized results of the moderating variables.

Categories, concepts, and dimension	Effect on information needs	References
Attitudes		
Cognition		
Correct cognition	Facilitated development of information needs	[9,26,28]
Misunderstanding	Inhibited development of information needs	[33,36]
Affect		
Shocked and worried	Facilitated development of information needs	[28,57]
Not worried	Inhibited development of information needs	[28]
Behavior		
Ready	Facilitated development of information needs	[8,35,46,48]
Not ready	Inhibited development of information needs	[25]
Health literacy		
Health knowledge		
Rich in health knowledge	Facilitated development of information needs	[33,40]
Lacking health knowledge	Inhibited development of information needs	[33]
Health style		
Positive coping strategies	Facilitated development of information needs	[33]
Negative coping strategies	Inhibited development of information needs	[34]
Health beliefs		
No fatalism	Facilitated development of information needs	[9,26,34]
Fatalism	Inhibited development of information needs	[28,33]
Demographic characteristics		
Age		
Younger	Paid more attention to information on new treatments, research advances, and effects of treatment on fertility and career, complementary therapies, dietary changes and exercise, possible carcinogens, and environmental factors	[8,25,44]
Older	Expressed less need for reconstructive surgery	[28,40,42]
Education level		
Higher	Paid more attention to information on medical terminology and medical information systems	[40]
Lower	Inhibited the development of information needs	[40]
Economic status		
Higher	Paid more attention to information on natural health products and healthy dietary changes	[57,64]
Lower	Expressed less need for healthy diet information and reconstructive surgery	[42]
Disease status		
Comorbidity		
With comorbidity	Paid more attention to information on secondary prevention; obese patients needed survival guidelines targeting their physical condition	[41,42]
Clinical stage		
Advanced breast cancer	Paid more attention to the experience of other advanced breast patients; information or support related to last will and testament and final arrangements	[60]
Political and cultural environment		
Cultural background		

XSL•FO
RenderX

Categories, concepts, and dimension	Effect on information needs	References
Chinese	Paid more attention to diet and exercise guidelines, less information on postoperative body changes	[33,38,49]
Turkish	Paid more attention to postoperative body changes and contraceptive information	[35]
Japanese	Paid more attention to information on postoperative body changes	[32]
Health care policy		
Policy changes	Paid more attention to information on changes in health care policy and practice (eg, frequency of routine examination)	[30]
Family factors		
Age of children		
Younger	Paid more attention to age-appropriate information on how to guide communication with children about the disease	[9]

Attitudes

Attitudes refer to the amount of affection for or against some object [71]. In this study, attitudes included affect, behavior, and cognition. Affect represented the strong feelings of breast cancer patients, such as being shocked and worried [28,57], while behavior described their readiness to receive information [25]. Cognition was the product of knowledge acquisition or application, such as an understanding of breast cancer.

Six studies showed that patients' cognition and affect influenced the development of information needs. For example, some patients misunderstood the causes and risks of breast cancer, believing that they would not have breast cancer after a certain age [33] or after breastfeeding [36], and thus they did not seek information. Correct cognition of patients facilitated the development of their need for information. For example, patients went to see their physicians after discovering a lump or feeling pain [9,26,28]. Two studies showed that affect also influenced their information needs. For instance, some participants who found a lump and felt shocked immediately went to the hospital to seek medical advice [28,57], while those who were not worried about the lump only described the symptoms when they went to see their general practitioner for other diseases or they chose to completely ignore the lump [28].

In addition, some patients who believed that they were still in the process of receiving diagnosis results were not willing to process a large amount of information [25]. Patients who lacked behavior intention chose to avoid information [25], while other patients asked for various types of information [8,35,46,48].

Health Literacy

Health literacy is defined as individuals' ability to obtain, understand, evaluate, and apply health information, which can affect individual judgments and decisions on health care and disease prevention [72]. In the study, this category included health beliefs, health knowledge, and health styles, respectively, representing patients' values toward breast cancer, awareness of cancer-related information in order to seek appropriate health services [73], and their coping strategies for breast cancer.

For example, one patient did not go to see a doctor after discovering a lump in her breast because she lacked health knowledge about the effect of screening tools (mistakenly believing that routine mammographic screening had a preventive effect), which inhibited the development of her information needs [33]. Relatively abundant health knowledge could promote the articulation of the need for information [33,40]. While one women who had a different health style also ignored the discovery of lump until she found that the lump was getting bigger 3 month later, other patients went to see a doctor immediately after feeling breast pain [33]. In terms of health beliefs, some of the patients suspected breast cancer immediately after finding the lump, but they were not worried or did not ask about it because of fatalism [28,33] whereas other patients sought further examination [9,26,34].

Demographic Characteristics

Eight studies reported that the patients' age, education level, and financial situation moderated the formation process of their information needs. For example, young patients cared more about new treatments, research advances, and effects of treatment on fertility and career compared to older patients [8,25,44]. Younger patients also expressed a greater need for information about complementary therapies, dietary changes and exercise, possible carcinogens, and environmental factors of breast cancer because of their interest or a sense of control and comfort [44]. In contrast, some older patients had less need for information and considered that the less they knew, the less they were bothered about it [28]. In general, patients with a lower education level were more likely to trust doctors and be satisfied with the information provided by their doctors [40]. These patients believed that they did not need more information and had fewer information needs than other groups [40]. In terms of education levels, having a high education level had a beneficial effect on information needs and they independently researched areas such as medical terminology and medical information systems [40]. Furthermore, patients with lower economic status showed no concern about healthy diet information due to the high costs as well as the conflicts between a special diet and their family's diet [42]. In contrast, others cared about natural health products and healthy dietary changes [57,64]. It was also found that worse economic status and older age inhibited their need for information on reconstructive surgery [42].

XSL•FO
RenderX

Disease Status

Disease status involved comorbidity and the clinical stage of patients. For patients with comorbidity, information on secondary prevention was needed to avoid complications [42]. For example, obese patients faced a higher risk of recurrence, and they needed targeted survival guidelines based on their physical condition [41]. Patients with advanced breast cancer also wanted to know about the experiences of other patients in the same clinical stage [60]. They also sought information or support related to the last will and testament and final arrangements [60].

Political and Cultural Environment

The political and cultural environment refers to the social and cultural context characterized by a community's values and beliefs [74]. Of the included studies, 6 suggested a moderating effect of cultural background and health care policy. Patients who were influenced by Chinese culture developed more information needs regarding specific recommendations on diet and exercise related to symptom management, promotion of rehabilitation, and prevention of recurrence [38,49] because of the traditional food therapy culture. Some Chinese patients reported no feeling about postoperative body changes because body image was not as important to them as returning home to take care of their families in the context of Chinese culture [33]. In contrast, patients in some other countries like Turkey [35] and Japan [32] attached great importance to the aesthetic needs of the female body image. In addition, for patients from some countries with religious beliefs like Turkey with strong opposition to abortion, the cultural beliefs directly affected the patients' need for contraceptive information [35]. Patients in countries that had experienced changes in the national health care policy (eg, national cost curtailment policy) and corresponding adjustments in practice (eg, frequency of routine examinations) needed more targeted information on these changes [30].

Family Factors

In addition to the moderating variables described above, one study found that patients who had younger children often talked about the lack of personalized information for their kids [9]. They particularly needed targeted information on how to guide their communication about breast cancer with their children because they had difficulty finding age-appropriate information [9].

Discussion

Principal Findings

This meta-synthesis strengthens our understanding of the formation and types of information needs of breast cancer patients. It also highlights the variables moderating the development of information needs.

Information Service Targeted on General Breast Cancer Patients

Figure 2 shows that breast cancer patients' information needs change over the three stages. Patients are often given substantial resources (eg, brochures, books, and nursing plans) but the information has disadvantages in both content and form. Targeted messages are often embedded in a large amount of irrelevant information. Breast cancer patients are also often given overwhelming information during hospitalization, putting extra pressure on them [75]. In addition, the information is not always delivered in expected, easy-to-digest forms, but is often filled with medical jargon, for instance [3]. This study provides some practical recommendations based on synthesized findings to guide medical professionals attempting to provide better information for breast cancer patients.

In this synthesis, breast cancer patients' information needs were systematically examined and classified into three cancer phases. This can help us understand the changing process of the specific information needed by patients at different phases. The synthesized findings also inform recommendations for health care staff to offer relevant and timely education materials at the appropriate stage of the cancer journey. A total of 94% (44/47) of the studies reported that patients were concerned about treatment information such as treatment procedures, side effects, and preoperative preparation. This type of universal information should be incorporated into education materials to help patients make proper treatment decisions. The synthesis also revealed that it is difficult for patients to obtain information about financial assistance [9,26,30,52,60], as well as complementary and alternative therapies [34,64]. More attention should be given to meet these information needs.

Easy-to-understand materials are also recommended since much the existing information and oral communication seems to contain complex medical language [76]. Education programs using virtual reality technology could also increase the knowledge and positive experience of patients [77]. Graphics and tables can also help patients understand the cancer care information.

Information Services Targeting Specific Groups of Breast Cancer Patients

Attitudes of Patients

The synthesis revealed that patients' attitudes greatly affected their information needs [9,25,28,46]. Numerous other studies have also verified this view. For instance, individual cultural beliefs, views on health and disease, and other factors could affect women's dealing of early diagnosis [78,79], which is an essential strategy for preventing breast cancer [80,81]. Emotional distress may also affect patients' ability to cope with the symptoms and treatment of cancer [82].

To date, doctors have not paid enough attention to the emotional pressure patients feel when confronted with a diagnosis and treatment of breast cancer [83]. Thus, social support (eg, support groups, online health communities, and eHealth mobile apps) can play an active role in assisting patients with stress management and providing high-quality care [84,85]. Health care staff can also provide psychological support for breast cancer patients [86] by providing more mental health information. They can also encourage patients to express their genuine emotions and be willing to listen [87]. Finally, it is necessary for medical staff to give full respect to patients and protect their privacy [61].

XSL•FO
RenderX

Health Literacy of Patients

This study found that breast cancer patients' information needs can be influenced by health literacy [28,33,34]. Patients with limited health literacy commonly had more unsatisfied information needs [88]. Health literacy could also influence patients' cancer screening knowledge, attitudes, and behavior [89]. Inaccurate health beliefs or lack of health knowledge could lead to patients overlooking physical abnormalities, which could affect early detection and treatment of diseases.

People can often detect and treat breast cancer early by improving their health literacy. A higher level of health literacy can also promote effective communication with medical staff and increase patients' engagement in medical decisions [90]. For patients, limited literacy is associated with a low education level and socioeconomic status [91]. For medical practitioners, the knowledge of prevention, causes, early symptoms, and self-examination methods could be communicated better through mass media to improve patients' awareness [92]. Furthermore, educational materials on breast cancer prevention should be sensitive to health beliefs [93] to reduce women's fears of breast cancer screening and fatalism toward the disease. Physicians and other health care providers should continue to give priority to patients' quality of life when determining treatment plans as patients' decisions may be influenced by health knowledge. Additional training would help doctors communicate more effectively with patients with low health literacy and reduce medical inequalities.

Demographic Characteristics, Disease Status, and Political and Cultural Environment of Patients

The significant increase in the number of breast cancer patients has resulted in diverse and complex information needs. For example, there were more unmet information needs among breast cancer patients who had a foreign native language background (eg, immigrants) [94]. Moreover, the age-specific information needs of younger breast cancer survivors is still far from satisfactory [95]. Therefore, tailored information is recommended to help subgroups of breast cancer patients cope with cancer and improve satisfaction levels [96].

As shown in Table 2, this synthesis reveals suggestions and implications for tailored information services. First, some studies suggested that information provided to younger patients could focus more on the impact of treatment on fertility [8,9,35,44], lifestyle, and career as well as corresponding coping strategies [25,44,52]. More detailed advice on diet could also be provided for patients including many Asian patients. Patients who have young children can also be educated on how to communicate with their children. If patients are adequately informed about body changes before treatment, those who care a lot about body image [32,35] might have more time to consider and prepare for the changes. Policy information should also be given to patients who live in a country with changing medical policies. For patients with comorbidities and other physical conditions, specific prevention and survival guidelines are vital. Since different treatment options have corresponding side effects, health prescriptions that are targeted to each type of treatment can be distributed to patients separately to help them improve symptom management.

Limitations

Whereas the findings of this study offer significant insights, the limitations can lead to recommendations for further study. Some of the included studies did not clearly report cancer stages, making it difficult to analyze the findings according to the timeline of the disease journey. In addition, this theory-generating meta-synthesis only focused on qualitative studies, which could lead to the omission of valuable findings in quantitative research. Further research is needed to fully understand the unique information needs of breast cancer patients at all stages of their journey.

Conclusions

This synthesis has identified different information needs at various stages of the cancer continuum. Among the information needs, patients generally pay more attention to information about treatment and prognosis, as well as impact and resumption of normal life. Moderating variables were also identified. The generated model describes a complete pattern of the formation of information needs. Thus, this study contributes to a deeper understanding of breast cancer patients' information needs and provides practical suggestions for health professionals and information service providers. In particular, health care providers can offer educational materials according to the information needs identified in this study. More personalized information can also be developed to tailor patients' needs with reference to the moderating variables that influence their information needs.

Acknowledgments

HL designed the search strings and conducted the search. HL, JX, YC, and Y Cheng screened the papers for inclusion and appraised the quality of the papers. All researchers read and extracted the results, findings, and discussion of the studies independently. HL, JX, Y Cheng, and LS coded and synthesized the studies through an inductive analysis. All of the authors contributed to the writing and refinement of the paper. The authors would also like to express appreciation for the reviewers' valuable suggestions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Search strategies and results.
[DOCX File , 13 KB - jmir_v22i7e17907_app1.docx]

Multimedia Appendix 2
Included articles and quality assessment results.
[DOCX File , 42 KB - jmir_v22i7e17907_app2.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019 Jan 08;69(1):7-34. [doi: 10.3322/caac.21551]
3. McRoy S, Rastegar-Mojarad M, Wang Y, Ruddy KJ, Haddad TC, Liu H. Assessing unmet information needs of breast cancer survivors: exploratory study of online health forums using text classification and retrieval. JMIR Cancer 2018 May 15;4(1):e10 [FREE Full text] [doi: 10.2196/cancer.9050] [Medline: 29764801]
4. Findik UY. The information needs of women who have undergone breast cancer surgery in the west of Turkey. J Cancer Educ 2017 Sep;32(3):432-437. [doi: 10.1007/s13187-015-0965-2] [Medline: 26691514]
5. Tucker CA, Martin MP, Jones RB. Health information needs, source preferences and engagement behaviours of women with metastatic breast cancer across the care continuum: protocol for a scoping review. BMJ Open 2017 Feb 17;7(2):e013619 [FREE Full text] [doi: 10.1136/bmjopen-2016-013619] [Medline: 28213599]
6. Rees CE, Bath PA. The information needs and source preferences of women with breast cancer and their family members: a review of the literature published between 1988 and 1998. J Adv Nurs 2000 Apr;31(4):833-841. [doi: 10.1046/j.1365-2648.2000.01341.x] [Medline: 10759979]
7. Qualizza M, Bressan V, Rizzuto A, Stevanin S, Bulfone G, Cadorin L, et al. Listening to the voice of patients with head and neck cancer: a systematic review and meta-synthesis. Eur J Cancer Care (Engl) 2019 May;28(3):e12939. [doi: 10.1111/ecc.12939] [Medline: 30284763]
8. Thewes B, Meiser B, Rickard J, Friedlander M. The fertility- and menopause-related information needs of younger women with a diagnosis of breast cancer: a qualitative study. Psychooncology 2003;12(5):500-511. [doi: 10.1002/pon.685] [Medline: 12833562]
9. Gould J, Grassau P, Manthorne J, Gray RE, Fitch MI. 'Nothing fit me': nationwide consultations with young women with breast cancer. Health Expect 2006 Jun;9(2):158-173 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00383.x] [Medline: 16677195]
10. Finfgeld-Connett D. A Guide to Qualitative Meta-Synthesis. 1st Edition. New York: Routledge; 2018.
11. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 3rd Edition. Los Angeles: Sage Publications; 2008.
12. Adams E, McCann L, Armes J, Richardson A, Stark D, Watson E, et al. The experiences, needs and concerns of younger women with breast cancer: a meta-ethnography. Psychooncology 2011 Aug;20(8):851-861. [doi: 10.1002/pon.1792] [Medline: 20597066]
13. Banning M. Employment and breast cancer: a meta-ethnography. Eur J Cancer Care (Engl) 2011 Nov;20(6):708-719. [doi: 10.1111/j.1365-2354.2011.01291.x] [Medline: 21933291]
14. MeSH vocabulary. URL: https://www.ncbi.nlm.nih.gov/mesh [accessed 2020-07-02]
15. Case DO, Given LM. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. 4th Edition. Bingley: Emerald Group Publishing Limited; 2016.
16. Catalano A. Patterns of graduate students' information seeking behavior: a meta‐synthesis of the literature. J Documentation 2013 Mar;69(2):243-274. [doi: 10.1108/00220411311300066]
17. Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. Int J Med Inform 2003 Aug;71(1):9-15. [doi: 10.1016/s1386-5056(03)00023-6] [Medline: 12909153]
18. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
19. Schultz-Jones B. Examining information behavior through social networks: an interdisciplinary review. J Documentation 2009;65(4):592-631. [doi: 10.1108/00220410910970276]
20. Saab MM, Reidy M, Hegarty J, O'Mahony M, Murphy M, Von Wagner C, et al. Men's information-seeking behavior regarding cancer risk and screening: a meta-narrative systematic review. Psychooncology 2018 Feb;27(2):410-419. [doi: 10.1002/pon.4506] [Medline: 28728212]
21. Evans D, FitzGerald M. Reasons for physically restraining patients and residents: a systematic review and content analysis. Int J Nurs Stud 2002 Sep;39(7):735-743. [doi: 10.1016/s0020-7489(02)00015-9] [Medline: 12231030]

XSL•FO
RenderX

22. Critical ASP. Critical Appraisal Skills Programme. CASP qualitative checklist. URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf [accessed 2020-07-02]
23. Hull C. Principles of Behavior: An Introduction to Behavior Theory Concerning the Individual Organism. New York: Greenwood Press; 1943.
24. Gradishar W, Anderson B, Abraham J, Aft R, Agnese D. Breast Cancer, Version 2, NCCN Clinical Practice Guidelines in Oncology. 2019. URL: https://www2.tri-kobe.org/nccn/guideline/archive/breast2019/english/breast_v2.pdf [accessed 2020-07-02]
25. Halkett GKB, Kristjanson LJ, Lobb E, O'Driscoll C, Taylor M, Spry N. Meeting breast cancer patients' information needs during radiotherapy: what can we do to improve the information and support that is currently provided? Eur J Cancer Care (Engl) 2010 Jul;19(4):538-547. [doi: 10.1111/j.1365-2354.2009.01090.x] [Medline: 19708930]
26. Balka E, Krueger G, Holmes B. Situating internet use: information-seeking among young women with breast cancer. J Comput Mediated Comm 2010:389-411. [doi: 10.1111/j.1083-6101.2010.01506.x]
27. Beatty L, Oxlad M, Koczwara B, Wade TD. The psychosocial concerns and needs of women recently diagnosed with breast cancer: a qualitative study of patient, nurse and volunteer perspectives. Health Expect 2008 Dec;11(4):331-342 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00512.x] [Medline: 19076662]
28. Burton M, Collins KA, Lifford KJ, Brain K, Wyld L, Caldon L, et al. The information and decision support needs of older women (>75 yrs) facing treatment choices for breast cancer: a qualitative study. Psychooncology 2015 Aug;24(8):878-884. [doi: 10.1002/pon.3735] [Medline: 25534045]
29. Cappiello M, Cunningham R, Knobf M. Breast cancer survivors: information and support after treatment. Clin Nurs Res 2007;16(4):278-293. [doi: 10.1177/1054773807306553]
30. Gray RE, Fitch M, Greenberg M, Hampson A, Doherty M, Labrecque M. The information needs of well, longer-term survivors of breast cancer. Patient Educ Couns 1998 Mar;33(3):245-255. [doi: 10.1016/s0738-3991(98)00024-x] [Medline: 9731162]
31. Halkett GKB, Arbon P, Scutter SD, Borg M. The phenomenon of making decisions during the experience of early breast cancer. Eur J Cancer Care (Engl) 2007 Jul;16(4):322-330. [doi: 10.1111/j.1365-2354.2007.00778.x] [Medline: 17587355]
32. Tsuchiya M, Horn SA. An exploration of unmet information needs among breast cancer patients in Japan: a qualitative study. Eur J Cancer Care (Engl) 2009 Mar;18(2):149-155. [doi: 10.1111/j.1365-2354.2008.00936.x] [Medline: 19267730]
33. Kwok C, White K. Cultural and linguistic isolation: the breast cancer experience of Chinese-Australian women—a qualitative study. Contemp Nurse 2011 Aug;39(1):85-94. [doi: 10.5172/conu.2011.39.1.85] [Medline: 21955269]
34. McWilliam CL, Brown JB, Stewart M. Breast cancer patients' experiences of patient-doctor communication: a working relationship. Patient Educ Couns 2000 Feb;39(2-3):191-204. [doi: 10.1016/s0738-3991(99)00040-3] [Medline: 11040719]
35. Karaöz B, Aksu H, Küçük M. A qualitative study of the information needs of premenopausal women with breast cancer in terms of contraception, sexuality, early menopause, and fertility. Int J Gynaecol Obstet 2010 May;109(2):118-120. [doi: 10.1016/j.ijgo.2009.11.027] [Medline: 20152978]
36. Obeidat RF, Lally RM. Health-related information exchange experiences of Jordanian women at breast cancer diagnosis. J Cancer Educ 2014 Sep;29(3):548-554. [doi: 10.1007/s13187-013-0574-x] [Medline: 24158903]
37. Burke NJ, Napoles TM, Banks PJ, Orenstein FS, Luce JA, Joseph G. Survivorship care plan information needs: perspectives of safety-net breast cancer patients. PLoS One 2016;11(12):e0168383 [FREE Full text] [doi: 10.1371/journal.pone.0168383] [Medline: 27992491]
38. Lee S, Chen L, Ma GX, Fang CY. What is lacking in patient-physician communication: perspectives from Asian American breast cancer patients and oncologists. J Behav Health 2012;1(2):1-10 [FREE Full text] [doi: 10.5455/jbh.20120403024919] [Medline: 24496377]
39. Radina ME, Ginter AC, Brandt J, Swaney J, Longo DR. Breast cancer patients' use of health information in decision making and coping. Cancer Nurs 2011;34(5):E1-E12. [doi: 10.1097/NCC.0b013e3181faab7c] [Medline: 21116176]
40. Clark J. Constructing expertise: inequality and the consequences of information-seeking by breast cancer patients. Illness Crisis Loss 2005;13(2):169-185. [doi: 10.1177/105413730501300207]
41. Royak-Schaler R, Passmore SR, Gadalla S, Hoy MK, Zhan M, Tkaczuk K, et al. Exploring patient-physician communication in breast cancer care for African American women following primary treatment. Oncol Nurs Forum 2008 Sep;35(5):836-843. [doi: 10.1188/08.ONF.836-843] [Medline: 18765331]
42. Latifi M, Salimi S, Barahmand N, Fahimnia F, Allahbakhshian Farsani L. Postmastectomy information needs and information-seeking motives for women with breast cancer. Adv Biomed Res 2018;7:75 [FREE Full text] [doi: 10.4103/abr.abr_187_17] [Medline: 29862224]
43. Haq R, Heus L, Baker NA, Dastur D, Leung F, Leung E, et al. Designing a multifaceted survivorship care plan to meet the information and communication needs of breast cancer patients and their family physicians: results of a qualitative pilot study. BMC Med Inform Decis Mak 2013 Jul 25;13:76 [FREE Full text] [doi: 10.1186/1472-6947-13-76] [Medline: 23883430]
44. Thewes B, Butow P, Girgis A, Pendlebury S. The psychosocial needs of breast cancer survivors: a qualitative study of the shared and unique needs of younger versus older survivors. Psychooncology 2004 Mar;13(3):177-189. [doi: 10.1002/pon.710] [Medline: 15022153]

XSL•FO
RenderX

45. Shaw CR, Wilson SA, O'Brien ME. Information needs prior to breast biopsy. Clin Nurs Res 1994 May;3(2):119-131. [doi: 10.1177/105477389400300205] [Medline: 8173518]
46. Robinson JD, Venetis M, Street RL, Kearney T. Breast cancer patients' information seeking during surgical consultations: a qualitative, videotape-based analysis of patients' questions. J Surg Oncol 2016 Dec;114(8):922-929. [doi: 10.1002/jso.24470] [Medline: 27734517]
47. Wong JJ, D'Alimonte L, Angus J, Paszat L, Soren B, Szumacher E. What do older patients with early breast cancer want to know while undergoing adjuvant radiotherapy? J Cancer Educ 2011 Jun;26(2):254-261. [doi: 10.1007/s13187-010-0188-5] [Medline: 21221883]
48. Wen K, Hu A, Ma GX, Fang CY, Daly MB. Information and communication needs of Chinese American breast cancer patients: perspectives on survivorship care planning. J Community Support Oncol 2014 Dec;12(12):439-445 [FREE Full text] [doi: 10.12788/jcso.0095] [Medline: 25811036]
49. Kwok C, White K. Perceived information needs and social support of Chinese-Australian breast cancer survivors. Support Care Cancer 2014 Oct;22(10):2651-2659. [doi: 10.1007/s00520-014-2252-x] [Medline: 24752565]
50. Gray R, Fitch M, Davis C, Phillips C. A qualitative study of breast cancer self-help groups. Psychooncology 1997 Dec;6(4):279-289. [doi: 10.1002/(SICI)1099-1611(199712)6:4<279::AID-PON280>3.0.CO;2-0] [Medline: 9451747]
51. Luker KA, Beaver K, Leinster SJ, Owens RG. Information needs and sources of information for women with breast cancer: a follow-up study. J Adv Nurs 1996 Mar;23(3):487-495. [doi: 10.1111/j.1365-2648.1996.tb00010.x] [Medline: 8655823]
52. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004 Apr;53(1):57-64. [doi: 10.1016/S0738-3991(03)00116-2] [Medline: 15062905]
53. Balneaves LG, Panagiotoglou D, Brazier ASA, Lambert LK, Porcino A, Forbes M, et al. Qualitative assessment of information and decision support needs for managing menopausal symptoms after breast cancer. Support Care Cancer 2016 Nov;24(11):4567-4575 [FREE Full text] [doi: 10.1007/s00520-016-3296-x] [Medline: 27278271]
54. D'Alimonte L, Angus J, Wong J, Paszat L, Soren B, Szumacher E. Working toward a decision: the development and first impressions of a decision aid for older women with early-stage breast cancer. J Med Imaging Radiat Sci 2012 Mar;43(1):60-65. [doi: 10.1016/j.jmir.2011.08.005] [Medline: 31052023]
55. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate: service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
56. Weber KM, Solomon DH. Locating relationship and communication issues among stressors associated with breast cancer. Health Commun 2008 Nov;23(6):548-559. [doi: 10.1080/10410230802465233] [Medline: 19089702]
57. Balneaves LG, Truant TLO, Kelly M, Verhoef MJ, Davison BJ. Bridging the gap: decision-making processes of women with breast cancer using complementary and alternative medicine (CAM). Support Care Cancer 2007 Aug;15(8):973-983. [doi: 10.1007/s00520-007-0282-3] [Medline: 17609997]
58. Kenny P, Quine S, Shiell A, Cameron S. Participation in treatment decision-making by women with early stage breast cancer. Health Expect 1999 Sep;2(3):159-168 [FREE Full text] [doi: 10.1046/j.1369-6513.1999.00050.x] [Medline: 11281892]
59. Reaby LL. The quality and coping patterns of women's decision-making regarding breast cancer surgery. Psychooncology 1998;7(3):252-262. [doi: 10.1002/(SICI)1099-1611(199805/06)7:3<252::AID-PON309>3.0.CO;2-O] [Medline: 9638786]
60. Kemp E, Koczwara B, Butow P, Turner J, Girgis A, Schofield P, et al. Online information and support needs of women with advanced breast cancer: a qualitative analysis. Support Care Cancer 2018 Oct;26(10):3489-3496. [doi: 10.1007/s00520-018-4206-1] [Medline: 29693203]
61. Boman L, Andersson JU, Björvell H. Needs as expressed by women after breast cancer surgery in the setting of a short hospital stay. Scand J Caring Sci 1997;11(1):25-32. [doi: 10.1111/j.1471-6712.1997.tb00427.x] [Medline: 9275819]
62. Wolf L. The information needs of women who have undergone breast reconstruction. Part I: decision-making and sources of information. Eur J Oncol Nurs 2004 Sep;8(3):211-223. [doi: 10.1016/j.ejon.2003.12.012] [Medline: 15304229]
63. Lee S, Chen L, Ma GX, Fang CY, Oh Y, Scully L. Challenges and needs of Chinese and Korean American breast cancer survivors: in-depth interviews. N Am J Med Sci (Boston) 2013 Feb 02;6(1):1-8 [FREE Full text] [doi: 10.7156/najms.2013.0601001] [Medline: 24019995]
64. Holmes MM, Bishop FL, Calman L. “I just googled and read everything”: exploring breast cancer survivors’ use of the internet to find information on complementary medicine. Complement Ther Med 2017 Aug;33:78-84. [doi: 10.1016/j.ctim.2017.06.007] [Medline: 28735830]
65. Nguyen SKA, Ingledew P. Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. J Cancer Educ 2013 Dec;28(4):662-668. [doi: 10.1007/s13187-013-0509-6] [Medline: 23832398]
66. Brown M, Koch T, Webb C. Information needs of women with non-invasive breast cancer. J Clin Nurs 2000 Sep;9(5):713-722. [doi: 10.1046/j.1365-2702.2000.00419.x]
67. Shands ME, Lewis FM, Zahlis EH. Mother and child interactions about the mother's breast cancer: an interview study. Oncol Nurs Forum 2000;27(1):77-85. [Medline: 10660925]

XSL•FO
RenderX

68. Barnes J, Kroll L, Burke O, Lee J, Jones A. Qualitative interview study of communication between parents and children about maternal breast cancer. BMJ 2000;321(7259):479-482. [doi: 10.1136/bmj.321.7259.479]
69. Ussher JM, Perz J, Gilbert E. Information needs associated with changes to sexual well-being after breast cancer. J Adv Nurs 2013 Feb;69(2):327-337. [doi: 10.1111/j.1365-2648.2012.06010.x] [Medline: 22500731]
70. Ham M, Jeger M, Frajman Ivković A. The role of subjective norms in forming the intention to purchase green food. Economic Research-Ekonomska Istraživanja 2015 Oct 09;28(1):738-748. [doi: 10.1080/1331677x.2015.1083875]
71. Breckler SJ. Empirical validation of affect, behavior, and cognition as distinct components of attitude. J Personality Soc Psychol 1984;47(6):1191-1205. [doi: 10.1037/0022-3514.47.6.1191]
72. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, et al. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
73. Zhao Q, Kulane A, Gao Y, Xu B. Knowledge and attitude on maternal health care among rural-to-urban migrant women in Shanghai, China. BMC Womens Health 2009 Mar 20;9:1-8 [FREE Full text] [doi: 10.1186/1472-6874-9-5] [Medline: 19298681]
74. Howell BM. Interactions between diet, physical activity, and the sociocultural environment for older adult health in the urban Subarctic. J Community Health 2019 Sep 11;9:1-12. [doi: 10.1007/s10900-019-00737-3] [Medline: 31512112]
75. Schmidt A, Kowalski C, Pfaff H, Wesselmann S, Wirtz M, Ernstmann N. The influence of health literacy on information needs among women newly diagnosed with breast cancer, with special reference to employment status. J Health Commun 2015;20(10):1177-1184. [doi: 10.1080/10810730.2015.1018626] [Medline: 26086892]
76. National Academies Press. Health literacy: a prescription to end confusion. National Academy of Medicine 2004;15(4):389-395. [Medline: 25009856]
77. Jimenez YA, Cumming S, Wang W, Stuart K, Thwaites DI, Lewis SJ. Patient education using virtual reality increases knowledge and positive experience for breast cancer patients undergoing radiation therapy. Support Care Cancer 2018 Aug;26(8):2879-2888. [doi: 10.1007/s00520-018-4114-4] [Medline: 29536200]
78. Kissal A, Beşer A. Knowledge, facilitators and perceived barriers for early detection of breast cancer among elderly Turkish women. Asian Pac J Cancer Prev 2011;12(4):975-984 [FREE Full text] [Medline: 21790237]
79. Champion VL, Skinner CS, Foster JL. The effects of standard care counseling or telephone/in-person counseling on beliefs, knowledge, and behavior related to mammography screening. Oncol Nurs Forum 2000;27(10):1565-1571. [Medline: 11103375]
80. Charkazi A, Samimi A, Razzaghi K, Kouchaki GM, Moodi M, Meirkarimi K, et al. Adherence to recommended breast cancer screening in Iranian turkmen women: the role of knowledge and beliefs. ISRN Prev Med 2013;2013:581027 [FREE Full text] [doi: 10.5402/2013/581027] [Medline: 24977094]
81. Duman N. Health beliefs of the female academicians about breast cancer, screening tests and the affecting factors. Int J Hematol Oncol 2013 Dec 1;23(4):233-241. [doi: 10.4999/uhod.13021]
82. Juarez G, Hurria A, Uman G, Ferrell B. Impact of a bilingual education intervention on the quality of life of Latina breast cancer survivors. Oncol Nurs Forum 2013 Jan;40(1):E50-E60 [FREE Full text] [doi: 10.1188/13.ONF.E50-E60] [Medline: 23269782]
83. Castillo A, Mendiola J, Tiemensma J. Emotions and coping strategies during breast cancer in Latina women: a focus group study. Hisp Health Care Int 2019 Sep;17(3):96-102. [doi: 10.1177/1540415319837680] [Medline: 30922185]
84. Kyngäs H, Mikkonen R, Nousiainen EM, Rytilahti M, Seppänen P, Vaattovaara R, et al. Coping with the onset of cancer: coping strategies and resources of young people with cancer. Eur J Cancer Care (Engl) 2001 Mar;10(1):6-11. [doi: 10.1046/j.1365-2354.2001.00243.x] [Medline: 11827269]
85. Dumrongpanapakorn P, Liamputtong P. Social support and coping means: the lived experiences of Northeastern Thai women with breast cancer. Health Promot Int 2017 Oct 01;32(5):768-777. [doi: 10.1093/heapro/dav023] [Medline: 25876908]
86. Drageset S, Lindstrøm TC, Giske T, Underlid K. Women's experiences of social support during the first year following primary breast cancer surgery. Scand J Caring Sci 2016 Jun;30(2):340-348. [doi: 10.1111/scs.12250] [Medline: 26179112]
87. Yoo GJ, Levine EG, Pasick R. Breast cancer and coping among women of color: a systematic review of the literature. Support Care Cancer 2014 Mar;22(3):811-824 [FREE Full text] [doi: 10.1007/s00520-013-2057-3] [Medline: 24389825]
88. Halbach SM, Ernstmann N, Kowalski C, Pfaff H, Pförtner T, Wesselmann S, et al. Unmet information needs and limited health literacy in newly diagnosed breast cancer patients over the course of cancer treatment. Patient Educ Couns 2016 Sep;99(9):1511-1518. [doi: 10.1016/j.pec.2016.06.028] [Medline: 27378079]
89. Peterson NB, Dwyer KA, Mulvaney SA, Dietrich MS, Rothman RL. The influence of health literacy on colorectal cancer screening knowledge, beliefs and behavior. J Natl Med Assoc 2007 Oct;99(10):1105-1112. [Medline: 17987913]
90. Shen H, Lin C, Hoffmann T, Tsai C, Hou W, Kuo KN. The relationship between health literacy and perceived shared decision making in patients with breast cancer. Patient Educ Couns 2019 Feb;102(2):360-366. [doi: 10.1016/j.pec.2018.09.017] [Medline: 30270171]
91. Castle M, Adams N, Coriell A, Buffinger A, Crabtree K, Liston J. Improving health literacy for positive patient outcomes. 2019 Presented at: Celebration of Scholarship Conference; 2019; University Heights.

XSL•FO
RenderX

92. Ramos Herrera IM. Policies and programs for the prevention and control of breast cancer in Mexican and Latin American women: protocol for a scoping review. JMIR Res Protoc 2019 Jun 12;8(6):e12624 [FREE Full text] [doi: 10.2196/12624] [Medline: 31199301]
93. Kissal A, Ersin F, Koç M, Vural B, Çetin ?. Determination of women's health beliefs, breast cancer fears, and fatalism associated with behaviors regarding the early diagnosis of breast cancer. Int J Cancer Manag 2018 Nov 17;11(12):1-5. [doi: 10.5812/ijcm.80223]
94. Kowalski C, Lee SD, Ansmann L, Wesselmann S, Pfaff H. Meeting patients' health information needs in breast cancer center hospitals—a multilevel analysis. BMC Health Serv Res 2014 Nov 25;14(1):1-9 [FREE Full text] [doi: 10.1186/s12913-014-0601-6] [Medline: 25422099]
95. Miyashita M, Ohno S, Kataoka A, Tokunaga E, Masuda N, Shien T, et al. Unmet information needs and quality of life in young breast cancer survivors in Japan. Cancer Nurs 2015;38(6):E1-E11 [FREE Full text] [doi: 10.1097/NCC.0000000000000201] [Medline: 25254410]
96. Hurren SJ, Yates K. At what point can I lift things? Women's satisfaction with lymphoedema prevention information after breast cancer surgery. Collegian 2019 Jun;26(3):335-340. [doi: 10.1016/j.colegn.2018.09.001]

Edited by G Eysenbach; submitted 21.01.20; peer-reviewed by M Behzadifar, S McRoy, T Haddad; comments to author 22.03.20; revised version received 07.05.20; accepted 03.06.20; published 28.07.20.

Please cite as:
Lu H, Xie J, Gerido LH, Cheng Y, Chen Y, Sun L
Information Needs of Breast Cancer Patients: Theory-Generating Meta-Synthesis
J Med Internet Res 2020;22(7):e17907
URL: https://www.jmir.org/2020/7/e17907
doi:10.2196/17907
PMID:32720899

©Hongru Lu, Juan Xie, Lynette Hammond Gerido, Ying Cheng, Ya Chen, Lizhu Sun. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Evidence Regarding Automatic Processing Computerized Tasks Designed For Health Interventions in Real-World Settings Among Adults: Systematic Scoping Review

Harshani Jayasinghe[1], BSc, MPhil; Camille E Short[2], DPhil, BPsyc; Annette Braunack-Mayer[3], DPhil, BMSc; Ashley Merkin[4], BSc; Clare Hume[1], DPhil, BASc

[1]School of Public Health, The University of Adelaide, Adelaide, Australia

[2]Melbourne Centre for Beahviour Change, Melbourne School of Psychological Science and Melbourne School of Health Science, University of Melbourne, Melbourne, Australia

[3]School of Health and Society, University of Wollongong, Wollongong, Australia

[4]School of Medicine, The University of Adelaide, Adelaide, Australia

Corresponding Author:
Harshani Jayasinghe, BSc, MPhil
School of Public Health
The University of Adelaide
Level 9, Adelaide Health and Medical Sciences Building
The University of Adelaide
Adelaide, 5000
Australia
Phone: 61 8313 4634
Email: harshani.jayasinghepedige@adelaide.edu.au

Abstract

Background: Dual process theories propose that the brain uses 2 types of thinking to influence behavior: automatic processing and reflective processing. Automatic processing is fast, immediate, nonconscious, and unintentional, whereas reflective processing focuses on logical reasoning, and it is slow, step by step, and intentional. Most digital psychological health interventions tend to solely target the reflective system, although the automatic processing pathway can have strong influences on behavior. Laboratory-based research has highlighted that automatic processing tasks can create behavior change; however, there are substantial gaps in the field on the design, implementation, and delivery of automatic processing tasks in real-world settings. It is important to identify and summarize the existing literature in this area to inform the translation of laboratory-based research to real-world settings.

Objective: This scoping review aims to explore the effectiveness of automatic training tasks, types of training tasks commonly used, mode of delivery, and impacts of gamification on automatic processing tasks designed for digital psychological health interventions in real-world settings among adults.

Methods: The scoping review methodology proposed by Arskey and O'Malley and Colquhoun was applied. A scoping review was chosen because of the novelty of the digital automatic processing field and to encompass a broad review of the existing evidence base. Electronic databases and gray literature databases were searched with the search terms "automatic processing," "computerised technologies," "health intervention," "real-world," and "adults" and synonyms of these words. The search was up to date until September 2018. A manual search was also completed on the reference lists of the included studies.

Results: A total of 14 studies met all inclusion criteria. There was a wide variety of health conditions targeted, with the most prevalent being alcohol abuse followed by social anxiety. Attention bias modification tasks were the most prevalent type of automatic processing task, and the majority of tasks were most commonly delivered over the web via a personal computer. Of the 14 studies included in the review, 8 demonstrated significant changes to automatic processes and 4 demonstrated significant behavioral changes as a result of changed automatic processes.

Conclusions: This is the first review to synthesize the evidence on automatic processing tasks in real-world settings targeting adults. This review has highlighted promising, albeit limited, research demonstrating that automatic processing tasks may be used effectively in a real-world setting to influence behavior change.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e17915) doi:10.2196/17915

KEYWORDS

automatic processing; computerized tasks; health interventions; real-world; adult; behavior change, mobile phone

Introduction

Background

Digital psychological health interventions, which encompass both behavior change and mental health interventions, are increasingly being adopted to prevent and manage chronic health conditions. This includes stroke [1], dementia [2], obesity [3], addiction [4], and mood disorder [5]. The effectiveness of these interventions may be increased by the adoption of dual process theories that propose that the brain uses 2 types of processes to influence behavior [6-8]: automatic processing and reflective processing.

Automatic processing is fast, immediate, unconscious, and unintentional [9,10]. It involves an appraisal of a stimulus (eg, a basketball) in terms of its affective and conditioned properties. A mental representation of the stimulus and the network of associated concepts is automatically activated. By activating these mental networks, stimuli can capture or repel attention (attentional bias), lead to an automatic judgment of the stimulus as good or bad (automatic evaluations), and elicit a tendency to approach or avoid the stimulus (approach or avoidance bias). Cognitive biases are mostly automatic processes that result in individuals giving increased attention to threatening stimuli, with the difficulties in disengaging from these stimuli [11]. In contrast, reflective processing focuses on logical reasoning; it is slow, step by step, voluntary, and intentional, such as self-regulatory processes [9,10]. These 2 types of processes can occur simultaneously and can have concordant or opposing influences on behavior [10]. For example, when a person sees a basketball, this will trigger a network of concepts in the procedural memory, such as *fun* and *good* or *hard* and *bad*, which can lead to an automatic response from the individual to either approach or avoid the basketball (approach-avoidance bias). If the stimuli are automatically perceived as positive, it will result in an approach response, whereas if the stimuli are automatically perceived as negative, it will result in an avoidance response [12,13]. In this instance, these automatic associations could influence the formation of intentions to play basketball in the first place or may contribute to an intention-behavior gap if the conscious, logical intention is at odds with the automatic association.

Given that both types of processes can occur simultaneously and that behavior change may be most likely when they are congruent, the dual process theory suggests that both the reflective and automatic processing pathways should be targeted to create and sustain behavior change [8,14]. However, most digital behavior change and mental health interventions tend to target the reflective system only, aiming to improve self-regulatory processes, for example, processes such as self-monitoring behavior, goal-setting, or cognitive behavioral therapy [10,15]. Although these strategies are useful for supporting intention or goal-directed behavior, there is also clear evidence of an intention-behavior gap [16].

Laboratory-based evidence suggests that digital interventions delivered via computerized tasks can change people's automatic processes to specific stimuli [14,17-19]. Automatic processes can be retrained using cognitive bias modifications (CBM) tasks, attentional bias modification tasks, and evaluative conditioning tasks [14,18,20-22]. CBM tasks have been extensively tested in the field of anxiety and depression in laboratory settings [14,18,20-22] and work by targeting attentional and interpretative biases away from threats [14]. The dot-probe test is often used to assess attentional biases [23] and involves the presentation of pairs of stimuli, one of which is threatening and one that is neutral. Participants are shown the stimuli simultaneously, one stimulus on either end of a computer screen, for a small amount of time (eg, 500 ms); a dot then appears in the place of one former stimulus, and participants are asked to indicate the location of this dot as quickly as possible. The computer automatically measures the speed of this reaction. The test is then repeated several times. Quicker response times to the dot when it appears in the location of threatening stimuli are interpreted as attentiveness to threat [23]. Dot-probe tasks can also be used to retrain attentional biases by replacing a targeted probe (designed to change attentional biases) with neutral or salient stimuli during all cycles of the task [19,20,24-26]. A commonly used task shows participants a smiling or neutral face alongside an unhappy or angry face; participants are then asked to select the positive image as fast as possible [14]. Through the repetitive nature of the task, a person's automatic processes can be reconditioned toward a particular stimulus [14].

Common criticisms of automatic processing tasks are that they are repetitive and boring [18,27]. Recently, the field has attempted to make these tasks more enjoyable and accessible by delivering these tasks on digitally ubiquitous technology such as smartphone apps and drawing on the developing realm of gamification. Gamification seeks to motivate and engage users by using gaming style elements often seen in games (rewards, points, and leader boards). A review by Boendermaker et al [19] into gamified CBM highlighted that there were many projects currently in progress; however, there still remained a lack of evidence to draw any firm conclusions regarding the effectiveness of gamification on increasing engagement. Similarly, Zhang et al [11] highlighted that understanding gamification approaches is crucial in future conceptualization and codesign of attention bias modification interventions.

Objectives

Despite the evidence of the importance of the automatic processing pathway for regulating behavior, the majority of reviews on automatic processing tasks have been limited, focusing mainly on mental health conditions delivered in laboratory settings [17,18,28-30]. In addition, existing reviews have routinely included studies with children whose brain development differs from that of adults [31-33]; thus, a need to map the existing literature examining automatic processing

XSL•FO
RenderX

pathways in digital health interventions in real-world settings among adult populations was identified [16,27,34,35]. This scoping review aimed to explore the effectiveness of automatic training tasks, types of training tasks commonly used, mode of delivery, and impacts of gamification on automatic processing tasks designed for digital psychological health interventions in real-world settings among adults.

Methods

Scoping Review Methodology

Scoping reviews aim to "map rapidly the key concepts underpinning a research area and the main sources and types of evidence available" [36]. This scoping review sought to address the question: "what is the current evidence base around design and effectiveness of automatic processing computerized tasks designed for health interventions in real-world settings among adults?" A scoping review was chosen because of the novelty of the digital automatic processing field and to encompass a broad review of the evidence. Scoping reviews allow the development of inclusion and exclusion criteria during the study selection, the inclusion of all types of studies, and extraction of data regarding key issues and themes, in contrast to systematic reviews that are much more stringent with synthesis [37,38].

This scoping review followed the framework described by Arksey and O'Malley [38], who provided a detailed description of how to conduct a methodologically rigorous scoping review. In addition, the current best practices of Colquhoun et al [37] for the conduct of scoping reviews' guidelines were also applied. Arksey and O'Malley's methodology [38] comprised the following steps: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. A number of additional recommendations from a study by Colquhoun et al [37] were also integrated: development of a protocol before the initial scoping study began, using Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Protocols, the use of 2 independent reviewers, use of same inclusion criteria during initial data screening, and full-text screening and pilot testing of the data extraction template.

Identifying Relevant Studies

Search Strategy

A systematic literature search was created with the assistance of an academic librarian from the University of Adelaide. It was applied to all databases by HJ. The search was completed using the electronic databases PubMed, Scopus, Excerpta Medica Database (EMBASE), Psychological Information Database (PsycINFO), Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Database of Systematic Reviews, and Google Scholar. The search was performed in September 2018. A gray literature search was also conducted to identify any published or unpublished data not found through the initial search. This included electronically searching the repositories holding theses and research papers: Trove, The Gray Literature Report, ProQuest, OpenGrey, and Grey Literature Network Service (GreyNet International). Reference lists of all included studies were manually screened to identify possible relevant citations. The search strategy created was based on 5 main components: (1) automatic processing, (2) computerized tasks, (3) health interventions, (4) real world, and (5) adult. Relevant keywords were identified using Medical Subjects Headings (MeSH) and Emtree terms, synonyms, and keywords from relevant articles (Multimedia Appendix 1).

The results of the search were uploaded into EndNote X 7.3.1 (Clarivate Analytics), and duplicates were then removed and exported into Rayyan. Rayyan is a web-based tool that assists in the completion of systematic reviews. Rayyan was initially used by HJ to screen titles search and determine eligibility. This was followed by an abstract screen using the inclusion and exclusion criteria. If eligibility was ambiguous, criteria were discussed with other coauthors until consensus was reached. An eligibility proforma was also used during this process. Articles matching the inclusion criteria were then selected for full-text analysis.

Inclusion and Exclusion Criteria

Inclusion criteria for the study were adults aged older than 18 years, designed to be a health intervention (defined as aiming to improve physical or mental health), delivered via a computerized (digital) task, and delivered in a real-world setting (a free-living environment, excluding a laboratory-based environment). Only studies published in English between the years 2000 and 2018 were included. Articles not published in English were excluded. Articles were also excluded if they were published before the year 2000 because of digital expansion in the field of automatic processing predominately occurring after the year 2000.

Charting the Data

Data extraction was performed using a standardized data extraction template with the following fields: author, year, country, aims of the study, setting and population, participant demographics, details of the intervention and comparators, study methodology, sampling and recruitment, completion rates, and intervention details (Multimedia Appendix 2). HJ, CES, CH, and AB all conducted data extraction for the first 2 publications using the original data extraction template. All reviewers then discussed any iterations, and the template was changed appropriately to reflect any inconsistencies. Data for all remaining studies were extracted by HJ, with 10% being verified by AM. Conflicts or concerns during this period were resolved through discussion with CES and CH.

Collating, Summarizing, and Reporting the Results

The data extraction forms were used to form quick overview summaries of the included studies. A descriptive numerical summary was used to create a numerical overview of general study characteristics, and then a narrative overview was conducted on the type of automatic processing task used and effectiveness.

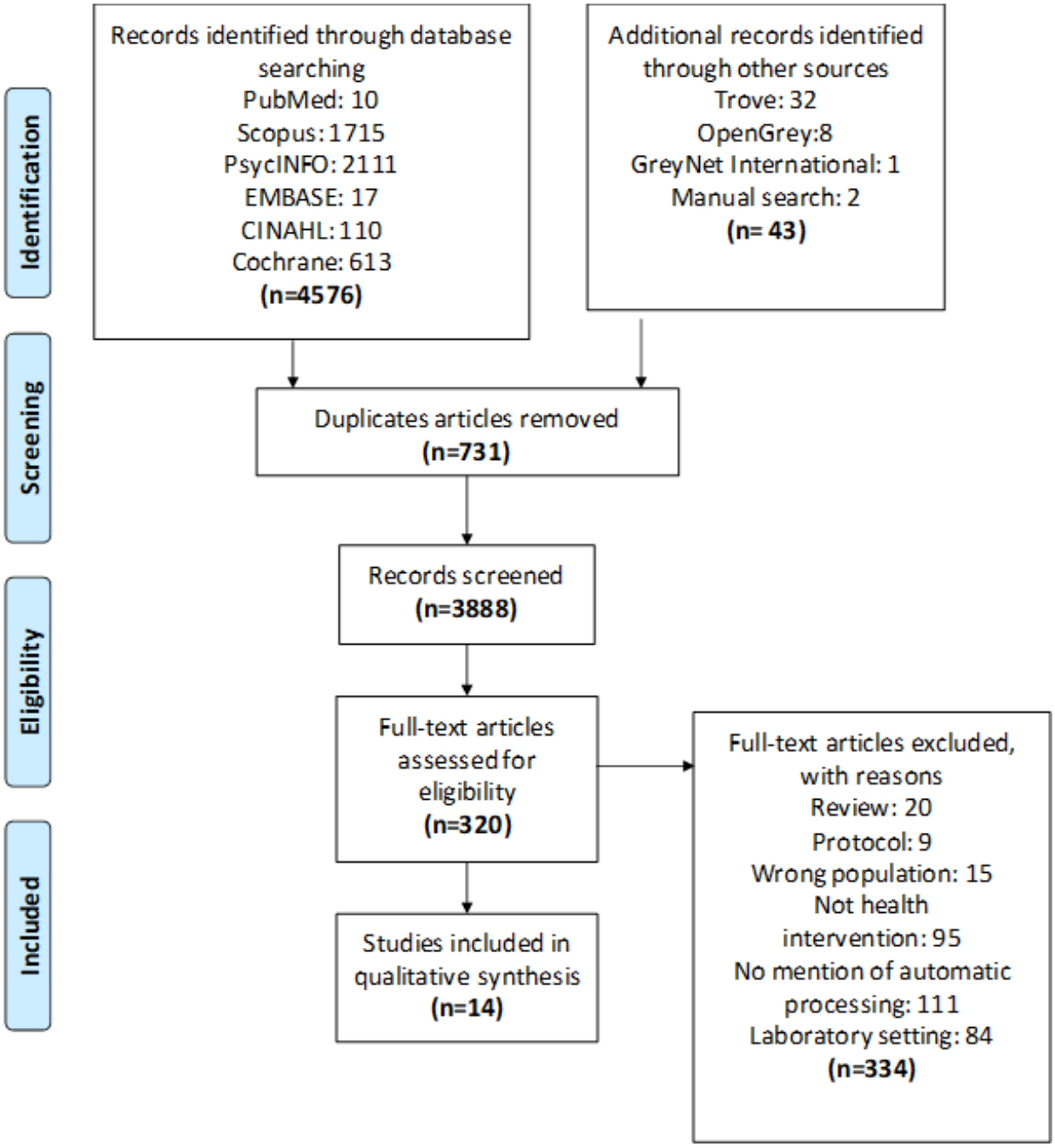
XSL•FO
RenderX

Results

PRISMA Results

The PRISMA flow chart for the results of the search is presented in Figure 1. The flowchart contains results from the initial search, how many studies were removed because of duplications, study selection, and amount selected for full-text analysis (Figure 1). The search is up to date until September 2018; 4576 studies were found to be eligible for inclusion (Figure 1). Of these, 320 were assessed for full-text inclusion, of which 14 studies met all inclusion criteria. The reasons for study exclusion included being a review paper, irrelevant to the topic, and not being set in a real-world setting.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses for the results of the search. CINAHL: Cumulative Index of Nursing and Allied Health Literature; EMBASE: Excerpta Medica Database; GreyNet: Grey Literature Network Service; PsycINFO: Psychological Information Database.



General Characteristics

Characteristics including the type of study, sample size, country, and follow-up are reported in Table 1. Most of the included studies were conducted in Europe and the United States, with 6 in Europe and 5 in the United States. One study was conducted in China [39] and another in Australia [34]. All studies were randomized controlled trials, except one that was a cross-sectional study. Sample sizes ranged from 21 to 672 participants. Five of the studies targeted automatic processes

XSL•FO
RenderX

regarding alcoholism [19,20,31,32,35]. Two studies targeted smoking attitudes and behavior [35,40]. Two studies targeted automatic processes toward anxiety [21,39], and one study targeted automatic processing toward both anxiety and depression [20]. In addition, one study was found for insomnia [34], self-injurious thoughts and behaviors [41], attitudes toward homosexuality and self-esteem in gay men [42], and relationship satisfaction in marriages [43]. Follow-up periods also varied and ranged from no follow-up (n=4 studies [39,42,44,45]) to the longest follow-up period of 6 months [35].

Table 1. General demographics of included studies (n=14).

Reference	Type of study	Population size (N)	Country	Follow-up
Robinson et al (2017) [40]	RCT[a]	64	The United States	Two weekly follow-ups
Clarke et al (2016) [34]	RCT	36	Australia	8 days
Yang et al (2017) [39]	RCT	76	China	No follow-up
Fleming and Burns (2017) [42]	RCT	274	The United States	No follow-up
van Deursen et al (2015) [45]	Cross-sectional	437	Germany	No follow-up
de Voogd et al (2016) [20]	RCT	340	The Netherlands	Baseline measures were taken before training and then 3, 6, and 12 months after training
Enock et al (2016) [21]	RCT	429	The United States	Pretest, posttest, and 1- and 2-month follow-ups
Monk et al (2017) [44]	RCT	62	The United Kingdom	No follow-up
Wiers et al (2015) [26]	RCT	615	The Netherlands	The first follow-up was 1 month after the posttest (N=109, 35.0%) and the second 2 months later (N=87, 28.0%)
Boendermaker et al (2016) [25]	RCT	96	The Netherlands	Four sessions conducted at least one day apart and 2-week follow-up after session 4
Crane et al (2018) [46]	RCT	672	The United Kingdom	The follow-up questionnaire was sent to participants 28 days after downloading the app
Elfeddali et al (2016) [35]	RCT	475	Amsterdam	A posttraining assessment (ie, manikin and visual probe tasks) followed the intervention sessions. An assessment of continued smoking abstinence followed 6 months after baseline
Franklin et al (2016) [41]	RCT	114, 131, and 163	The United States	1-month follow-up
McNulty et al (2017) [43]	RCT	288	The United States	2 weeks

[a]RCT: randomized controlled trial.

Intervention Characteristics

Types of Digital Technology

Table 2 reports the intervention characteristics of the included studies. The type of digital technology most often used to execute the intervention was a personal computer or laptop via a website (n=8) [20,25,26,35,42-45], followed by the use of a smartphone app (n=5) [21,34,39,41,46]. One study was delivered on a personal digital assistant device [40].

XSL•FO
RenderX

Table 2. Intervention characteristics of included studies (n=14).

Reference	Type of digital technology	Real-world setting
Robinson et al (2017) [40]	Hewlett-Packard iPAQ Personal Digital Assistant	Laboratory and home
Clarke et al (2016) [34]	Smartphone	Laboratory and home
Yang et al (2017) [39]	Smartphone	Home
Fleming and Burns (2017) [42]	Functioned only on desktop or laptop personal computers	Web-based
van Deursen et al (2015) [45]	Personal computer or laptop	Web-based
de Voogd et al (2016) [20]	Computer	Web-based at school and home
Enock et al (2014) [21]	Smartphone	Web-based
Monk et al (2017) [44]	Computer, mobile prohibited	Web-based
Wiers et al (2015) [26]	Computer	Web-based
Boendermaker et al (2016) [25]	Computer	Laboratory and home
Crane et al (2016) [46]	Smartphone app	Web-based
Elfeddali et al (2016) [35]	Computer	Web-based
Franklin et al (2016) [41]	Mobile app but can be used on phones, tablets, laptops, and desktops	Web-based
McNulty et al (2017) [43]	Computer	Laboratory and home

Use of Dual Process Theory

A total of 5 of the 14 included studies referred to dual process theory when explaining the theory behind the use of bias modification tasks [25,26,35,45,46]. Moreover, 7 studies did not mention any theories guiding their research [20,21,34,39-41,44]. Furthermore, 2 studies did not mention dual process theories but referred to other theories as a guide to their research [42,43].

Only 2 of the included studies [45,46] focused their interventions on targeting both parts of the dual process theory, targeting both automatic processing and reflective processing pathways. Crane et al [46] evaluated module components of an alcohol reduction app called *Drink Less* to change participants' drinking behavior in a factorial trial. Of the 5 modules tested, 1 (cognitive bias retraining module) targeted automatic processes regarding the impulse to drink alcohol. There was a significant two-way interaction between theis module on weekly alcohol consumption (P=.03), indicating that enhanced normative feedback (targeting misperceptions via reflective processes) led to a significant reduction in weekly alcohol consumption only when combined with enhanced cognitive bias retraining. van Deursen et al [45] examined the relationship among problem drinkers seeking web-based help to change their alcohol use, hypothesizing that executive functions would moderate the relationship between automatic associations and drinking and that this effect would be stronger in individuals with strong motivation to change. A brief *Implicit Association Test* was used to test valence and approach associations, whereas executive (reflective) pathways were assessed via a number of smaller tests. The study results provided partial support for the moderating role of motivation in the interplay between automatic processes and executive functions.

Types of Automatic Processing Tasks

Three types of automatic processing tasks were used in the included studies: attentional bias modification tasks, CBM tasks, and evaluative conditioning tasks. Attentional bias modification tasks were the most commonly delivered intervention through the use of dot-probe, visual probe, and visual search tasks. CBM tasks were delivered via dot-probe test [39], word sentence association paradigms [39], attentional control training, approach bias retraining [26], and cognitive bias retraining [46]. Evaluative conditioning was delivered via evaluative conditioning tasks [42,43], a stop-signal task, and a game-like therapeutic evaluative conditioning task.

Usage

Intervention durations varied considerably between the studies that reported on how long the intervention was delivered for, ranging from the shortest intervention period of 75 min to the longest intervention period of 6 weeks. All studies delivered at least some of their interventions in a real-world setting (as per inclusion criteria), with 4 studies delivering interventions in a mixed setting (both laboratory and home environments). Some studies have reported problems with usage. This included the number of training trials used over sessions being too low, preventing the training from potentially changing attentional biases [25]; low training compliance, with some participants dropping out of the study midway effecting overall results and limiting generalizability of findings [20,26,40,46]; and having self-selected use and dosage during the intervention, making it hard to determine the optimal dosage needed for intervention [41].

Gamification Strategies

Three of the included studies incorporated elements of gamification to enhance engagement of their intervention [25,41,46]. Boendermaker et al gamified [25] an existing automatic processing task, which aimed to train attention away from pictures of alcoholic beverages via a visual probe task.

XSL•FO
RenderX

They found that the gamified automatic processing task did not improve participant motivation to train as compared with the usual ungamified task. In fact, some aspects of motivation appeared to deteriorate rather than improve. Crane et al [46] and Franklin et al [41] saw positive effects with the use of gamified elements in their interventions. However, the effect that gamification had on influencing study outcomes was not investigated, making it difficult to deduce which aspects of gamification produced different effects and overall how effective gamification was.

Outcome Assessments

Two studies reported that it was possible that the option to train at home had a negative effect on the final results, as it may have made participants take the training less seriously. It was also noted that the web-based nature of web-based assessments resulted in issues with standardization of interventions, making it hard to control how measures were completed [45]. The large number of measures included in some studies may have also resulted in participant fatigue [45], and the use of similar types of tasks testing different outcomes may have resulted in practice effects that may have affected the overall findings.

Effectiveness of Interventions

A detailed description of the aims, measures, and effectiveness of included studies can be found in Multimedia Appendix 3 [14,15,19,20,29-34,36-39]. A total of 14 studies assessed changes in automatic processes, whereas 11 studies assessed how changes in automatic processing contributed to changes in behavior. Of the 14 studies that assessed changes in automatic processes, 8 reported a statistically significant effect in the direction expected [21,25,26,34,40,43,44,46]. Of the remaining 6 studies evaluating automatic processes, 4 reported no significant effects on changes in automatic processes [20,35,42,45].

Of the 11 studies that assessed behavioral or mental health outcomes, 4 reported a significant intervention effect [21,26,34,43] and 3 of these included automatic processing intervention strategies only. Five others reported on behavioral or mental outcomes but saw no significant results on changes to outcomes [20,25,35,42,46].

Three studies reported mixed findings [39-41] on changes to automatic processes, with 2 of the studies also further detailing changes to behavior. Robinson et al [40] found that a mobile-delivered attentional bias intervention could reduce attentional bias toward thoughts about smoking but had mixed effects on changing smoking behavior. It was thought that these findings may have been impacted by participant attrition affecting overall results and the sample being nontreatment seeking participants [40]. Yang et al found [39] that 1 of the 3 CBM tools assessed could be used effectively to reduce anxiety and mood problems. However, the other 2 methods assessed yielded limited effectiveness. The authors reported that a low sample size and a lack of engagement elements to make the tasks *fun* may have impacted the results [39]. Franklin et al [41] conducted 3 evaluative conditioning studies designed to reduce self-injurious thoughts and behaviors, finding that 2 of the studies successfully reduced self-injurious thoughts and behaviors. It was suggested that identifying additional treatment targets, such as other self-injurious thoughts and behaviors not covered by the study, and increasing digital engagement strategies for users may yield better results across all 3 studies in the future [41].

Some of the authors of the studies that showed no significant effects on changes in automatic processes to behaviors outlined possible reasons within their manuscripts. Boendermaker et al [25] attributed these findings to small attentional biases at baseline and low numbers of training trials as compared with other trials in the areas affecting dose-response relationships and having a web-based intervention in a real-world setting, which may have impacted participants' motivation by making them take the training less seriously. Fleming and Burns [42] attributed the null findings to having biased unrepresentative sample populations and the web-based nature of the intervention being available only via a personal computer or laptop limiting those who had an affinity for mobile use. de Voogd et al [20] inferred that the negative findings may have been because of participant dropout, as most adolescents did not complete all 8 intended training sessions. Elfeddali et al [35], Boendermaker et al [25], and de Voogd et al [20] also highlighted that motivation for web-based training appeared low, which they partly attributed to the repetitive nature of the training tasks and the web-based nature of the training tasks, which were completed at home and resulted in a lack of supervision or standardization of training circumstances.

Discussion

Principal Findings

This scoping review aimed to explore the effectiveness of automatic training tasks, types of training tasks commonly used, mode of delivery, and impacts of gamification on automatic processing tasks designed for digital psychological health interventions in real-world settings among adults. A small but developing evidence base was found. Of the 14 studies reviewed, only under half of the interventions resulted in positive changes to automatic processes. The positive trials provide some evidence that this approach may be possible in the real world, although many trials produced mixed results and issues with compliance and engagement were commonly described.

Types of Training Tasks

The review identified 2 main types of tasks commonly used in the field to change automatic processing: attentional bias modification tasks and CBM tasks. These tasks can be delivered via a variety of methods, but the most popular in the health domain are dot-probe, visual probe, and visual search tasks. All 3 tasks have been extensively used and reviewed in the literature [19,47,48]; however, historically, these tasks have been used in a laboratory setting, particularly in mental health interventions seeking to change anxiety and depression. This review revealed an expansion in the field, both in a real-world setting and in other health fields such as problem drinking, smoking, and suicide.

XSL•FO
RenderX

Usability

In the last few years, there has been a shift toward the use of smartphone apps for the delivery of automatic processing interventions. All automatic processing tasks contained within this review were deployed over the web, which made them easily accessible for use in the real world. They were most commonly delivered via a computer or laptop, with over half of the studies using this as the mode of delivery. This aligns with other reviews in the field, such as a recent review of attention and CBM apps by Zhang et al [11], which found 24 CBM apps that were commercially available. Although app usage is increasing, Zhang et al [11] found that most apps (n=8) were not rigorously evaluated, whereas the other 17 were all commercial apps, of which only one was evaluated in published literature. This review has been able to add to the work by Zhang et al [11], who found that 5 smartphone-based CBM task studies have been scientifically evaluated.

Digital advancements are increasingly facilitating pathways into real-world investigations in this field. Although digital platforms do increase participant accessibility in real-world environments, there are limitations to this approach. Three studies noted that the web-based nature of tasks, the lack of supervision, and standardization from external distractors may have negatively impacted the results [20,25,35]. Wiers et al [26] highlighted that this may have been because of large dropout rates commonly seen in web-based experiments and suggested making interventions more engaging to combat this. Boendermaker et al [25] and Elfeddali et al [35] suggested that allowing participants to do the training part of the intervention at home may have affected their motivation levels by them taking the task less seriously. de Voogd et al [20] proposed that the mixed results in that study may have been because of the lack of *stress* imposed by the laboratory environment, where most studies in this area have been conducted traditionally. Indeed, the stress of laboratory environments may, in fact, be beneficial to study outcomes, as participants may have taken the training task less seriously in their home environment, thereby negatively affecting conditioning effects. A review by Santarossa et al [49] in the field of health behavior change also shows that digital interventions are more effective when they have a human support element.

Effectiveness of Changing Automatic Processes

Similar to other reviews in the area [32,33], this review also found mixed findings on the effectiveness of automatic processing tasks in real-world conditions. Of the effective interventions, over half of the studies targeted changing automatic processes toward alcohol. The successful characteristics of these interventions included the use of evaluative conditioning or CBM tasks for intervention delivery, the use of personal computers for mode of delivery, and the use of both elements of the dual process theory.

Mixed results on effectiveness were found for smoking, social anxiety, and self-injurious thoughts and behaviors. This may have been because of participant attrition affecting overall results, and Franklin et al [41] suggested that mixed findings in that study may have been because of issues with a lack of engagement elements. This is a common criticism of bias modification tasks, as they are often reported to be quite boring and repetitive by participants. There has been development in the field to make these tasks more engaging by adding elements of gamification that use visuals, sound effects, point systems, and rewards to make the tasks more engaging [19,24-27,50]. Three of the included studies incorporated elements of gamification to enhance engagement of their interventions [25,41,46]. Boendermaker et al [25] gamified an existing automatic processing task that aimed to train attention away from pictures of alcoholic beverages via a visual probe task. They found that the gamified automatic processing task did not improve participants' motivation to train as compared with the usual nongamified task. In fact, some aspects of motivation appeared to deteriorate rather than improve, suggesting that gamification could have drawbacks if not done optimally.

Crane et al [46] and Franklin et al [41] observed positive effects of study outcomes with the use of games and gamified elements in their interventions. However, the effect that gamification had on influencing outcomes was not investigated. Other reviews have similarly found mixed findings with the use of gamification in automatic and reflective processing interventions [11,19,26,51].

Strengths and Limitations

To our knowledge, this is the first study in the field to review automatic processing studies that specifically focus on real-world settings and adults. The expansive search that included both database and gray literature searching was a strength of this study. This allowed an extensive gathering of evidence to map key concepts and ideas in the field currently.

The eligibility criteria may have limited findings from key studies that did not meet the eligibility requirements, for example, many papers in the field that focus on evaluating automatic processing tasks in real-world settings among children, which may collectively hold key insights into the field at large [24,31,52]. In addition, following the best practice recommendations of conducting scoping reviews from Colquohon et al [37] and Arksey and O'Malley [38], included studies were not assessed for quality in relation to areas of bias such as randomization. Scoping reviews provide a breadth of information rather than the assessment of quality. The disadvantage of this is that it makes it difficult to gain an insight into the robustness and generalizability of the findings. However, the benefit of this method is that it allows the mapping of a wider range of available resources, painting an overall picture of the field at large, as the guidelines for inclusion are not as stringent as a systematic review. The results of a scoping review can, however, sometimes inform the development of a systematic review, which is better placed to deliver an assessment of quality. Finally, only primary studies published in English were included, resulting in a small number of studies for inclusion, which is a common limitation in scoping review [53].

Future Directions

The increasing use of digital platforms to deliver automatic processing tasks, while increasing population reach and accessibility, does have drawbacks. Although monitoring and

XSL•FO
RenderX

standardization levels are relatively achievable in laboratory-based environments, it is often difficult to monitor compliance and ensure adherence in real-world studies. Future studies could experiment with different instructions or persuasion techniques for completing the training as well as different training paradigms to increase compliance. Furthermore, there may be concerns about privacy and confidentiality issues and require further research [54]. Engagement of automatic processing tasks remains a prominent issue because of their inherent boring nature. Gamification offers promising capabilities, and future research should further investigate how its incorporation can enhance enjoyment in the field. Although research into gamification is mixed, studies contained within this review have highlighted gamification as an important engagement strategy [20,26,35], which if implemented correctly could enhance the enjoyment of traditionally mundane tasks. Finally, it was unfortunate that there were few studies in the field that targeted both processes during their interventions; both processes alone have shown significant ability to change behavior, and combining these processes could improve the design and effectiveness of future health interventions and could be a crucial missing link.

Conclusions

This is the first review to synthesize the evidence for published and gray literature on automatic processing tasks set in real-world settings targeting adults. This review has highlighted promising, albeit limited, research demonstrating that automatic processing tasks may be used effectively in a real-world setting to influence behavior change. Given that several trials with negative findings were also identified, future research is needed to understand why significant effects are observed in some contexts and not others and how to optimize delivery for optimal engagement and efficacy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 25 KB - jmir_v22i7e17915_app1.docx]

Multimedia Appendix 2
Data extraction template.
[DOCX File , 26 KB - jmir_v22i7e17915_app2.docx]

Multimedia Appendix 3
Automatic processing targeted health outcomes and effectiveness.
[DOCX File , 36 KB - jmir_v22i7e17915_app3.docx]

References

1. Zhang MW, Ho RC. Harnessing the potential of the Kinect sensor for psychiatric rehabilitation for stroke survivors. Technol Health Care 2016 Mar 4;24(4):599-602. [doi: 10.3233/THC-161147] [Medline: 27061386]
2. Zhang MW, Ho RC. Personalized reminiscence therapy m-health application for patients living with dementia: innovating using open source code repository. Technol Health Care 2017;25(1):153-156. [doi: 10.3233/THC-161253] [Medline: 27689553]
3. Zhang MW, Ho RC, Cassin SE, Hawa R, Sockalingam S. Online and smartphone based cognitive behavioral therapy for bariatric surgery patients: initial pilot study. Technol Health Care 2015;23(6):737-744. [doi: 10.3233/THC-151026] [Medline: 26409514]
4. Zhang MW, Ho RC. Tapping onto the potential of smartphone applications for psycho-education and early intervention in addictions. Front Psychiatry 2016;7:40 [FREE Full text] [doi: 10.3389/fpsyt.2016.00040] [Medline: 27014103]
5. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017 Nov 12;7(11):- [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
6. Corr PJ. Automatic and controlled processes in behavioural control: implications for personality psychology. Eur J Pers 2010 Jul 20;24(5):376-403. [doi: 10.1002/per.779]
7. Carver CS, Johnson SL, Joormann J. Serotonergic function, two-mode models of self-regulation, and vulnerability to depression: what depression has in common with impulsive aggression. Psychol Bull 2008 Nov;134(6):912-943 [FREE Full text] [doi: 10.1037/a0013740] [Medline: 18954161]
8. Schneider W. Controlled & automatic processing: behavior, theory, and biological mechanisms. Cogn Sci 2003 Jun 11;27(3):525-559 [FREE Full text] [doi: 10.1016/S0364-0213(03)00011-9]
9. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004;8(3):220-247. [doi: 10.1207/s15327957pspr0803_1] [Medline: 15454347]

XSL•FO
RenderX

10. Cheval B, Radel R, Neva JL, Boyd LA, Swinnen SP, Sander D, et al. Digital platforms in the assessment and monitoring of patients with bipolar disorderbehavioral and neural evidence of the rewarding value of exercise behaviors: a systematic review. Sports Med 2018 Jun;48(6):1389-1404. [doi: 10.1007/s40279-018-0898-0] [Medline: 29556981]
11. Zhang MW, Ying JB, Song G, Ho RC. A review of gamification approaches in commercial cognitive bias modification gaming applications. Technol Health Care 2018;26(6):933-944. [doi: 10.3233/THC-181313] [Medline: 30040771]
12. Conroy DE, Berry TR. Automatic affective evaluations of physical activity. Exerc Sport Sci Rev 2017 Oct;45(4):230-237. [doi: 10.1249/JES.0000000000000120] [Medline: 28704217]
13. Rebar AL, Schoeppe S, Alley SJ, Short CE, Dimmock JA, Jackson B, et al. Automatic evaluation stimuli - the most frequently used words to describe physical activity and the pleasantness of physical activity. Front Psychol 2016;7:1277 [FREE Full text] [doi: 10.3389/fpsyg.2016.01277] [Medline: 27602014]
14. Beard C. Cognitive bias modification for anxiety: current evidence and future directions. Expert Rev Neurother 2011 Feb;11(2):299-311 [FREE Full text] [doi: 10.1586/ern.10.194] [Medline: 21306216]
15. Hofmann W, Friese M, Wiers RW. Impulsive versus reflective influences on health behavior: a theoretical framework and empirical review. Health Psychol Rev 2008 Sep;2(2):111-137. [doi: 10.1080/17437190802617668]
16. Rhodes RE, de Bruijn GJ. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013 May;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
17. Beard C, Sawyer AT, Hofmann SG. Efficacy of attention bias modification using threat and appetitive stimuli: a meta-analytic review. Behav Ther 2012 Dec;43(4):724-740 [FREE Full text] [doi: 10.1016/j.beth.2012.01.002] [Medline: 23046776]
18. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
19. Boendermaker WJ, Prins PJ, Wiers RW. Cognitive bias modification for adolescents with substance use problems--can serious games help? J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):13-20 [FREE Full text] [doi: 10.1016/j.jbtep.2015.03.008] [Medline: 25843611]
20. de Voogd E, Wiers R, Prins P, de Jong P, Boendermaker W, Zwitser R, et al. Online attentional bias modification training targeting anxiety and depression in unselected adolescents: short- and long-term effects of a randomized controlled trial. Behav Res Ther 2016 Dec;87:11-22 [FREE Full text] [doi: 10.1016/j.brat.2016.08.018] [Medline: 27585484]
21. Enock PM, Hofmann SG, McNally RJ. Attention bias modification training via smartphone to reduce social anxiety: a randomized, controlled multi-session experiment. Cogn Ther Res 2014 Mar 4;38(2):200-216. [doi: 10.1007/s10608-014-9606-z]
22. Blankers M, Salemink E, Wiers R. Cognitive behavioural therapy and cognitive bias modification in internet-based interventions for mood, anxiety and substance use disorders. In: Mucic D, Hilty DM, editors. e-Mental Health. New York, USA: Springer; 2016:193-215.
23. MacLeod C, Mathews A, Tata P. Attentional bias in emotional disorders. J Abnorm Psychol 1986 Feb;95(1):15-20. [doi: 10.1037//0021-843x.95.1.15] [Medline: 3700842]
24. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
25. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 6;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
26. Wiers RW, Houben K, Fadardi JS, van Beek P, Rhemtulla M, Cox WM. Alcohol cognitive bias modification training for problem drinkers over the web. Addict Behav 2015 Jan;40:21-26. [doi: 10.1016/j.addbeh.2014.08.010] [Medline: 25218067]
27. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
28. Allen L, Mulgrew KE, Rune K, Allen A. Attention bias for appearance words can be reduced in women: results from a single-session attention bias modification task. J Behav Ther Exp Psychiatry 2018 Dec;61:97-103. [doi: 10.1016/j.jbtep.2018.06.012] [Medline: 29990684]
29. Förderer S, Unkelbach C. Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS-US relations. Cogn Emot 2012;26(3):534-540. [doi: 10.1080/02699931.2011.588687] [Medline: 21770728]
30. Förderer S, Unkelbach C. Attribute conditioning: changing attribute-assessments through mere pairings. Q J Exp Psychol (Hove) 2015;68(1):144-164. [doi: 10.1080/17470218.2014.939667] [Medline: 24985786]
31. Lau JY. Commentary: a glass half full or half empty? Cognitive bias modification for mental health problems in children and adolescents--reflections on the meta-analysis by Cristea et al (2015). J Child Psychol Psychiatry 2015 Jul;56(7):735-737. [doi: 10.1111/jcpp.12436] [Medline: 26058924]
32. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]

XSL•FO
RenderX

33. Price RB, Wallace M, Kuckertz JM, Amir N, Graur S, Cummings L, et al. Pooled patient-level meta-analysis of children and adults completing a computer-based anxiety intervention targeting attentional bias. Clin Psychol Rev 2016 Dec;50:37-49 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.009] [Medline: 27693664]
34. Clarke PJ, Bedford K, Notebaert L, Bucks RS, Rudaizky D, Milkins BC, et al. Assessing the therapeutic potential of targeted attentional bias modification for insomnia using smartphone delivery. Psychother Psychosom 2016;85(3):187-189. [doi: 10.1159/000442025] [Medline: 27043732]
35. Elfeddali I, de Vries H, Bolman C, Pronk T, Wiers RW. A randomized controlled trial of web-based attentional bias modification to help smokers quit. Health Psychol 2016 Aug;35(8):870-880. [doi: 10.1037/hea0000346] [Medline: 27505210]
36. Dijkers M. What is a Scoping Review? KTDRR. 2015. URL: https://ktdrr.org/products/update/v4n1/dijkers_ktupdate_v4n1_12-15.pdf [accessed 2020-07-07]
37. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
38. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
39. Yang R, Cui L, Li F, Xiao J, Zhang Q, Oei TP. Effects of cognitive bias modification training via smartphones. Front Psychol 2017;8:1370 [FREE Full text] [doi: 10.3389/fpsyg.2017.01370] [Medline: 28855880]
40. Robinson CD, Muench C, Brede E, Endrighi R, Szeto EH, Sells JR, et al. Effect of attentional retraining on cognition, craving, and smoking in African American smokers. Psychol Addict Behav 2017 Aug;31(5):636-646 [FREE Full text] [doi: 10.1037/adb0000286] [Medline: 28627913]
41. Franklin JC, Fox KR, Franklin CR, Kleiman EM, Ribeiro JD, Jaroszewski AC, et al. A brief mobile app reduces nonsuicidal and suicidal self-injury: evidence from three randomized controlled trials. J Consult Clin Psychol 2016 Jun;84(6):544-557. [doi: 10.1037/ccp0000093] [Medline: 27018530]
42. Fleming JB, Burns MN. Online evaluative conditioning did not alter internalized homonegativity or self-esteem in gay men. J Clin Psychol 2017 Sep;73(9):1013-1026. [doi: 10.1002/jclp.22388] [Medline: 27755649]
43. McNulty JK, Olson MA, Jones RE, Acosta LM. Automatic associations between one's partner and one's affect as the proximal mechanism of change in relationship satisfaction: evidence from evaluative conditioning. Psychol Sci 2017 Aug;28(8):1031-1040. [doi: 10.1177/0956797617702014] [Medline: 28562159]
44. Monk RL, Qureshi A, Pennington CR, Hamlin I. Generalised inhibitory impairment to appetitive cues: from alcoholic to non-alcoholic visual stimuli. Drug Alcohol Depend 2017 Nov 1;180:26-32. [doi: 10.1016/j.drugalcdep.2017.07.038] [Medline: 28858670]
45. van Deursen DS, Salemink E, Boendermaker WJ, Pronk T, Hofmann W, Wiers RW. Executive functions and motivation as moderators of the relationship between automatic associations and alcohol use in problem drinkers seeking online help. Alcohol Clin Exp Res 2015 Sep;39(9):1788-1796. [doi: 10.1111/acer.12822] [Medline: 26247799]
46. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
47. Zhang M, Ying J, Song G, Fung DS, Smith H. Attention and cognitive bias modification apps: review of the literature and of commercially available apps. JMIR Mhealth Uhealth 2018 May 24;6(5):e10034 [FREE Full text] [doi: 10.2196/10034] [Medline: 29793899]
48. Jones EB, Sharpe L. Cognitive bias modification: a review of meta-analyses. J Affect Disord 2017 Dec 1;223:175-183. [doi: 10.1016/j.jad.2017.07.034] [Medline: 28759865]
49. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
50. Lumsden J, Skinner A, Woods A, Lawrence N, Munafò M. The effects of gamelike features and test location on cognitive test performance and participant enjoyment. PeerJ 2016;4:e2184 [FREE Full text] [doi: 10.7717/peerj.2184] [Medline: 27441120]
51. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
52. Jacobus J, Taylor CT, Gray KM, Meredith LR, Porter AM, Li I, et al. A multi-site proof-of-concept investigation of computerized approach-avoidance training in adolescent cannabis users. Drug Alcohol Depend 2018 Jun 1;187:195-204 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.03.007] [Medline: 29679914]
53. Lim CR, Barlas J, Ho RC. The effects of temperament on depression according to the schema model: a scoping review. Int J Environ Res Public Health 2018 Jun 11;15(6):- [FREE Full text] [doi: 10.3390/ijerph15061231] [Medline: 29891793]
54. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HL, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health 2019 Jun 18;16(12):- [FREE Full text] [doi: 10.3390/ijerph16122150] [Medline: 31216619]

XSL•FO
RenderX

Abbreviations

CBM: cognitive bias modification
CINAHL: Cumulative Index of Nursing and Allied Health Literature
EMBASE: Excerpta Medica Database
GreyNet: Grey Literature Network Service
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PsycINFO: Psychological Information Database
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by R Ho, K Fox; comments to author 10.03.20; revised version received 24.05.20; accepted 03.06.20; published 29.07.20.

Please cite as:
Jayasinghe H, Short CE, Braunack-Mayer A, Merkin A, Hume C
Evidence Regarding Automatic Processing Computerized Tasks Designed For Health Interventions in Real-World Settings Among Adults: Systematic Scoping Review
J Med Internet Res 2020;22(7):e17915
URL: http://www.jmir.org/2020/7/e17915/
doi:10.2196/17915
PMID:32499213

©Harshani Jayasinghe, Camille E Short, Annette Braunack-Mayer, Ashley Merkin, Clare Hume. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Design, Use, and Effects of Sex Dolls and Sex Robots: Scoping Review

Nicola Döring[1], Prof Dr; M Rohangis Mohseni[1], Dr; Roberto Walter[1], MA

Institute of Media and Communication Science, Ilmenau University of Technology, Ilmenau, Germany

Corresponding Author:
Nicola Döring, Prof Dr
Institute of Media and Communication Science
Ilmenau University of Technology
Ehrenbergstr. 29
Ilmenau, 98693
Germany
Phone: 49 3677694703
Fax: 49 3677694645
Email: nicola.doering@tu-ilmenau.de

Abstract

Background: Although sex toys representing human body parts are widely accepted and normalized, human-like full-body sex dolls and sex robots have elicited highly controversial debates.

Objective: This systematic scoping review of the academic literature on sex dolls and sex robots, the first of its kind, aimed to examine the extent and type of existing academic knowledge and to identify research gaps against this backdrop.

Methods: A comprehensive multidisciplinary, multidatabase search strategy was used. All steps of literature search and selection, data charting, and synthesis followed the leading methodological guideline, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist. A total of 29 (17 peer reviewed) and 98 publications (32 peer reviewed) for sex dolls and sex robots, respectively, from 1993 to 2019 were included.

Results: According to the topics and methodologies, the sex doll and sex robot publications were divided into 5 and 6 groups, respectively. The majority of publications were theoretical papers. Thus far, no observational or experimental research exists that uses actual sex dolls or sex robots as stimulus material.

Conclusions: There is a need to improve the theoretical elaboration and the scope and depth of empirical research examining the sexual uses of human-like full-body material artifacts, particularly concerning not only risks but also opportunities for sexual and social well-being.

(J Med Internet Res 2020;22(7):e18551) doi:10.2196/18551

KEYWORDS

sex toys; sexual objectification; anthropomorphization; embodied sexual fantasies; parasocial interactions and relationships, mobile phone

Introduction

Background

In 2050, it will be perfectly normal for women and men to experience love and sex with robots. This bold prediction from roboticist David Levy [1] started a debate, now more than a decade after, on the ethics, design, use, and effects of human-like, anatomically correct sex robots and of sex dolls, their noninteractive, immobile precursors. Futurologist Ian Pearson [2] went further by predicting that by 2050, women and men will have more sex with robots than with their conspecifics. One may question the validity of these predictions, but there is no doubt that technological change affecting all areas of life will not leave human sexualities unaffected. Significant changes in sexual behavior because of digital media and technologies are already well established [3,4]. Embodied technologies such as sex dolls and sex robots should not be overlooked in this context, especially as the popularization of the sexual uses of human-like material artifacts has long since begun.

So-called *sex toys* representing human body parts (eg, penis-shaped dildos and vibrators) are widely used and

XSL•FO
RenderX

normalized. The lifetime prevalence of vibrator use, for example, is approximately 50% for heterosexual-identified women and men in the United States and Germany [5,6]. Sex toys are also popular in noncisgender and nonheterosexual populations [7]. Through web-based retailers, the sex toy market has expanded and diversified in recent decades, successfully targeting female customers in particular [8]. In the digital age, sex toys are becoming increasingly technologically advanced. Vibrators having integrated cameras are now available that can be remotely controlled by a partner in a long-distance relationship or that can synchronize with the user's digital music playlist or preferred virtual reality porn. The development of innovative sex toys is, at least in part, pushed by customer demand. This is demonstrated by crowdfunding projects in which future customers grant investment money to entrepreneurs who bring new sex toys to the market (eg, the *Ambrosia Vibe*, a so-called *bionic dildo* providing haptic biofeedback while strapped on). There is also growing interest in the development of sex toys for aging populations and for people with disabilities, for instance, sex toys that are mind-controlled and therefore do not require hand function [9,10].

Despite the broad acceptance of sex toys representing human body parts, the development and marketing of human-like full-body sex dolls and of interactive and moving full-body sex robots have elicited great controversy in both public and academic discourses [11,12]. The discrepancies begin with the clarification of the central concepts. Should sex dolls and sex robots simply be regarded as next-level, high-end sex toys? Do they play a different and more important role in the sexual and social lives of their owners and users? Are they treated as surrogates for real partners or even accepted as fully adequate posthuman synthetic partners? This would raise questions not only about their impact on sexual and overall health but also about the future of intimate relationships.

Domestic Use of Sex Dolls and Sex Robots

A typical usage scenario for sex dolls and sex robots is the domestic context in which the artifacts—after purchase—are available for recreational and long-term use at home. Some authors predict strong positive effects of sex dolls and sex robots, including social companionship, sexual exploration, pleasure, and increased satisfaction for individuals and couples [1]. Others, focusing on male users, predict strong negative effects in terms of objectification of and violence against women [13]. They assume that the men using women-like sex dolls or sex robots will be trained to sexually objectify real women and to disregard sexual consent. Furthermore, they predict that women and adolescent girls, already harmed by ubiquitous exposure to unrealistic beauty standards in the media, will feel even more inadequate when exposed to a consumer culture marketing perfectly beautiful, eternally youthful, and completely submissive female-gendered sex dolls and sex robots. Are we looking into an even more gender-unequal future?

Or are we just creating it with one-sided, male-centered, and sex-negative predictions? Why do we not ask different questions [11], such as: What do women want from innovative sex technologies? How could we design and market women- and couple-friendly, feminist, queer, empowering, and inclusive sexual health– and well-being–promoting sex dolls and robots? Most of the claims about current and future effects of sex dolls and sex robots are purely speculative so far because design studies and empirical use and effect studies are scarce.

Commercial Use of Sex Dolls and Sex Robots

The same holds true for the commercial use of sex dolls and sex robots. The first so-called *sex doll brothels* have already opened in Asia, North America, and Europe, accompanied by strong media publicity (HJ Nast, unpublished data, 2019) [14]. In *sex doll brothels*, customers pay an hourly fee to be in a room with a human-like sex doll of their choice. Some authors argue that dolls and robots used as *surrogate prostitutes* are a good thing as they relieve women from prostitution and could reduce sexual violence [15]. However, we have not yet seen data collected from sex workers' perspectives on the issue. Do they want to be relieved of their jobs or are they more afraid of dolls and robots as new competition? Furthermore, anecdotal evidence shows that some customers are now booking both a sex worker and a sex doll. This points to possible commercial use scenarios marked by neither substitution nor competition but collaboration between human sex workers and sex dolls or robots.

Again, the conceptualization is unclear here. What are the practical, legal, and ethical implications of framing short-term commercial use options as a *brothel* or *escort* instead of a *rental* business? Who are the customers—that is, will regular customers switch to dolls, or will we see new *technophilic* customers specifically requesting dolls or robots? Will demand for short-term commercial use of sex dolls and sex robots persist, increase, or wane? Budget restrictions (life-like sex dolls and robots are very expensive), need for discretion (hiding a full-body sex doll or robot from other household members is nearly impossible), and media-induced curiosity (seeing sex dolls and robots represented in pornography and fictional and nonfictional media can be intriguing) might be factors motivating a trip to the *sex doll brothel* today. Will these factors still play a role tomorrow when markets, media representations, and attitudes change?

Commercial short-term use of sex dolls and sex robots is barely understood, but is so highly controversial that some of the first *sex doll brothels* in Europe, North America, and Asia had to close shortly after opening because of community protests and police raids, as reported in numerous news media.

Therapeutic Use of Sex Dolls and Sex Robots

The perspectives of clinicians are also divided. Some therapists, based on first case studies, explain how living with a *love doll* (as doll owners often prefer to call them) can be a helpful and healing transitional process after traumatic experiences, especially when accompanied by professional therapeutic care [16]. Other clinical authors warn their colleagues that products from the *sex robot industry* are marketed with health claims that are *rather specious* [17]. Even more heated are debates about childlike sex dolls produced in Asia and shipped worldwide. Some ethicists and clinicians argue that people with pedophilic preferences could use such dolls or robots as substitutes to prevent them from committing actual child sexual abuse and that therapeutic use might be promising [18]. Other

XSL•FO
RenderX

ethicists and therapists completely reject this idea and warn that childlike sex dolls or robots are very harmful as they normalize and foster child sexual abuse in both pedophilic individuals and the society at large [19]. Legal bans against child sex dolls and robots are not only campaigned for (*Campaign against sex robots*) but, in some countries, also already in preparation or in effect (eg, the *Curbing Realistic Exploitative Electronic Pedophilic Robots Act of 2017—CREEPER Act of 2017* for short—in the United States) [20].

Again, conflicting approaches are visible in clinical, ethical, and legal debates. Should sex dolls and sex robots of all kinds be explored as possible therapeutic tools in the context of different paraphilic disorders and other sexual pathologies? Or should at least some of them be criminalized immediately, with the implication that new forms of doll- and robot-related sexual deviance have been introduced and must be prosecuted?

Objectives, Questions, and Purpose of the Scoping Review

Against this backdrop of highly polarized debates, this scoping review study aimed to examine the extent and type of existing academic knowledge on sex dolls and sex robots and to identify gaps in theory and evidence as well as areas for further inquiry. In accordance with the leading methodological guideline for scoping reviews, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist [21,22], we will proceed to explain, separately and in detail, the review objectives, questions, and purpose.

Review Objectives

To comprehensively map the state of research on sex dolls and robots, it is necessary to cover academic literature from different disciplines and address various dimensions of the issue. Often, debates firstly and primarily focus on the negative or positive *effects* of sex dolls and sex robots. Effects are an important dimension of this issue. However, it is crucial to be aware that effects always depend on the users and *use* in different settings (eg, domestic, commercial, and therapeutic) as well as on the selected sex doll's or sex robot's *design* (eg, gender, age, race, body type, and sexual and nonsexual functionalities). Hence, this review, as indicated by its title, addresses the design, use, and effects of both sex dolls and sex robots.

Definition of Sex Dolls

Sex dolls are defined as human-like, full-body, anatomically correct anthropomorphic dolls of different materials (eg, rubber, plush, silicone, and thermoplastic elastomer) and price ranges that are designed for sexual use [11]. Sex dolls have at least one penetrable orifice (mouth, vagina, or anus) and/or one body part that can be inserted by the user (tongue or penis). The doll parts for sexual penetration or insertion are usually designed to be removable for cleaning. The special thing about sex dolls is their sexual functions, but this does not mean that they are used exclusively for sexual purposes. They can also serve as artificial love partners, social companions, or photo models, which is why their owners often call them *love dolls* or simply *dolls*. The term is also used by most scientists and parts of the media.

Sex dolls come in different genders (female, male, or trans), races (eg, white, Asian, or African), ages (adult, adolescent, or child), body types, skin, hair, and eye colors. High-end sex doll manufacturers (eg, RealDoll Abyss Creations, Sinthetics, and Orient Industry) offer ample options for selection and customization and also produce custom-made sex dolls. Therefore, *abstract sex dolls* with no resemblance to a specific real person need to be differentiated from *portrait sex dolls* designed in the likeness of a real person (eg, porn star, celebrity, or ex-partner). True-to-life sex dolls like RealDolls from Abyss Creations are delicate, need care and repair, and are not easy to handle because of their weight of approximately 65-70 lbs for female dolls and 85-105 lbs for male dolls.

Thus far, the sex doll market—determined by customer demand—offers mainly female sex dolls with highly sexualized looks that meet traditional feminine beauty standards (young, slim, pretty face, long hair, and large breasts). However, customization already allows for more body diversity (eg, androgynous or gender-queer looks), including the deliberate design of so-called bodily flaws (eg, moles, scars, stretch marks, belly fat, or body hair). Hence, the sex doll industry caters to different appearance-related customer demands (eg, the illusion of perfect supernatural beauty, resemblance to a real person, specific body-related preferences, or fetishes).

Definition of Sex Robots

Sex robots (sexbots) are defined as human-like, full-body, anatomically correct humanoid service robots of different materials, technologies, and price ranges that are designed for sexual use [11]. Sex robots look like sex dolls but are equipped with sensors, actuators, and artificial intelligence (AI). Sometimes, they are called AI sex dolls or robotic sex dolls to characterize them as upgrades of their noninteractive, immobile precursors. For gender sex robots, the binary terms fembot and malebot or gynoid and android are used. Sex robots come with all the attributes and functionalities of sex dolls and, in addition, can display conversation skills, emotions, and preprogrammed personalities. Furthermore, they can perform partially autonomous behaviors such as sexual movement (eg, hand movement for masturbation) or simulation of orgasm. However, the range of behaviors of existing sex robots is very limited. It can be assumed that the handling and maintenance of sex robots as large, heavy, and technically advanced products is demanding. Like sex dolls, sex robots are defined by their sexual functions but are also suitable for other functions in addition to sex (eg, social companionship).

Sex robots marketed today should not be confused with *concepts of future advanced sex robots* that are envisioned as having sentience, consciousness, free will, morality, and possibly even the legal status of citizens. There are also visions of future multifunctional assistance robots for domestic use that will do housework and errands, look after children, provide elderly care services, and offer sexual services. These imagined advanced sex robots or multifunctional robots with sexual functions appear in science fiction (eg, the Swedish television series *Real Humans* or the US movie *Ex Machina*) and in recent philosophical and legal sex robot debates [23,24], but are far away from the current state of technological development.

XSL•FO
RenderX

The Relevance of Sex Dolls and Sex Robots

Although high-end, true-to-life sex dolls have been on the market for more than 20 years (the leading US manufacturer Abyss Creations, creator of RealDoll, was founded in 1997), sex robots are still in a very early stage of development. The manufacturer TrueCompanion claims to have brought the world's very first sex robots to the market. It presented its female-gendered sex robot *Roxxxy* to the public in 2010 and later announced the male-gendered sex robot *Rocky*, stirring a media frenzy [25]. However, it is assumed today, for good reasons, that *Roxxxy* and *Rocky* have never been more than overhyped prototypes [14,26]. Thus far, not a single customer has surfaced, and the TrueCompanion webshop has not changed over the years. The established RealDoll manufacturer Abyss Creations launched its first sex robot *Harmony* in 2018, followed by *Solana* and *Henry*. Sex robot *Samantha* by the Spanish manufacturer Synthea Amatus and sex robot *Emma* by the British-Chinese manufacturer AI Tech UK have likewise been sold since 2018. All these sex robots are sex dolls enhanced with some very limited AI and interactive features. Hence, although supposedly thousands of experienced sex doll owners exist worldwide, who have built their own distinct doll owner communities with online forums and offline meet-ups, there is, by comparison, only a very small number of pioneer users of sex robots. This limits the options for empirical research on long-term sex robot users, use, and effects.

However, as we are transitioning into the age of the robot, and sex robots provide interactivity, AI, and partly autonomous behavior, sex robots have been attracting much more public and scholarly attention than sex dolls. After all, they have been an integral part of science fiction literature for decades [27]. Considering the history of and relation between sex dolls and sex robots, it seems reasonable to address them collectively in this research review concerned with the sexual uses of human-like full-body material artifacts.

Review Questions

In mapping the current state of academic knowledge on sex dolls and sex robots, the scoping review aimed to answer the following 4 review questions (RQ):

> *RQ1: What is the state of sex doll and sex robot research in terms of the overall amount and type of research?*
>
> *RQ2: What is (not) known about the design of sex dolls and sex robots?*
>
> *RQ3: What is (not) known about the users and uses of sex dolls and sex robots?*
>
> *RQ4: What is (not) known about the effects of sex doll and sex robot use?*

These 4 RQs will be addressed separately for sex dolls and sex robots based on the respective literature searches.

Review Purpose

By systematically mapping the current state of academic knowledge on sex dolls and sex robots, this scoping review aimed to advance the understanding of sex researchers and practitioners and foster their professional involvement in the field of sexual uses of human-like material artifacts. Technicization and digitalization are fundamental societal processes that affect all areas of life, including human sexualities. Sex researchers and practitioners must be prepared to deal with these transformations in an informed and professional way, reflecting their own knowledge gaps, prejudices, and projections. Sex dolls and sex robots seem to be a particularly fruitful field of inquiry and professional development, as they often elicit very strong emotions that need to be recognized, worked through, and questioned with the help of clear conceptualizations, sound theories, and solid empirical evidence.

Methods

A scoping review is "a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting and synthesizing existing knowledge" [28]. As the body of academic literature on sex dolls and sex robots has not yet been comprehensively reviewed and exhibits a broad and heterogeneous nature that is not amenable to a more precise systematic review, a scoping review is of particular use [29]. Our procedure follows current methodological guidelines for conducting systematic scoping reviews [29], particularly the PRISMA-ScR checklist [21,22].

Literature Search

To search for relevant academic publications on sex dolls and sex robots, the following 5 scientific literature databases covering different disciplines were used to ensure a multidisciplinary, multidatabase search strategy:

1. *Scopus* (largest academic literature database, approximately 57 million references, covering different disciplines, 1960-current),
2. *Medical Literature Analysis and Retrieval System Online (MEDLINE*; approximately 28 million references, focus on medicine, 1950-current),
3. *PsycINFO* (approximately 4 million references, focus on psychology, 1806-current),
4. *Institute of Electrical and Electronics Engineers (IEEE) Xplore* (approximately 4.5 million references, focus on technology, 1872-current), and
5. *Association for Computing Machinery (ACM) Digital Library*—Guide to Computing Literature (approximately 3 million references, focus on computing, 1950-current).

For sex dolls, the search terms "sex doll," "sex dolls," "doll sex," "love doll," "love dolls," and "doll love" were used. For sex robots, the search terms "sex robot," "sex robots," "sexbot," "sex bot," "robot sex," "love robot," "love robots," "lovebot," "love bot," and "robot love" were used. Search terms were applied to publication titles, abstracts, and keywords. Searches were limited to the English language, without publication date, publication type, or study type restrictions.

The search strategy was validated through the retrieval of a key set of relevant publications in Scopus, where 24 citations for sex dolls and 73 citations for sex robots were identified. The

XSL•FO
RenderX

Scopus search strategy was then translated to the other 4 databases and executed between August 6, 2019, and August 9, 2019 (Multimedia Appendix 1 shows the full documentation of the electronic search strategy). Bibliographic information for all search results was exported from the databases into the citation management software Citavi 5.7.1 (Swiss Academic Software GmbH).

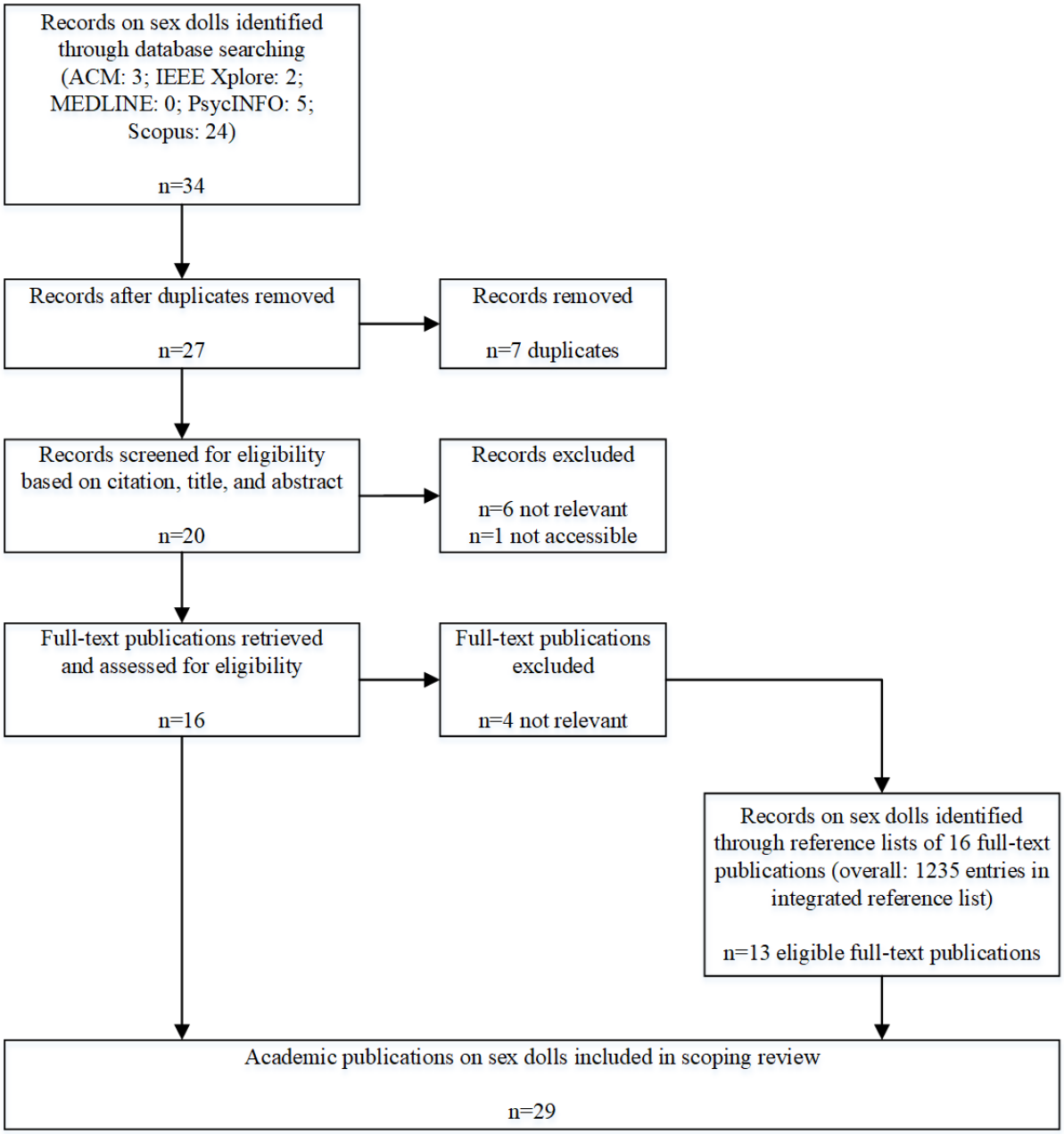
Literature Selection

The literature selection included 3 steps: (1) the removal of duplicates among identified records; (2) the scanning of citations, titles, and abstracts for eligibility; and (3) the retrieval of full texts and assessment of eligibility. As we are reviewing an innovative emerging research field, we included all study and publication types from all available publication years. The only 2 exclusion criteria applied were lack of topical relevance (irrelevant were all publications that did not provide substantial knowledge about sex dolls or sex robots, that is, publications that only mentioned but did not investigate the topic or only referred to relevant publications) and nonaccessibility of published full text. We used the reference lists of all eligible full texts found through the databases to systematically search for further publications. Screening, assessing, and inclusion were performed in duplicate.

As can be seen in Figure 1, for sex dolls, we identified 16 eligible publications through the databases and 13 additional publications through their reference lists, resulting in 29 included sex doll publications (Figure 1).

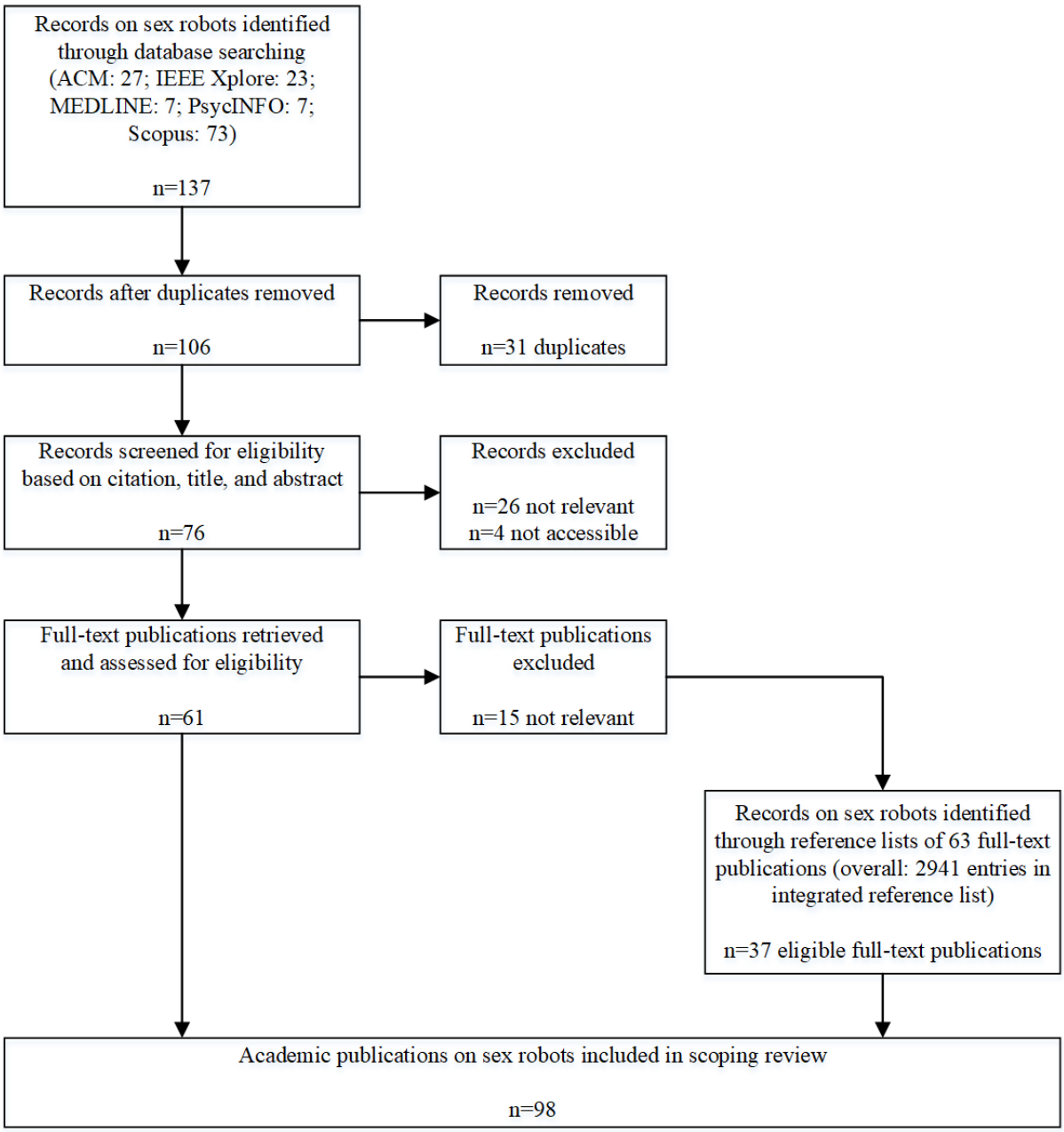
The same procedure was used for sex robot publications, resulting in 98 included publications (Figure 2).

Figure 1. Flowchart of the scoping review procedure for literature identification and selection of sex dolls. ACM: Association for Computing Machinery. IEEE: Institute of Electrical and Electronics Engineers; MEDLINE: Medical Literature Analysis and Retrieval System Online.



XSL•FO
RenderX

Figure 2. Flowchart of the scoping review procedure for literature identification and selection of sex robots. ACM: Association for Computing Machinery; IEEE: Institute of Electrical and Electronics Engineers; MEDLINE: Medical Literature Analysis and Retrieval System Online.



Data Charting

During the data charting phase, all included publications were reviewed and charted using a data charting form that was pilot tested for all 29 included sex doll publications, discussed within the team, and revised 3 times. The final charting form has a sex doll and a sex robot version, each including 10 variables: (1) citation (author and year), (2) citation count (derived from Google Scholar), (3) publication type, (4) peer review, (5) academic discipline (derived from the first author's academic position and/or education), (6) study type (derived from topic and methodology and used for grouping of sex doll/sex robot publications), (7) key findings regarding the study's research question, (8) key findings—if applicable—regarding sex doll/robot design, (9) key findings—if applicable—regarding sex doll/robot use, and (10) key findings—if applicable—regarding sex doll/robot effects. Variables 1 to 7 were used to answer RQ1 in the overall state of the research, variable 8 addressed RQ2, variable 9 addressed RQ3, and variable 10 addressed RQ4. The data were charted in duplicate.

Synthesis and Reporting of Results

First, a numeric overview of the number, type, and distribution of included publications was created using 2 summary tables and a chart of the timeline of publication activity. Second, a narrative synthesis of the results of the previous studies was created, focusing particularly on their insights regarding design, use, and effects of sex dolls and sex robots. Third, to fully answer the RQs on "what is (not) known" about sex dolls and sex robots, the state of research was critically assessed for research gaps, and recommendations for future research were included. To avoid vagueness and to achieve maximum

XSL•FO
RenderX

usefulness, recommendations for future research with specific references to applicable theories, relevant methods, and related research fields were backed up. All steps of data synthesis and reporting were discussed within the team and performed in duplicate.

Results

State of Research on Sex Dolls

To summarize the state of research on sex dolls, we first map the number and type of publications and then report their main results regarding the design, use, and effects of sex dolls before coming to the research gaps and recommendations on how to fill them.

Amount and Type of Research on Sex Dolls

During the scoping review literature identification process, we included 29 academic publications on sex dolls (Figure 1). This body of literature consists of 5 distinct groups of studies according to both their topics and methodologies, which are closely linked (Table 1).

The body of academic literature contains 2 published monographs that exclusively focus on sex dolls [33,43]. Approximately 50% (17/29) of the included sex doll publications were peer reviewed. The Google Scholar citation count reveals a range from 0 to 46 citations; the most cited publication was the monograph *The Sex Doll: A History* by Anthony Ferguson [33]. It is noteworthy that all of the most cited publications within each of the 5 groups of sex doll publications were not peer reviewed. Regarding the timeline, the oldest sex doll publication identified in the databases and included in our review is a 1993 clinical case study from medicine [50] that deals with the shared use of an inflatable sex doll. However, this is an outlier, with >85% (25/29) of the sex doll publications having been published within the past 10 years (2010-2019; Figure 3).

XSL•FO
RenderX

Table 1. Amount and type of research on sex dolls (N=29 included academic publications, based on literature search in August 2019).

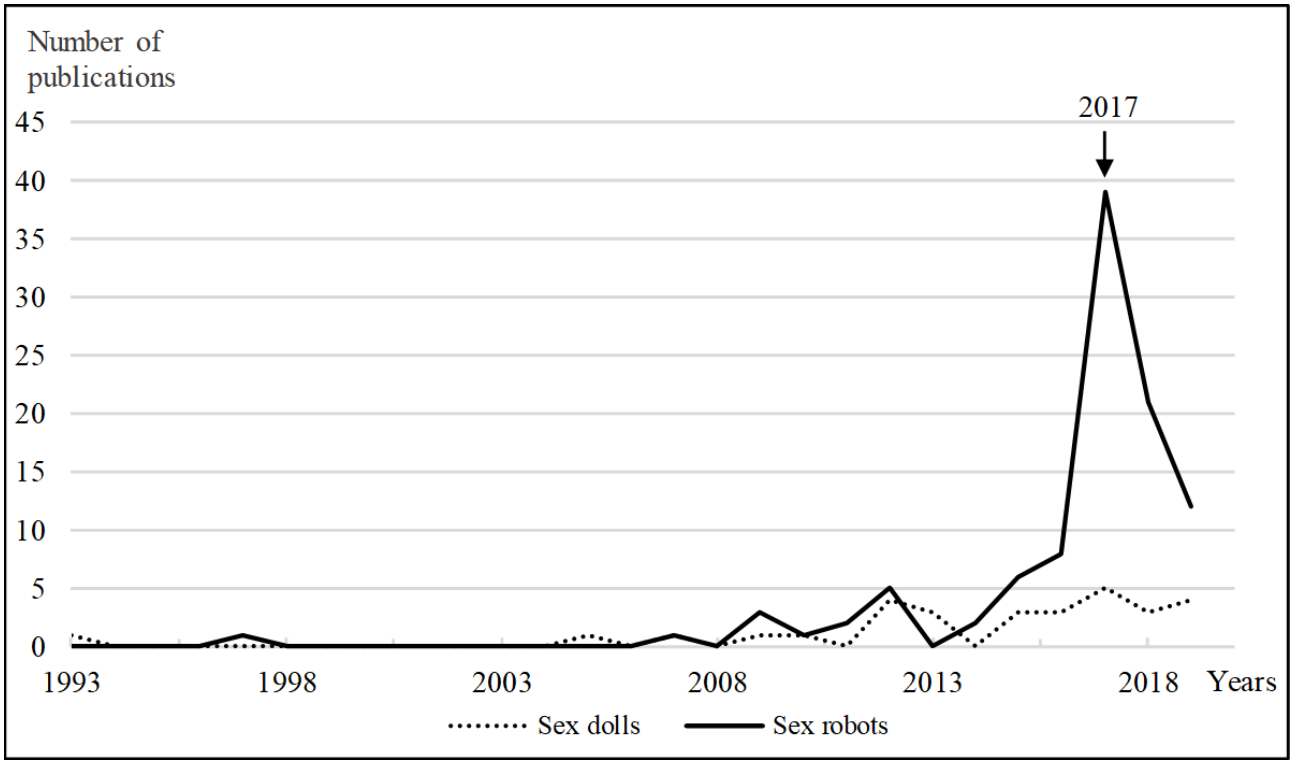
Reference	Citation count[a]	Peer review	Academic discipline
Sex doll conceptualization and theory (n=11)			
Blizzard (2015) [30]	__[b]		Science and technology studies
Blizzard (2018) [31]	0		Science and technology studies
Cassidy (2016) [32]	5	✓	English
Döring and Pöschl (2018) [11]	6	✓	Psychology
Ferguson (2010) [33]	46		Unknown
Kim (2012) [34]	5	✓	Women's and gender studies
Levy (2012) [14]	27		Artificial intelligence
Nast (2017) [35]	7	✓	International studies
Nast (2019) (HJ Nast, unpublished data, 2019)	—	✓	International studies
Ray (2016) [36]	3		English
Wong (2015) [37]	0		Sociology
Sex doll representations in art and media (n=7)			
Burr-Miller and Aoki (2013) [38]	7	✓	Communication and media studies
Connor (2015) [39]	2	✓	English
Getsy (2013) [40]	6	✓	Art history
Koné (2016) [41]	0		German
Roos (2005) [42]	9	✓	English
Smith (2013) [43]	21		Visual arts
Weisel-Barth (2009) [44]	1	✓	Psychoanalysis
Empirical studies on sex doll use and effects (n=5)			
Ciambrone et al (2017) [45]	5	✓	Sociology
Knox et al. (2017) [46]	3	✓	Sociology
Langcaster-James and Bentley (2018) [47]	1	✓	Anthropology
Su et al (2019) [48]	0	✓	Human-computer interaction
Valverde (2012) [49]	15		Psychology
Clinical case studies on sex doll use and effects (n=3)			
Kleist and Moi (1993) [50]	18		Medicine
Knafo (2015) [16]	3	✓	Clinical psychology/psychoanalysis
Knafo and Lo Bosco (2017) [51]	12		Clinical psychology/ psychoanalysis
Legal regulation of child sex dolls (n=3)			
Brown and Shelling (2019) [19]	0	✓	Criminology
Chatterjee (2019) [52]	0	✓	Criminology and law
Maras and Shapiro (2017) [20]	7		Criminology and law

[a]Citation count according to Google Scholar in August 2019.

[b]Google Scholar did not list the reference.

XSL•FO
RenderX

Figure 3. Timeline of publications for sex dolls (N=29) and sex robots (N=98).



Research Findings on Sex Dolls

We have summarized the main findings of previous sex doll research separately for the 5 groups of sex doll publications (Table 1).

Sex Doll Conceptualization and Theory

The largest group of sex doll publications (11/29, 38%) consists of theoretical studies that aim to conceptualize the human–sex doll relationship. Half of them (6/11) promoted a critical feminist conceptualization of the female sex doll as an expression and affirmation of patriarchal gender power relations and women's sexual objectification by men (ie, by male sex doll producers, owners, users, and observers). The publications deal with the production and use of female sex dolls in Western [30,32,33,36] and Asian (HJ Nast, unpublished data, 2019) [35] countries and sometimes discuss gender issues of female sex doll use in relation to economic, cultural, and racial issues as well as in relation to recent crises of masculinity. Their overall assessment of female sex dolls and their effects is very negative. A typical example of the critical feminist conceptualization of female sex dolls is as follows [33]:

> *The female sex doll is man's ultimate sexually idealized woman. It is never more than the sum of its fully functional parts. A woman rendered harmless, it is immobile, compliant, and perhaps most importantly, silent. What the user of the sex doll seeks is the negation of change and the comfort of always retaining control of the relationship.*

The other half of the theoretical papers (5/11) conceptualize human–sex doll relations, mainly in a positive way [11,14,31,34,37]. These papers do not limit their focus to female dolls or (supposedly heterosexual, sexist, and misogynist) male doll users only, but they address the already observable and potentially growing diversity of both dolls and doll users (eg, including women, queers, older people, and people with disabilities). Furthermore, they reject the 2 key assumptions of the critical feminist conceptualizations that dolls are inanimate objects for mere (and questionable forms of) male sexual gratification (eg, acting out sexual fantasies of subjugation and violence against women) and that they are surrogates for real women. Instead, dolls are conceptualized as *new types of social actors*, neither inanimate objects nor surrogate humans but as *posthuman partners* or as *interanimated beings* [31,34,37].

What dolls *are* and what human-doll relationships *mean* is, therefore, not predefined by attributes of the doll, but is the result of the *connections* between the human beings and the *doll beings*. It is up to the users if they *abuse* or *take care of* their dolls if they act out hatred or love. The anthropomorphic, anatomically correct full-body doll in this context might appear passive. The papers, however, argue that in its passivity lies agency and even power [34]: The *doll being*, although vulnerable to abuse just as the human being, is easily able to elicit attention, care, love, and long-term relationships. The conceptualization of dolls as interanimated beings covers rather than denies situations of doll objectification and abuse. However, it also covers situations of doll appreciation, care, and love. Most importantly, such a conceptualization covers complex situations of mixed and ambivalent connections between dolls and their users.

Sex Doll Representations in Art and Media

The second largest group of sex doll publications (7 out of 29; Table 1) analyzes sex doll representations in art and media. Several studies explain that men creating idealized and sexualized female statues, mannequins, or dolls is a *common*

XSL•FO
RenderX

trope in the history of art and culture that can be understood as an expression of patriarchal gender relations, objectification, and fetishization of women [39,42,43]. One notorious example is Ovid's poem about the sculpturer and ancient Greek mythical king of Cyprus, *Pygmalion*. Mythical Pygmalion, appalled by female sexual permissiveness, turned away from real women and created an ivory sculpture of his ideal woman. He physically loved the sculpture, and it later came to life.

A very famous example from modern cultural history is the Austrian artist *Oskar Kokoschka*, who in 1919 commissioned an anatomically correct sex doll in the likeness of his former lover Alma Mahler after she had ended both the relationship with him and her pregnancy. The *Alma Mahler doll* is an example of a portrait sex doll produced without the consent of the person portrayed. Kokoschka lived with the Alma Mahler doll, hired a maid for her, brought her to public spaces like the opera, and created numerous drawings and paintings of her before he destroyed the doll [42,53]. Kokoschka's strange and scandalous actions were often dismissed as a private matter of grief, trauma, or insanity. However, they can also be read as an expression of male entitlement and an attempt to exercise revenge by publicly shaming Alma Mahler. Last but not least, according to the literature, there is also good reason to consider this case as some sort of performance art [42]. Within the sex doll literature, the Alma Mahler portrait sex doll is addressed the most in papers interested in sex doll representations in art and media (4 out of 7) [39,41-43] but is also mentioned in theoretical [33,37], empirical [49], and case study [16,51] sex doll publications.

Although the feminist critique of sex dolls plays a role, most publications in this group provide more complex interpretations. They point to the fact that in creating sexualized female dolls, male artists deal with more than gender relations, also dealing, for example, with their own fear of death [39] or with their own object status [42]. Furthermore, some authors in their cultural analyses point to sex doll–related artwork that reveals additional and potentially emancipatory dimensions of sex doll use. An artist in the United States, Amber Hawk Swanson, who identifies as a lesbian commissioned a RealDoll sex doll in her own likeness from Abyss Creations, married her, and lived and collaborated with her in video and performance artwork. Amber Swanson's *Amber Doll Project* (2006-2011) triggered and disrupted the audience's clichéd (heterosexual) fantasies about lesbian desire, twin sexuality, and the role of females as sexual objects [40].

Another cultural analysis stresses the 2 main functions of dolls: they are made to be looked at and to be played with [41]. Although a feminist critique often assumes a rigid misogynist meaning and use of female sex dolls, art projects demonstrate more complex, creative games to be played with dolls. An artist in the United States Laurie Simmons brought back a female sex doll from Japan and created a series of photographs of her. *The Love Doll* (2009-2011) project goes beyond affirmation and deconstruction of sexual objectification as the female artist casts a loving, *maternal gaze* on her doll, thus inventing "a novel game to play with the doll" [41].

The last 2 publications deal with the representation of sex dolls in movies and television. The first one, the US movie *Lars and the Real Girl* (2007), is interpreted from a psychoanalytic perspective as an "inspiring tale of healing" [44]. The movie tells the story of withdrawn single 28-year old *Lars* who starts living with sex doll *Bianca* to end his loneliness. His family and the whole town play along by treating the doll as his legitimate partner and welcoming her as a new community member. This magically transforms everyone for the better. Ultimately, Lars can let go of the doll and turns toward a real woman. The movie deals with the contested topic of men's relations with sex dolls in a very empathetic and romantic way. Interestingly, fictional Lars never has sex with his doll because *Bianca* is very religious, and thus, premarital sex is unthinkable. Tellingly, the acceptance of the *Bianca* doll by both the fictional community in which Lars lived and by the mainstream cinema audience required her to be a sexually abstinent sex doll [44].

The British Broadcasting Corporation (BBC) documentary *Guys and Dolls* (2007) portrays 4 men (*Davecat*, *Everard*, *Gordon*, and *Mike*) living with their female sex dolls. As media analysis reveals [38], the documentary explains the doll owners' unusual lifestyle as a result of their *heteronormative shortcomings*. The heterosexual men were not able to create relationships with real women and hence settled with dolls. However, a queer reading of the documentary is also possible, as the lifestyle of a doll lover allows men to express their sexualities differently. In the context of doll care, a variety of feminine connotated sensual activities are legitimized and carried out (eg, washing, drying, powdering, dressing, and putting make-up on the doll). Thus, the documentary unintentionally illustrates that the doll owner identity can also be read as a *queer sexual identity* [38].

Empirical Studies on Sex Doll Use and Effects

The third group of sex doll publications (5 out of 29; Table 1) contains empirical studies on sex doll use based on potential future or on current doll users' subjective accounts. We could not find empirical papers dealing with the prevalence of sex doll use. As reported in one of the above-cited theoretical papers [11], in a national web-based survey conducted in 2016 in Germany (N=2000; 50% female, 18-69 years), the lifetime prevalence of sex doll use was 9% for men and 2% for women. A web-based survey of 345 (81% female) undergraduate students in sexuality courses at a university in the United States revealed that 8% of the respondents would use a sex doll and 17% could understand a sex doll user [46]. The authors interpret the result as an indicator of the widespread stigmatization of sex doll use as opposed to the widespread acceptance of sex toy use.

To survey doll owners on their first-hand, long-term experiences with sex dolls, some researchers successfully turned to sex doll owner online community forums for recruitment. A psychological survey with 52 doll owners (6 female) of an English language international online doll owner forum showed that respondents used their dolls for solo and partnered sexual activities and evaluated the sexual experiences with their dolls as enjoyable [49]. Contrary to common belief, the surveyed doll owners (mean age, 43 years) did not show below-average mental health or life satisfaction on standardized scales; however, they

XSL•FO
RenderX

reported possibly above-average problems with sexual functioning. Human-doll relationships are not always monogamous. A considerable number of surveyed male doll owners were in a relationship with a human partner (21%) and/or had more than one doll (39%). The author calls for more research on the psychologically adaptive and maladaptive uses of sex dolls.

An anthropological survey with 83 members (3 female, 2 gender fluid, 2 trans, and 1 other gender) of 2 English language international online sex doll forums revealed that most respondents characterize the relationship with their doll as a *sexual relationship* (50/83, 77%). At the same time, many respondents also describe their relationship with the doll as "companionship" (47/83, 57%) and as a "loving relationship" (39/83, 47%). The researchers conclude that so-called sex dolls not only provide sexual gratification but also serve as multifunctional dolls (they suggest the label *allodolls*) that can provide *posthuman kinship* and alleviation of loneliness.

Instead of using survey methodology, 2 other studies chose a nonreactive approach and collected sex doll owners' publicly available web-based content. One study qualitatively analyzed 68 customer testimonials (4 written by females) published between 2006 and 2016 on the website of the RealDoll manufacturer Abyss Creations [45]. The researchers found that dolls foster the commodification of female bodies because (1) the manufacturer offers many options for customization that are in line with stereotypical beauty standards, and consequently, (2) the users write a lot and in great detail about their preferences regarding the bodily appearance of their female dolls. Apart from the physical beauty of the doll, emotional closeness to the doll also plays an important role in doll owners' testimonials. They write extensively about the comforting effect of the doll's mere presence [45]:

> *My doll arrived four days ago and my home has a new, warm feeling to it.*

They also stress how much they enjoy taking care of the doll (doll maintenance includes regular washing, drying, powdering, and dressing) and, thus, feeling needed [45]:

> *She's coming to life for me, I want to take care for [sic] her all the time. Yes, my life has become much fuller.*"

The authors' assessment of men's attachments to sex dolls is ambivalent—concern about objectification of women's bodies is mixed with acknowledgment of the creation of supportive emotional intimacy.

The same main result is reported by a qualitative content analysis of 316 discussion threads with 7775 posts from the Abyss Creations RealDoll online forum [48]. Sex doll owners create *embodied intimate fictions* with the dolls. They often praise their idealized bodily attributes and supernatural feminine beauty. However, they do not reduce the dolls to mere sex toys but create rich narratives (in both text and photographs) about their dolls' personalities, backstories, and experiences, integrating domestic life, outdoor trips, and sexual encounters. Furthermore, the lively discussions in the online forum illustrate that doll owners not only bond with their dolls but also with other members of the doll owner community. As hobbyists, not unlike pet owners, they share tips and tricks around doll purchase, doll care, and doll photography.

Clinical Case Studies on Sex Doll Use and Effects

The fourth group of sex doll publications involves 3 clinical case studies, 2 from psychology, and 1 from medicine (3 out of 29; Table 1). The clinical-psychological case study of psychoanalyst Danielle Knafo [16] deals with a *48-year-old psychotherapy patient Jack*, an actuary by profession. He had suffered a problematic childhood with a derogatory mother, and his 2 marriages had failed. Deeply hurt by the most important women in his life and inspired by an online forum, he had bought RealDoll *Maya* for over US $10,000. Reluctantly, he shares with his psychoanalyst that *Maya* has now been his *girlfriend* for 2 years and that she is "beautiful" and "super in bed". He adds how much he enjoys her company, how harmonious their relationship is ("we never fight"), and that he thinks he might be in love with her. However, he is also conflicted about his unusual lifestyle and therefore seeks therapeutic help. The feminist identified psychoanalyst reports how she was initially repulsed at the idea of a man choosing a sexist object as his girlfriend [16]. However, her "own perversity kicked in" along with a kind of "voyeuristic curiosity," as she describes it [16]. She manages not to judge Jack but to understand him. She concludes that RealDoll *Maya* is more than a "perversion" and that she is an "invention" and a "lifesaver for Jack" [16]. During psychoanalysis with an accepting female therapist, Jack gains enough self-confidence and optimism to retire *Maya* and return to relationships with real women. In this case study, the sex/love doll served as a soothing and healing *transitional object* in the sense of Donald Winnicott's [54] theory of transitional objects.

The second case study from the sex doll literature comes again from Danielle Knafo [51] and is based on 7 hours of personal interviews that she led with *Davecat, a 42-year-old African American self-proclaimed doll lover* in his Michigan home. Davecat has lived with RealDoll *Sidore Kuroneko* (nickname *Shi-chan*) since 1998 and regards her as his wife. They wear matching wedding rings inscribed with the words "Synthetic love lasts forever" [51]. In 2012, Davecat ordered a second doll, this time from the Russian manufacturer *Anatomical Dolls* and named her *Elena Vostrikova* (nickname *Lenka*). *Elena* has the status of a "mistress, plaything and companion" for both Davecat and his bisexual synthetic wife *Sidore*. *Elena* is built lighter with looser joints. "Elena is more built for sex whereas Sidore is built for love" as Davecat puts it [51]. *Muriel Noonan* (nickname *Mew-Mew*), his third doll, is made of wood, leather, Lycra skin, and cotton batting. She is least used for sex and mostly serves as a *flatmate*. Davecat has given all of his dolls complex backstories and personalities and lives with them in what he describes as a harmonious polyamorous *family* to which, at the time of the interview, he plans to add 2 more dolls in the future. Davecat explains how he experiences sex with a doll (for him a "synthetik [sic] partner") in comparison to sex with a human (for him an "organik [sic] partner") [51]:

> *Dolls overall are simultaneously robustly made and fragile. They're ostensibly made for sex, but they're*

XSL•FO
RenderX

> *also sculpture pieces. With an organik partner, obviously you can be a bit rougher, but I take care to be gentle with Shi-can and Lenka when we're in bed. Another remarkable difference between organik and synthetik women is that when you're inside a doll's vagina or anus, there's a vacuum effect that's pretty… breathtaking. I'd say sex flows a little better with an organik, as she's able to move herself, whereas changing positions with a doll requires you to pause and rearrange everything. Overall, though, personally, I'd rate sex with a synthetik woman to be as good, if not better, than with an organik woman. Mainly as a doll's artificiality is a huge draw with me....*

The psychoanalyst characterizes Davecat as a man who has been struggling all his life with intimacy issues and has found dolls as a viable alternative to having a human partner. At the same time, she acknowledges that Davecat feels sexually and emotionally attracted to the artificial aspects of dolls [51]. He self-identifies as an *iDollator*, a doll lover who prefers dolls to humans. This identity is so meaningful to Davecat that he serves as an activist and spokesperson for the doll lover community. He has participated in numerous press interviews, photoshoots, and television documentaries (eg, the earlier mentioned BBC documentary *Guys and Dolls*). The psychoanalyst, dissecting Davecat's biography, neither stigmatizes nor pathologizes him. Although she assumes that his unconventional lifestyle is rooted in anxieties and conflicts [51], in her evaluation, it appears to be a viable solution. After all, Davecat is not harming anybody, is able to work, well-integrated socially, and satisfied with his life.

The third and last case study is a medical one. It proves that the shared use of an inflatable sex doll can lead to the transmission of a sexually transmitted disease (ie, gonorrhea) if the doll is not cleaned or no protection is used [50]. In this case, a male sailor had found the sex doll of a colleague on board by chance and used it secretly.

Legal Regulation of Child Sex Dolls

The fifth and final group of sex doll publications covers 3 publications on child sex dolls (3 out of 29; Table 1). All 3 call unanimously for a legal ban and explore the implementation of such a ban in different legal systems, namely, in Australia [19], the United Kingdom [52], and the United States [20]. They reject the idea of possible therapeutic value and stress that the production, marketing, and use of child sex dolls would normalize and foster child sexual abuse. The publications point to different harmful uses of child sex dolls (eg, the use of child sex dolls during grooming or during the abuse of children or the exploitation of individual children by producing portrait sex dolls in their likeness). The most important reason given for banning child sex dolls is the assumption that acting out child sexual abuse with a doll would rehearse, train, and trigger real child sexual abuse. Abstract child sex dolls are compared with computer-generated or so-called *fantasy child pornography* [52]. In both cases, no children are directly harmed in the process of production, but the dissemination and marketing of two-dimensional or three-dimensional depictions of sexualized children is still considered harmful and exploitative and should therefore be criminalized according to all three studies.

Research Gaps in Sex Dolls

There is a considerable discrepancy between the great media interest in the topic of sex dolls and sex robots mentioned in the introduction and the limited amount of scientific knowledge. Overall, the interdisciplinary field of sex doll research is fairly small (RQ1). *Empirical and clinical* studies on doll use, in particular, are scarce (5 peer-reviewed papers in total) and often have limited generalizability due to small convenience samples or single case studies. Accordingly, it is not surprising that many research gaps exist.

Research Gaps in Sex Doll Design

Regarding *sex doll design* (RQ2), many publications agree that the sexualized and idealized looks of female sex dolls pose a problem in terms of further sexual objectification of women within a patriarchal consumer culture already saturated with unrealistic beauty standards for women's bodies. However, previous research falls short on conceptualizing the *sexual fantasy dimension of sex dolls*. Understanding dolls as embodied sexual fantasies, it is neither surprising nor questionable that dolls do not mirror reality as it is, or as it ethically ideally should be, but unapologetically express unrealistic, exaggerated, clichéd, and thus exciting and satisfying fantasies. Research on sexual fantasies has revealed that humans of all genders are usually not particularly turned on by morality or normality but often by the direct opposite [55,56]. What technological change brings about is ample new possibilities to express and materialize sexual fantasies formerly enjoyed purely privately so that they now become readable, audible, visible, and—with dolls and robots—even tangible in the public realm.

Although a sex-positive perspective usually acknowledges the value of fantasy, creativity, play, provocation, and pleasure, a critical perspective usually warns against the expression and dissemination of fantasies whose content is not in line with the ethical standards applied to real life. Obviously, child sex dolls are regarded as a hard limit in the academic sex doll literature. However, for other types of fantasies that dolls can and could embody, there is no consensus and not even a rational debate.

If the breast size of female sex dolls poses a problem (many authors complain about the female dolls' *pornographic looks*), what range of breast sizes would be ethically correct and/or harmless enough regarding the prevailing beauty norms for female bodies? Do we need size norms for ethical dildos and vibrators as well? Questions like these are both banal and profound at the same time: meaningful critical evaluations of sex doll design should go beyond the trivial observation that sex dolls look like sexual clichés because that is exactly the point of sexual fantasy products. Young-looking sex dolls and related products like full-body cushions depicting sexualized young women (so-called *dakimura*) are often criticized, but, in Japan, for example, their main target group is young men and adolescent boys [35]. Is it inappropriate that they desire same-age dolls? Do we want older-looking dolls to be marketed to them? Racial issues are also very confusing. Regarding racial prejudices and privileges, one might problematize that in Japan,

XSL•FO
RenderX

for example, exclusively Japanese-looking sex dolls that emphasize skin whiteness are marketed [35], whereas one may also problematize the marketing of Japanese-looking sex dolls to non-Japanese customers. Is there any way to criticize (and improve) sexual fantasy product designs and marketing strategies that take into consideration both the concern about social inequalities and vulnerabilities and the concern for sexual rights and freedom of sexual fantasy and expression?

The same issues have been discussed for decades regarding pornography [57]. Although some authors still claim that all pornography is inherently inhumane and sexist, just as some authors and activists claim that all sex dolls are inherently inhumane and sexist, other authors and activists accept that sexual explicitness and lack of realism are necessary ingredients of sexual fantasy products. However, they push for a greater variety of fantasies to be represented in the products. This is why female-friendly, couple-friendly, feminist, and queer pornographies have been produced and investigated since the 1980s [58]. The sex doll market could also be diversified. Exploring directions for diverse sex doll designs and their implications could be a task for future research. Design studies could bring together sex researchers, current and future customers with different gender and sexual identities and lifestyles (including older people and people with disabilities), sexual health experts, designers, and/or industry representatives. Collaborations with the sex doll industry promise new insights and, against common belief, do not imply the abandonment of critical analysis. Indeed, critical analysis is often much sharper and more to the point if researchers are closely familiar with the research subject and its context instead of only looking at it from a distance.

Research Gaps in Sex Doll Use

Although pornography use has become mainstream among men and women, it is unclear how large the sex doll user population is and whether it will grow or stay a niche market. Systematic analyses of market data and representative surveys of national populations regarding the prevalence and acceptance of sex doll use are widely lacking. In China, for example, due to the former one-child policy, there is a demographic surplus of millions of men—will they become a target group for sex dolls (HJ Nast, unpublished data, 2019)? With aging societies and a persistent gender gap in life expectancy, we will see a surplus of millions of widows and single older females in developed countries—perhaps another target group for sex dolls and further innovative sexual technologies.

Previous studies on sex doll owners' experiences have demonstrated that men create complex, multi-dimensional relationships with their dolls that include, but are not limited to, the search for sexual gratification. To further explore the *psychology and sexuality of doll play and human-doll relationships*, theories, methods, and results from related research fields should be considered.

Although sex doll conceptualization struggles with the passivity and object status of dolls and the one-sidedness of human-doll relationships, in the field of media research, the concept of one-sided "parasocial relationships" between humans and media figures has been well developed for more than half a century [59]. It is also established that parasocial interactions and relationships are linked to well-being [60]. Romantic and erotic relationships between humans and media figures are common (eg, adolescent girls falling in love with members of boy groups from the music industry) and psychologically meaningful and helpful [61]. Established measures for parasocial interactions and relationships between humans and media personas could be adapted to investigate human–sex doll interactions and relationships.

Surprisingly, research on men's play with female sex dolls has widely ignored the research on children's play with childlike dolls and research on women's play with babylike dolls. Children love, kiss, cuddle, talk to, and sleep with their dolls, and sometimes, they poke their dolls' eyes, cut their hair without consent, or open their stomachs during questionable operations [62-64]. However, nobody assumes that children's use of childlike dolls makes them antisocial or encourages them to treat other children like objects. The same holds true for the female adult doll owner community that uses realistic baby dolls (so-called *Reborn Dolls*,). Here, women use doll play to express sexuality-related fantasies of procreation and motherhood without being accused of antisocial inclinations or objectification of babies, although their behavior is criticized and scandalized in the media [51,65]. Last but not least, research on sex doll use could learn from research on so-called *doll therapy* [66]. Doll therapy addresses dementia patients and encourages holding, kissing, cuddling, talking to, feeding, or dressing an anthropomorphic doll because interactions and relationships with dolls provide comfort, control, and peace as well as feelings of pride, purpose, and bonding that can alleviate agitation and other symptoms [66]. Such soothing and healing effects of dolls have also been reported by sex doll owners. Theoretical elaboration is needed to link and/or differentiate the various user groups and uses of different types of dolls. Why is men's play with sex dolls so outstanding in its assumed connections with antisocial tendencies and an unhealthy confusion of play and reality? Are male gender and sexual fantasy dolls such a dangerous coupling and/or are we dealing with sex-negative and gendered projections?

The previous literature points to different types of sex doll owners like the passionate, possibly paraphilic, lifelong *iDollator*; the misogynist doll owner, the possibly sadistic doll owner striving for complete dominance; the pedophilic doll owner; the transient doll user working through hurt and heartbreak or through teenage angst; the unattractive, old, or disabled user with very limited prospects of success in the real partner market; the doll photographer and hobbyist; or the sexually experimental female user and couple. However, a definitive typology is missing. According to the literature, approximately 20% of the sex doll owner community are couples and females [67], and thus far, we do not know much about them.

Research Gaps in Sex Doll Effects

Sex doll effects of both long-term and short-term sex doll use are under-researched. Long-term domestic use by doll owners has only been explored with small convenience samples and mostly without the use of established and validated measures

XSL•FO
RenderX

for predictors and outcomes of sex doll use, for example, measures of sexual and mental health, personality, sociability, sexism, doll-related paraphilias (eg, objectophilia and doll fetishism), and new sexual identities (eg, digisexuality) [68]. Short-term commercial uses of sex dolls and their effects are completely unexplored. Interviews with customers of *sex doll brothels* and expert interviews with sex doll brothel staff could be helpful. The therapeutic uses and effects of sex dolls have also been under-researched. More clinical case studies are necessary.

What is special about sex dolls as sexual fantasy products is their materiality: they are embodied sexual fantasies, and their use demands specific sexual skills–fantasy skills to enrich the parasocial interaction and practical skills in positioning and moving the heavy doll to create an enjoyable and satisfying sexual experience. Thus far, no observational or experimental studies of social or sexual interactions between humans and sex dolls and their outcomes have been conducted.

State of Research on Sex Robots

We have summarized the state of research on sex robots by mapping the number and type of publications, reporting their main results and indicating the research gaps.

Amount and Type of Research on Sex Robots

During the scoping review literature identification process, we included 98 academic publications on sex robots (Figure 2). This body of literature consists of 6 distinct groups of publications according to both their topics and their methodologies (Table 2). The groups of sex robot publications are similar to those of sex doll publications, the main difference being unavailability of clinical case studies for sex robot, but the availability of many ethical studies and some design studies.

The largest group of sex robot publications (40/98, 41%) deals with sex robot conceptualization and theory, written by authors from social and life sciences, humanities, philosophy, and engineering. The second largest group of publications (28/98, 29%) addresses the ethics of sex robots and is mainly rooted in philosophy. The third group of publications contains empirical studies on sex robot use and effects (12/98, 12%), mainly from the fields of psychology and human-computer interaction. The fourth group of publications addresses sex robot representations in art and media (8/98, 8%), the fifth group of publications looks at child sex robots and their legal regulation (6/98, 6%), and the sixth and final group of publications involves sex robot design studies (4/98, 4%).

The body of academic literature contains 3 published monographs focusing exclusively on sex robots [1,69,70]. Approximately one-third of the included sex robot publications are peer reviewed (32 out of 98). Many sex robot publications are papers from the international conference series LSR (*Love and Sex with Robots*), initiated by David Levy (LSR1 2014 in Funchal, Portugal; LSR2 2016 in London, United Kingdom; LSR3 2017 in London, United Kingdom; and LSR4 2019 in Brussels, Belgium). The Google Scholar citation count reveals a range from 0 to more than 500 citations, the latter for David Levy's [1] seminal book *Love and Sex with Robots*. Heavily cited sex robot publications are often not peer reviewed. Regarding the timeline, the oldest sex robot publication identified in the databases and included in our review is a 1997 comment of a sociologist on the impact of future sex robots [71] that raises questions still discussed today. However, it is an outlier, with approximately 85% (83/98) of the sex robot publications having been published in the last 5 years (2015-2019; Figure 3).

XSL•FO
RenderX

Table 2. Amount and type of research on sex robots (N=98 included academic publications, based on literature search in August 2019).

Reference	Citation count[a]	Peer review	Academic discipline
Sex robot conceptualization and theory (n=40)			
Adshade (2017) [72]	0		Economics
Barber (2017) [73]	2		Creative arts, film, and media
Bołtuć (2017) [74]	3		Philosophy
Carpenter (2017) [75]	2		Human-technology interaction
Cheok et al (2017) [26]	3		Pervasive computing
Cox-George and Bewley (2018) [17]	6		Medicine
Cranny-Francis (2016) [76]	1	✓	Gender studies
Danaher (2017) [77]	4		Ethics and law
Danaher (2017) [78]	3		Ethics and law
Danaher et al (2017) [79]	8		Ethics and law
Devlin (2015) [80]	13		Computer science
Devlin (2018) [69]	7		Computer science
Döring and Pöschl (2018) [11]	6	✓	Psychology
Eggleton (2019) [81]	1		Medicine
Evans (2010) [82]	10		Robotics
Facchin et al (2017) [83]	6		Clinical psychology
Goldfeder and Razin (2015) [84]	7	✓	Law and religion
Gutiu (2016) [85]	7		Law
Hall (2017) [86]	2		Computer science
Hauskeller (2017) [87]	1		Philosophy
Herzfeld (2017) [88]	1		Science and religion
Klein and Lin (2018) [89]	1		Technology ethics
Kolivand et al (2018) [90]	1		Computer science
Lee (2017) [70]	11		Media studies
Levy (2007) [1]	531		Artificial intelligence
Levy (2017) [91]	5		Artificial intelligence
Mackenzie (2018) [24]	4	✓	Law and medical ethics
McArthur and Twist (2017) [68]	11	✓	Philosophy
Migotti and Wyatt (2017) [92]	0		Philosophy
Musiał (2019) [93]	0		Philosophy
Nyholm and Frank (2017) [94]	8		Philosophy
Pearson (2015) [2]	8		Futurology
Richardson (2016) [13]	77	✓	Social anthropology
Rousi (2018) [95]	1		Cognitive science
Rousi (2018) [96]	0	✓	Cognitive science
Sharkey et al (2017) [97]	35		Computer science
Snell (1997) [71]	9		Sociology
Søraa (2017) [98]	9	✓	Interdisciplinary studies of culture
Wennerscheid (2018) [99]	0		Literary studies
Yeoman and Mars (2012) [15]	89	✓	Tourism management
Ethics of sex robots (n=28)			

XSL•FO
RenderX

Reference	Citation count[a]	Peer review	Academic discipline
Amuda and Tijani (2012) [100]	14	✓	Law and theology
Bendel (2015) [101]	23		Technical philosophy
Bendel (2017) [102]	9		Technical philosophy
Beschorner and Krause (2018) [103]	1		Business ethics
Carvalho Nascimento et al (2018) [104]	0	✓	Bioethics
Coeckelbergh (2009) [105]	62	✓	Philosophy of media and technology
Di Nucci (2016) [106]	3		Ethics
Di Nucci (2017) [107]	3		Ethics
Frank and Nyholm (2017) [108]	13	✓	Philosophy and ethics
Goldstein (2017) [109]	0		Political science
González-González et al (2019) [110]	0		Gender studies
Levy (2012) [14]	27		Artificial intelligence
Mackenzie (2014) [111]	6		Law and medical ethics
Mackenzie (2018) [23]	1	✓	Law and medical ethics
McArthur (2017) [112]	2		Philosophy
Petersen (2017) [113]	2		Philosophy
Richardson (2016) [114]	2		Social anthropology
Richardson (2016) [115]	27	✓	Social anthropology
Russell (2009) [116]	8	✓	Law
Shen (2019) [117]	2		Law
Simmons (2016) [118]	1	✓	Law
Sparrow (2017) [119]	25	✓	Philosophy
Spencer (2011) [120]	1		Theology
Sullins (2012) [121]	75	✓	Philosophy
Wagner (2018) [122]	0		Robotics
Welsh (2015) [123]	—[b]		Robot ethics
Whitby (2012) [124]	0		Philosophy and ethics
Ziaja (2011) [125]	7		Law
Empirical studies on sex robot use and effects (n=12)			
Appel et al (2019) [126]	0	✓	Psychology
Bartneck and McMullen (2018) [67]	3		Human-computer interaction
Edirisinghe and Cheok (2017) [127]	2		Human-robot interaction
Edirisinghe et al (2018) [128]	1		Human-robot interaction
Korn et al (2018) [129]	0		Human-computer interaction
Richards et al (2017) [130]	8		Communication
Scheutz and Arnold (2016) [131]	55		Computer science
Scheutz and Arnold (2017) [132]	2		Computer science
Szczuka and Krämer (2017) [133]	2		Psychology
Szczuka and Krämer (2018) [134]	1	✓	Psychology
Szczuka and Krämer (2019) [135]	0	✓	Psychology
Yulianto and Shidarta (2015) [136]	4	✓	Human-robot interaction
Sex robot representations in art and media (n=8)			

XSL•FO
RenderX

Reference	Citation count[a]	Peer review	Academic discipline
Barber (2009) [137]	6		Creative arts, film, and media
Beggan (2017) [138]	0	✓	Sociology
Conn (2017) [139]	0		Comparative literature
Döring and Poeschl (2019) [27]	0	✓	Psychology
Gevers (2018) [140]	0		Art
Hasse (2019) [141]	1	✓	Anthropology
Hauskeller (2014) [142]	29		Philosophy
Hawkes and Lacey (2019) [143]	0	✓	Media studies
Legal regulation of child sex robots (n=6)			
Behrendt (2018) [18]	3		Philosophy
Chatterjee (2019) [52]	0	✓	Criminology
Danaher (2017) [144]	33	✓	Ethics and law
Danaher (2019) [145]	0	✓	Ethics and law
Maras and Shapiro (2017) [20]	7		Criminology and law
Strikwerda (2017) [146]	5		Law and ethics
Design of sex robots (n=4)			
Bendel (2018) [147]	3		Technical philosophy
Danaher (2017) [148]	1		Ethics and law
Gomes and Wu (2018) [9]	0	✓	Engineering
Su et al (2019) [48]	0	✓	Human-computer interaction

[a]Citation count according to Google Scholar in August 2019.

[b]Google Scholar did not list the reference.

Research Findings on Sex Robots

In the following sections, the main findings of previous sex robot research will be reported separately for the 6 groups of sex robot publications (Table 2).

Sex Robot Conceptualization and Theory

The largest group of sex robot publications (40/98, 41%) deals with the conceptualization of sex robots and of human–sex robot relationships. Within this group, 2 issues are predominant: the (non)inherent sexism of sex robots and the (non)humanness of sex robots.

Echoing the critical feminist conceptualization of female sex dolls, several publications on sex robots characterize the female sex robot as an *inherently sexist object*. The most cited author of this position is Kathleen Richardson [13], who conceptualizes the female sex robot as a representative or surrogate of a sexually objectified woman, a female (forced) porn actor, a female (forced) prostitute, or a female sex slave. Following this conceptualization, the production and use of female sex robots is regarded as harmful for individual male users, their female partners, and society at large, as female sex robots symbolically reinvent and reaffirm the status of women as sex slaves [85]. However, this conceptualization operates more with metaphorical equations than established theories and is challenged by other publications as vague and unconvincing [79,80,89]. Although existing sex robots might appear sexist, different designs are possible; therefore, sex robots are not inherently sexist, according to other authors [11,69,76,78,98].

Further theoretical publications deal with the question of the humanness of the sex robot. Several publications stress that, by their definition, sexual interactions and intimate relationships are bidirectional and require a consenting human partner. Humanness—by the definition of these publications—implies sentience, first-person consciousness, and free will; none of these attributes can be ascribed to current sex robots. Consequently, the authors conclude that current sex robots are *nonhuman pseudo persons*. Accordingly, relations with robots are only *pseudorelationships* that inherently lack mutual concern for the welfare of each other [74] and do not lead to personal or spiritual growth [88]. Following this conceptualization, there is no sexual interaction possible *with* a robot or *between* a human and a robot, only robot-enhanced solipsistic masturbation [83].

There are also publications that focus on *future advanced sex robots* and their humanness. Several authors assume that, in the foreseeable future, sex robots could be produced that are sentient, self-conscious, and have a free will [24]. They might even have the legal status of citizens so that humans can legally marry them [72,84,91]. Such advanced humanoid robots will be so human-like that they must be conceptualized as persons and relations with them as interpersonal relationships. Advanced sex robots with excellent social and sexual skills and perfect looks who enter relationships with humans out of their own free

XSL•FO
RenderX

will could be very attractive for many people [82]. Advanced sex and love robots could bring more love to the world [94], but they could, at the same time, devalue real humans [93]. However, the concept of an advanced sex robot that is almost indistinguishable from a real human paradoxically makes it seem pointless to build sex robots. If the advanced sex robot acts like a self-determined, willful human, if it consequently lies, cheats, criticizes, disregards, rejects, and leaves the human, what is the merit of creating it in the first place [87,92,95,96]?

Obviously, there is an inherent tension in the conceptualization of the degree and quality of humanness of interactive humanoid robots. A dumb robot is easy to control but lacks autonomous capabilities and sociability; hence, it cannot bring much additional value to traditional sex dolls. An intelligent advanced robot provides true sociability but lacks the manageability and obedience that we expect from a service technology.

A third conceptualization overcomes the divide between a sex robot as a mere masturbation aid and a sex robot as a quasihuman and stresses that successful sex robots can easily be imagined as purposefully designed to be nonhuman like regarding appearance, functionalities, and social role. Possibilities might be sex robots as synthetic animals [77], as fantasy creatures, or as interfaces to other types of sexual entertainment technology (*fantasy hardware*) [86]. Sex robots could be cherished and desired by humans, particularly by *digisexuals* or *technophilics*, precisely because of their *fundamental otherness* [68,99]. As humans can bond with sex dolls, it is even easier for them to form meaningful emotional attachments with interactive sex robots [26,73,75]. Instead of insisting that advanced sex robots be as human-like as possible to legitimize sexual interactions and emotional attachments with them, robots could also be accepted as nonhuman social agents, for example, to provide safe sex work [15], alleviate social and sexual deprivation [81], or allow for safe explorations of sexual fantasies [2].

The different implications of a wider use of sex robots are addressed by many theoretical publications as unanswered questions [71,97], for example, regarding health [17], social norms [70], and religious beliefs [90].

Ethics of Sex Robots

What is the *right thing* to do in view of the emergence of sex robots? The second largest group of sex robot publications (28/98, 29%; Table 2) attempts to tackle this core question of sex robot ethics. Although some authors stick to metareflection and debate which ethical approach to use [104], other authors provide answers of 3 different types:

- *Sex robots should not be built and used at all*. Starting from the assumption that human-human sexual and romantic relationships are most healthy and ethically superior to all human-robot pseudorelations and to the use of sexualized and sexist robotic objects, authors with different political [109], philosophical [119], feminist [114,115], theological [100,120], engineering [122], and legal [118] academic backgrounds reject further developments in this field. They call for bans and boycotts, stigmatization of, and abstinence from sex robots. Quite popular and often quoted in the media are the arguments of the earlier mentioned Kathleen Richardson, founder of the *Campaign against sex robots*, who compares sex robots with killer robots and with female sex slaves [114,115].
- *Sex robots should be built in an ethical way to avoid harm to humans, especially vulnerable humans*. Starting from the assumption that sex robots can be a good thing if they alleviate loneliness and/or sexual deprivation and contribute to the sexual and social well-being of individuals and couples, authors with different backgrounds encourage ethical design [105]. Publications are very diverse and often vague as to what exactly they expect from ethical sex robot design. One author explains that she wants sex robots designed in such a way that they do not get involved in acts of infidelity because they have learned the concept of heartbreak [125]. Other authors explicitly do not want sex robots to be *love robots* because they fear humans could be too easily manipulated by robots that fake romantic attachment [121,123]. Others want sex robots designed in a women-friendly [110] and disability-inclusive [106,107] way or demand design that is more focused on consumer safety [117]. Some authors point out the many different questions for ethical design ranging from "Should the robot become active on its own and entice the partner to have sex?" to "How should the robot collect an evaluate patient data to better satisfy its partner's sexual needs?" [101,102]. As some authors assume that the development of sex robots is driven by a profit-oriented *uncaring industry* [124], there is a need for more involvement of ethically responsible entrepreneurs and designers from different backgrounds who aim to develop and market sex robots for sexual well-being, pleasure, fun, and play while taking into consideration the concerns and desires of diverse user and stakeholder groups. Some authors are very optimistic that sex robots will bring a lot of pleasure and happiness and are, therefore, ethically a good thing [112], although some ethical issues are unresolved (eg, regarding *robot prostitutes*) [14]. Other authors stress that robots are a good thing only for a very small group of people who absolutely cannot find a human sex partner [124].
- *Sex robots should be built in an ethical way to avoid harm to robots, especially for advanced sentient robots*. Starting from the assumption that future humanoid robots will be advanced to a very high degree of human likeness, according to several authors, their sexual and other citizen's rights must be protected with a *nonanthropocentric* but *robocentric* ethic [23,24,103,111,116]. For example, sex with an advanced sex robot should only be acceptable if the robot has given explicit consent [108]. Although some authors stress the relevance of a robocentric ethic for sex robots to protect them from anticipated harm and exploitation, other authors argue that sentient robots designed as sex robots could have a "good life" and experience pleasure and satisfaction from fulfilling their tasks [113].

Empirical Studies on Sex Robot Use and Effects

The third group of sex robot publications contains empirical studies (12/98, 12%; Table 2). Thus far, not a single empirical

XSL•FO
RenderX

study has been published that deals with the small but presumably growing number of pioneer users of sex robots. All existing studies address nonusers and investigate their attitudes toward sex robots and their reactions to sex robot–related stimuli.

Most empirical studies (8/12, 67%) are small web-based surveys on sex robot acceptance using convenience samples from the United States (N=261: [126], N=133: [130], N=100: [131], and N=198: [132]), Germany (N=263: [133]), Indonesia (N=380: [136]), and Malaysia (N=32: [127]). Their results show diverse rates of sex robot acceptance. For example, 40% of male and 17% of female respondents in the United States (mean age, 33 years) reported willingness to try out a sex robot [131] in comparison with 16% of Indonesian respondents [136] and 9% of Malaysian respondents [127]. Cultural background, male gender, positive attitudes toward robots in general, interest in manga and games, sensation seeking, and shyness appeared to be predictors of sex robot acceptance. Interestingly, sexual and relationship satisfaction did not predict sex robot acceptance [130,133]. However, because of the small nonrepresentative samples, the generalizability of existing sex robot acceptance data is very limited. Another problem is the varying operational definitions of sex robots given to respondents in the surveys. A Delphi survey explored the predictions of 20 social robot experts who were reluctant regarding sexual apps [129], whereas 1 expert interview explored the sex robot predictions of the founder of sex doll and sex robot manufacturer Abyss Creations, Matthew McMullen [67].

In addition to the survey and interview studies, 3 experimental studies were found (3/12, 25%). They investigated how heterosexual women experience their male partner's imagined infidelity with a female robot vs a real woman [134], at which body parts of female robots vs female humans, both represented in pictures, male and female gaze [135], and how humans physically react when they touch different, including private, body parts of a robot that is not a sex robot [128]. Overall, these experiments show differences and similarities in humans' sexuality-related reactions to humanoid robots and fellow humans. So far, no experimental study exists that uses an actual sex robot as the stimulus material.

Sex Robot Representations in Art and Media

The fourth group of sex robot publications concerns sex robot representations in art and media (8/98, 8%; Table 2). In their selected and analyzed examples from the science fiction literature, some studies from humanities focus on fictional female sexbots that seem to embody male fantasies of the ideal woman but who, in the course of the action, become *feminist robots* striving for independence from their male human partner or creator by leaving or even killing him, for example, the robot *Ava* in the 2015 UK movie *Ex Machina* [141,143], the virtual *Samantha* in the 2013 US movie *Her* [143], or the robotic wife in the 1981 Chinese story *Conjugal Happiness in the Arms of Morpheus* [139]. The famous US television series *Star Trek Voyager* presented the character *Seven of Nine*, a cybernetic organism and former Borg drone, who—although embodying traditional feminine beauty—challenged traditional ideas of gender and sexuality [137]. Other dystopian science fiction representations, selected and analyzed by the academic literature, illustrate the female sex robots' sexual exploitation and victimization, for example, as porn actors in the 2009 US movie *2040* [138] or as rape victims in the US television series *Westworld* [143]. One monograph critically analyzes posthuman utopias in sex robot representations [142], and one editorial volume documents the *Robot Love 2018* International Expo of the *Niet Normaal Foundation* in the Netherlands that brought together researchers and artists [140].

A quantitative media content analysis examined the representation of human–sex robot relationships in 370 fictional and 340 nonfictional media examples [27]. The results of this study indicate that media representations of intimate human-robot relationships tend to portray the human partner as a man who is disadvantaged in interpersonal relationships. At the same time, media often portray the involved robot partner as a humanoid female sex robot. Although nonfictional media describe intimate human-robot relationships more often in sexual terms, fictional media focus more on emotional aspects, cohabitation, and even procreation between humans and robots. Overall, media representations of intimate human-robot relationships reveal stereotypical gender roles, heteronormativity, and a focus on sexual vs emotional intimacy [27].

Legal Regulation of Child Sex Robots

The fifth group of sex robot publications covers child sex robots (6/98, 6%; Table 2). All the 6 publications [18,20,52,144-146] characterize child sex robots as harmful and unethical and call for a legal ban that is already in preparation or in effect in several countries (eg, the aforementioned *CREEPER Act of 2017* in the US) [20]. In all, 2 publications speculate on the possible therapeutic uses of child sex robots. Although one of them assumes that their exploration would be too risky [145], the other encourages their exploration only in certain, controlled circumstances under strict medical supervision and in accordance with guidelines issued by an ethics committee [18].

Design of Sex Robots

Only 4 publications in the sex robot literature focus mainly on design (4/98, 4%; Table 2): 1 on erotic voice output [147], 1 on a mind-controlled neurodildo to be used separately or implemented in robots [9], 1 on general design aspects based on results about sex doll use [48], and 1 on feminist sex robot design in an analogy of initiatives for feminist pornography [148].

Research Gaps in Sex Robots

Although the body of sex robot publications is 3 times as large as that of sex doll publications, *empirical* studies on sex robot use and effects are equally scarce (4 peer-reviewed papers in total). Fundamental questions regarding the sexual use of human-like full-body material artifacts that remained unanswered for sex dolls also remain unanswered for sex robots.

Research Gaps in Sex Robot Design

Just as with sex dolls, the question of how much fantasy, and which and whose fantasies should legitimately be implemented in sexual fantasy products to make them socially acceptable,

XSL•FO
RenderX

harmless, and still sexually desirable and exciting, remains unanswered with sex robots as well. Although there is a lot of speculation on the possible therapeutic uses of sex robots to be found in public and academic debates, the literature fails to provide design guidelines for therapeutic sex robots informed by evidence from sex and relationship therapy and focused on specific problems (eg, on sexual shyness and anxiety, sexual dysfunctions, sexual trauma, paraphilias, and paraphilic disorders). Design studies for current sex robots hardly exist, and the literature predominantly speculates about imagined future sex robots. Instead of researchers, it was journalists who first dealt with the question, "What would sex robots for women look like?" [149] and who let women and men draw and explain their *ideal sex robots* [150].

Research Gaps in Sex Robot Use

Thus far, no empirical study has investigated experienced sex robot users or interactions of unexperienced participants with actual sex robots. Results from research on sex doll use are, therefore, the best available proxy for sex robot use. Regarding the potential market size and user population, there are also no data available that allow for sound predictions. One might speculate that sex robots could overcome some of the stigmatization of sex dolls as sex robots can be framed as cutting edge, high-tech products. Thus, their users might appear more modern, future oriented, and competent in comparison with traditional sex doll owners. Against this backdrop, one would expect more growth for the sex robot market than for the sex doll market, but data are needed. To further explore the sexual appeal of robots, insights from research on objectophilia [151,152] and technofetishism [153] could be helpful.

Surprisingly, the sex robot literature falls short in conceptualizing and investigating interactions and relationships between humans and current sex robots in a psychologically nuanced way. Whereas the sex doll literature has already established that dolls easily trigger humans to build meaningful, caring, loving, long-term relationships with them, the sex robot literature often falls back on binary thinking. It categorizes the current sex robot as an inanimate object and mere masturbation aid without any sociability and is only willing to ascribe sociability to future imagined sex robots that are advanced to the point of indistinguishability from humans. Hence, the literature on sex robots often misses the key point that robots are more than mere masturbation aids due to anthropomorphization and that they are meaningful and possibly helpful precisely because they are not substitutes for real humans but are sociotechnical entities for parasocial use and play. Parasocial interactions and play with sexual fantasy products grant more degrees of freedom in sexual expression and allow to take a break from all of the norms, ethics, expectations, and responsibilities of human-human interactions.

Research Gaps in Sex Robot Effects

As sex robot users and use are completely unknown thus far, any claims about positive and/or negative effects are mere speculation. Although some authors are so convinced of their speculations on strong to catastrophic negative effects that they demand immediate boycotts and bans of sex robots, others urgently call for empirical research on sex robot effects. The idea that sex robots allow humans to indulge in interactive embodied sexual fantasies elicits strong projections of lust and fear. Most likely, empirical research will help us overcome exaggerated projections and understand the diversity and ambivalence of effects on different types of sex robot users.

Discussion

Main Results of the Review

In conclusion, the main results of the whole review are summarized, its limitations are indicated, and a roadmap for future research is drawn. The body of sex robot literature, with approximately 100 academic publications in total, is more than 3 times larger than that on sex doll literature, with approximately 30 publications (RQ1). However, only a handful of peer-reviewed empirical papers on both sex doll use and sex robot use are available thus far. No sex robot study exists that investigates people experienced in sex robot use and/or introduces actual sex dolls or sex robots as stimulus material. Regarding the first RQ, one must concede that sex dolls and sex robots, although attracting growing public and scholarly attention, are heavily under-researched. Both sex doll and sex robot research are fields characterized by disciplinary diversity, with notable participation from philosophy, humanities, and engineering, and a conspicuous lack of participation from sex researchers.

Sex doll and sex robot designs (RQ2) are often critically assessed in the literature, mainly because the bodies of women-like dolls and robots are usually designed in sexualized ways following and exaggerating traditional feminine beauty ideals. However, when understanding sex dolls and sex robots as sexual fantasy products, it makes sense that they do not imitate reality but cater to sexual fantasy. Often, it is exactly the point of sexual fantasies to be unrealistic. Thus far, the literature has not addressed the core question of how we could and should assess designs of sexual fantasy products such as sex dolls and sex robots, considering both social inequalities and vulnerabilities and the freedom of sexual fantasy and expression. Regarding future advanced sex robots, the literature presents various requirements for ethical design, which—at the current state of robot development—are very speculative. Systematic design studies that work with current and future users (eg, private sex doll owners, sex workers, and sex therapists) and address different use scenarios (eg, domestic, commercial, or therapeutic) are lacking.

Although previous research has provided some insights into the domestic long-term use of sex dolls (with or without parallel psychotherapy), no data have been collected thus far on the short-term interactions or long-term relationships between humans and sex robots. Thus, the best proxy for sex robot use and users today is the limited data on sex doll use and users (RQ3).

Considering the lack of empirical knowledge about sex doll users and sex robot users and use, it is obvious that the predictions of positive and negative effects found in the literature can only be speculative (RQ4). It is striking that authors still provide very strong and contradictory effect claims ranging

XSL•FO
RenderX

from utopian visions of improved sexual satisfaction and overall happiness to dystopian visions of dehumanization, objectification, and isolation. Predictions of small and/or ambivalent effects might be more realistic but are seldom discussed in the academic literature thus far, which seems to mirror some of the hype and scandalization observable in public media discourses.

Limitations

This scoping review addressed sex doll and sex robot research as far as it is represented in the accessible literature published before August 2019. We were particularly careful to retrieve publications not only from the databases but also, in a systematic way, from all the included publications' reference lists. Nevertheless, it must be taken into consideration that further studies that have not (yet) been published and/or were not (yet) accessible (eg, conference presentations, qualifications theses, and journal articles under peer review) could exist. However, we are confident that our systematic literature identification strategy covered previous research thoroughly enough, especially as this is the very first systematic review of the field.

To map previous research in a comprehensible and useful manner, we organized the body of literature by building distinct groups of publications according to their key topics and methodologies. We discussed data charting and synthesis within the team and checked everything in duplicate. However, some decisions might be questionable. There is an inherent tension between the aim of providing a clear and comprehensible structure, which requires a reduction in complexity, and the aim of doing justice to the individual publications, which requires a representation of their complexities. Due to space constraints, we were forced to reduce the complexity much more than we would have wished. Therefore, we encourage readers to consult the original publications whenever in doubt and apologize to fellow researchers in case they feel our review misrepresents their work.

Another inherent problem of a multidisciplinary review lies in the tension between the aim of doing justice to discipline-specific styles of knowledge production and communication and the aim of presenting existing knowledge in a consistent, readable, and generally understandable way. We deliberately simplified concepts and streamlined discipline-specific jargon to improve consistency. We provide a broad overview spanning from ancient Greek myths to contemporary web-based surveys, and spanning from the psychoanalyst's office to the robotics lab. We agree with many authors we cite in this review that a deeper understanding of sex dolls and sex robots and their meanings for human sexuality can only be achieved through more interdisciplinary collaboration. We hope that our prioritization of disciplinary width over depth will inspire this collaboration. However, we are aware of the risks and limitations of simplification.

Roadmap for Future Research

Hopefully, the many and diverse research gaps pointed out in this review can serve as starting points for future research projects on sex dolls and sex robots, their design, use, and effects. To conclude, we suggest 4 selected, particularly urgent research strands:

1. *Public debates* about and *media representations* of sex dolls and sex robots, polarized and scandalizing as they are, attract much attention, shape public opinions, and influence research activity. They deserve more scholarly analysis and participation by the sex research community. This includes traditional mass media and social media. Mass media tend to assume dramatic positive or negative effects, while often completely ignoring the fact that sex dolls and sex robots, overall, could have only small and/or ambivalent effects. Social media, sometimes, offer more nuanced views, but expert statements and documentaries about sex robots on YouTube, for instance, are met by a noteworthy amount of misogynist comments that welcome female sex robots as substitutes for women. On Twitter and Instagram, we see sex dolls communicating to the public, their accounts steered by doll owners (eg, the earlier mentioned *Davecat*) and by doll manufacturers. These examples illustrate that we need to know more about media representations as they are an important element of the cultural context in which sex dolls and sex robots are developed, marketed, discussed, used, and investigated today.
2. Research on the sexual uses of human-like *material artifacts* such as sex dolls and sex robots needs to be advanced and connected to research on human-like *digital artifacts* such as chat bots, avatars, holograms, or immersive virtual reality pornography. After all, a sexual AI system trained by a particular user could be used on different technological and media platforms such as a full-body sex robot, an immersive virtual reality system, or a smartphone. Although the materiality of dolls and robots offers new possibilities in terms of embodied sexual fantasies (eg, physical presence, physical care, physical touch, and physical stimulation), it also creates boundaries (eg, through the body weight and difficult handling of the dolls and robots at home, limited mobility outdoors, high visibility, and risk of social stigmatization). For sexual fantasy products that aim to enhance their users' sexual and social experiences, the right degree and mixture of materiality and virtuality is an open question for research and design.
3. Although sex robots have triggered the publication of many theoretical and ethical papers, we urgently need *empirical data* on actual sex doll and sex robot users and uses. Different study designs (nonexperimental and experimental, cross-sectional, and longitudinal) and data collection methods (qualitative interviews, focus group discussions, surveys, psychological tests, and physiological measurements) are suitable for research with actual sex doll and sex robot users. Instead of using only imagined or visually depicted artifacts as stimulus material, some of the real sex dolls and sex robots should be incorporated in empirical studies.
4. Despite the relatively large number of theoretical papers, the degree of *theoretical elaboration* of human–sex doll/sex robot relations and their consequences is not yet very high [12]. Commonly used theoretical concepts are objectification, gratification, and pseudorelationships. For a more thorough understanding, we suggest including

XSL•FO
RenderX

theoretical concepts from the field of doll play and doll therapy and from the field of human interaction with media personas (parasocial interactions and relationships), with digital technologies (computers as social actors and media equation theory) and with social robots (uncanny valley concept and anthropomorphization) as well as from social psychology (social cognitive learning theory), clinical and developmental psychology (transitional objects, objectophilia, and robophilia), and sexuality research (sexual scripts theory and theories on sexual fantasies). It is not yet clear which theories from the different related research fields on dolls, robots, sexuality, gender relations, well-being, and health are best applicable to human–sex doll/robot relationships and if and how they can be combined to best explain the complex intimate engagements of humans with artifacts.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature searches as of August 6 to 9, 2019.
[DOC File , 35 KB - jmir_v22i7e18551_app1.doc]

References

1. Levy D. Love and Sex with Robots: The Evolution of Human-Robot Relationships. New York, USA: Harper; 2007.
2. Pearson I. The Future of Sex Report: The Rise of the Robosexuals. A Futurizon Report - Bondara. 2015. URL: http://graphics.bondara.com/Future_sex_report.pdf [accessed 2020-06-09]
3. Döring N. The internet's impact on sexuality: a critical review of 15 years of research. Comput Hum Behav 2009 Sep;25(5):1089-1101. [doi: 10.1016/j.chb.2009.04.003]
4. Castro-Calvo J, Giménez-García C, Gil-Llario MD, Ballester-Arnal R. Motives to engage in online sexual activities and their links to excessive and problematic use: a systematic review. Curr Addict Rep 2018 Nov 6;5(4):491-510. [doi: 10.1007/s40429-018-0230-y]
5. Döring N, Poeschl S. Experiences with diverse sex toys among German heterosexual adults: findings from a national online survey. J Sex Res 2019 Feb 26 epub ahead of print. [doi: 10.1080/00224499.2019.1578329] [Medline: 30806076]
6. Herbenick D, Reece M, Sanders SA, Dodge B, Ghassemi A, Fortenberry JD. Women's vibrator use in sexual partnerships: results from a nationally representative survey in the United States. J Sex Marital Ther 2010;36(1):49-65. [doi: 10.1080/00926230903375677] [Medline: 20063237]
7. Wood J, Crann S, Cunningham S, Money D, O'Doherty K. A cross-sectional survey of sex toy use, characteristics of sex toy use hygiene behaviours, and vulvovaginal health outcomes in Canada. Can J Hum Sex 2017 Dec;26(3):196-204. [doi: 10.3138/cjhs.2017-0016]
8. Attwood F. Fashion and passion: marketing sex to women. Sexualities 2016 Aug 15;8(4):392-406. [doi: 10.1177/1363460705056617]
9. Gomes LM, Wu R. User evaluation of the neurodildo: a mind-controlled sex toy for people with disabilities and an exploration of its applications to sex robots. Robotics 2018 Aug 20;7(3):46. [doi: 10.3390/robotics7030046]
10. Morales E, Gauthier V, Edwards G, Courtois F, Lamontagne A, Guérette A. Co-designing sex toys for adults with motor disabilities. Sex Disabil 2017 Oct 17;36(1):47-68. [doi: 10.1007/s11195-017-9506-8]
11. Döring N, Pöschl S. Sex toys, sex dolls, sex robots: our under-researched bed-fellows. Sexologies 2018 Jul;27(3):e51-e55. [doi: 10.1016/j.sexol.2018.05.009]
12. Döring N. [Vom internetsex zum robotersex: forschungsstand und herausforderungen für die sexualwissenschaft]. Z Sex-Forsch 2017 Mar 22;30(1):35-57. [doi: 10.1055/s-0043-101471]
13. Richardson K. The asymmetrical 'relationship': parallels between prostitution and the development of sex robots. SIGCAS Comput Soc 2016 Jan 5;45(3):290-293. [doi: 10.1145/2874239.2874281]
14. Levy D. The ethics of robot prostitutes. In: Lin P, Abney K, Bekey GA, editors. Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge, UK: MIT Press; 2012:223-231.
15. Yeoman I, Mars M. Robots, men and sex tourism. Futures 2012 May;44(4):365-371. [doi: 10.1016/j.futures.2011.11.004]
16. Knafo D. Guys and dolls: relational life in the technological era. Psychoanalytic Dialogues 2015 Aug 12;25(4):481-502. [doi: 10.1080/10481885.2015.1055174]
17. Cox-George C, Bewley S. I, Sex Robot: the health implications of the sex robot industry. BMJ Sex Reprod Health 2018 Jul;44(3):161-164. [doi: 10.1136/bmjsrh-2017-200012] [Medline: 30012720]
18. Behrendt M. Reflections on moral challenges posed by a therapeutic childlike sexbot. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:96-113.
19. Brown R, Shelling J. Exploring the implications of child sex dolls. Trends Issues Crime Crim Justice 2019 Mar;570:1-13 [FREE Full text]

XSL•FO
RenderX

20. Maras M, Shapiro L. Child sex dolls and robots: more than just an uncanny valley. J Internet Law 2017;21(5):3-21. [doi: 10.1093/oso/9780190647162.003.0010]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. The EQUATOR Network. 2018. URL: http://www.equator-network.org/reporting-guidelines/prisma-scr/ [accessed 2019-08-28]
23. Mackenzie R. Sexbots: customizing them to suit us versus an ethical duty to created sentient beings to minimize suffering. Robotics 2018 Nov 11;7(4):70. [doi: 10.3390/robotics7040070]
24. Mackenzie R. Sexbots: sex slaves, vulnerable others or perfect partners? Int J Technoethics 2018;9(1):1-17. [doi: 10.4018/ijt.2018010101]
25. Karlin S. Red-Hot Robots: Roxxxy and Rocky, the World's First Sex Robots, Are Ready to Leave the Lab. IEEE Spectrum. 2010. URL: https://spectrum.ieee.org/robotics/humanoids/redhot-robots [accessed 2020-06-17]
26. Cheok A, Karunanayaka K, Zhang E. Lovotics: Human-robot love and sex relationships. In: Abney K, Lin P, Jenkins R, editors. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. New York, NY: Oxford University Press; 2017:193-213.
27. Döring N, Poeschl S. Love and sex with robots: a content analysis of media representations. Int J of Soc Robotics 2019 Feb 6;11(4):665-677. [doi: 10.1007/s12369-019-00517-y]
28. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
29. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
30. Blizzard D. Making the fantasy: consumption, relationships, and the RealDoll. In: Larsson M, Johnsdotter S, editors. Sexual Fantasies: At the Convergence of the Cultural and the Individual. New York, USA: Peter Lang Publishers; 2015:57-77.
31. Blizzard D. The next evolution: the constitutive human-doll relationship as companion species. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018.
32. Cassidy V. For the love of doll(s): a patriarchal nightmare of cyborg couplings. Eur Soc Cardiol 2016;42(1-2):203-215. [doi: 10.1353/esc.2016.0001]
33. Ferguson A. The Sex Doll: A History. Jefferson, NC: McFarland & Company; 2010.
34. Kim E. Why do dolls die? The power of passivity and the embodied interplay between disability and sex dolls. Rev Educ Pedagog Cult Stud 2012 Jul;34(3-4):94-106. [doi: 10.1080/10714413.2012.686852]
35. Nast HJ. Into the arms of dolls: Japan’s declining fertility rates, the 1990s financial crisis and the (maternal) comforts of the posthuman. Soc Cult Geogr 2016 Sep 26;18(6):758-785. [doi: 10.1080/14649365.2016.1228112]
36. Ray P. 'Synthetic love lasts forever': sex dolls and the (post)human condition. In: Banerji D, Paranjape MR, editors. Critical Posthumanism and Planetary Futures. New Delhi, India: Springer India; 2016:91-112.
37. Wong N. Social Surrogates or Posthuman Lovers?: Love Dolls in the Robotic Moment. Open ULeth Scholarship (OPUS). 2015. URL: https://opus.uleth.ca/bitstream/handle/10133/3753/WONG_NATHAN_MA_2015.pdf [accessed 2020-06-10]
38. Burr-Miller A, Aoki E. Becoming (hetero) sexual? The hetero-spectacle of idollators and their real dolls. Sex Cult 2013 May 18;17(3):384-400. [doi: 10.1007/s12119-013-9187-0]
39. Connor S. Guys and dolls. Women 2015 Jul 16;26(1-2):129-142. [doi: 10.1080/09574042.2015.1035053]
40. Getsy DJ. Queer exercises: Amber Hawk Swanson's performances of self-realization. GLQ 2013 Sep 6;19(4):465-485. [doi: 10.1215/10642684-2279987]
41. Koné CA. Such a Doll!: Man-made Dolls in German Modern Culture and Their Afterlife in Postmodern Visual Culture. New Brunswick, NJ: Rutgers University; 2015.
42. Roos B. Oskar Kokoschka's sex toy: the women and the doll who conceived the artist. Modernism/Modernity 2005;12(2):291-309. [doi: 10.1353/mod.2005.0067]
43. Smith M. The Erotic Doll: A Modern Fetish. New Haven: Yale University Press; 2013.
44. Weisel-Barth J. Loneliness and the creation of realness in lars and the real girl. Int J Psychoanal Self Psychol 2008 Dec 18;4(1):111-118. [doi: 10.1080/15551020802527837]
45. Ciambrone D, Phua V, Avery E. Gendered synthetic love: real dolls and the construction of intimacy. Int Rev Mod Sociol 2017;43(1):59-78 [FREE Full text]
46. Knox D, Huff S, Chang I. Sex dolls - creepy or healthy?: attitudes of undergraduates. J Positive Sexuality 2017;3(2):32-37 [FREE Full text]
47. Langcaster-James M, Bentley G. Beyond the sex doll: post-human companionship and the rise of the ‘allodoll’. Robotics 2018 Oct 8;7(4):62. [doi: 10.3390/robotics7040062]
48. Su NM, Lazar A, Bardzell J, Bardzell S. Of dolls and men. ACM Trans Comput-Hum Interact 2019 Jun 6;26(3):1-35. [doi: 10.1145/3301422]

XSL•FO
RenderX

49. Valverde S. The Modern Sex Doll-owner: a Descriptive Analysis. California State Polytechnic University. 2012. URL: https://digitalcommons.calpoly.edu/theses/849/ [accessed 2020-06-10]
50. Kleist E, Moi H. Transmission of gonorrhoea through an inflatable doll. Genitourin Med 1993 Aug;69(4):322 [FREE Full text] [doi: 10.1136/sti.69.4.322] [Medline: 7721299]
51. Knafo D, Lo Bosco R. The Age of Perversion: Desire and Technology in Psychoanalysis and Culture. New York, USA: Routledge/Taylor & Francis Group; 2017.
52. Chatterjee BB. Child sex dolls and robots: challenging the boundaries of the child protection framework. Int Rev Law Comput Technol 2019 Apr 21;34(1):22-43. [doi: 10.1080/13600869.2019.1600870]
53. ALMA History: The Puppet. ALMA. 2019. URL: https://www.alma-mahler.at/engl/almas_life/puppet.html [accessed 2018-08-30]
54. Winnicott DW. Transitional objectstransitional phenomena. In: Winnicott DW, editor. Through Pediatrics to Psychoanalysis: Collected Papers. New York, USA: Routledge; 1951:229-242.
55. Joyal CC. Defining 'normophilic' and 'paraphilic' sexual fantasies in a population-based sample: on the importance of considering subgroups. Sex Med 2015 Dec;3(4):321-330 [FREE Full text] [doi: 10.1002/sm2.96] [Medline: 26797067]
56. Joyal CC, Cossette A, Lapierre V. What exactly is an unusual sexual fantasy? J Sex Med 2015 Feb;12(2):328-340. [doi: 10.1111/jsm.12734] [Medline: 25359122]
57. Smith C, Attwood F. Anti/pro/critical porn studies. Porn Stud 2014 Mar 21;1(1-2):7-23. [doi: 10.1080/23268743.2014.887364]
58. Williams L. Hard Core: Power, Pleasure, and the 'Frenzy of the Visible'. Berkeley, CA: University of California Press; 1989.
59. Horton D, Wohl RR. Mass communication and para-social interaction; observations on intimacy at a distance. Psychiatry 1956 Aug;19(3):215-229. [doi: 10.1080/00332747.1956.11023049] [Medline: 13359569]
60. Hartmann T. Parasocial interaction, parasocial relationships, and well-being. In: Reinecke L, Oliver MB, editors. The Routledge Handbook of Media Use and Well-Being: International Perspectives on Theory and Research on Positive Media Effects. London, UK: Routledge Taylor & Francis Group; 2016:131-144.
61. Erickson S, Harrison K, Dal CS. Toward a multi-dimensional model of adolescent romantic parasocial attachment. Commun Theory 2018;28(3):376-399. [doi: 10.1093/ct/qtx006]
62. Sears PS. Doll play aggression in normal young children: influence of sex, age, sibling status, father's absence. Psychol Monogr 1951;65(6):1-42. [doi: 10.1037/h0093598]
63. Liss MB. Patterns of toy play: an analysis of sex differences. Sex Roles 1981 Nov;7(11):1143-1150. [doi: 10.1007/bf00287590]
64. Vig S. Young children's object play: a window on development. J Dev Phys Disabil 2007 Apr 20;19(3):201-215. [doi: 10.1007/s10882-007-9048-6]
65. Fitzgerald L. 'Let's play mummy': simulacrum babies and reborn mothers. Eur J Cult Stud 2011 Mar 10;14(1):25-39. [doi: 10.1177/1367549410377142]
66. Ng QX, Ho CY, Koh SS, Tan WC, Chan HW. Doll therapy for dementia sufferers: a systematic review. Complement Ther Clin Pract 2017 Feb;26:42-46. [doi: 10.1016/j.ctcp.2016.11.007] [Medline: 28107848]
67. Bartneck C, McMullen M. Interacting with Anatomically Complete Robots: A Discussion About Human-Robot Relationships. In: ACM/IEEE International Conference on Human-Robot Interaction. 2018 Presented at: HRI'18; March 5-8, 2018; Chicago, IL. [doi: 10.1145/3173386.3173387]
68. McArthur N, Twist ML. The rise of digisexuality: therapeutic challenges and possibilities. Sex Relatsh Ther 2017 Nov 17;32(3-4):334-344. [doi: 10.1080/14681994.2017.1397950]
69. Devlin K. Turned On: Science, Sex and Robots. New York, USA: Bloomsbury; 2018.
70. Lee J. Sex Robots: The Future of Desire. London, UK: Palgrave Macmillan; 2017.
71. Snell J. Impacts of robotic sex. Futurist 1997;31(4) [FREE Full text]
72. Adshade M. Sexbot-induced social change: an economic perspective. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:15.
73. Barber T. For the love of artifice 2: an extension of the paper 'for the love of artifice: why we need robot sex dollswhy there is a growing sub culture of real people trying to become them'. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017.
74. Bołtuć P. Church-turing lovers. In: Abney K, Lin P, Jenkins R, editors. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. New York, USA: Oxford University Press; 2017:214-228.
75. Carpenter J. Deus sex machina: loving robot sex workers and the allure of an insincere kiss. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:14-34.
76. Cranny-Francis A. Is data a toaster? Gender, sex, sexuality and robots. Palgrave Commun 2016 Oct 7;2(1). [doi: 10.1057/palcomms.2016.72]
77. Danaher J. Should we be thinking about robot sex? In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:1-15.
78. Danaher J. The symbolic-consequences argument in the sex robot debate. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:7-37.

XSL•FO
RenderX

79. Danaher J, Earp B, Sandberg A. Should we campaign against sex robots? In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:4-33.
80. Devlin K. In Defence of Sex Machines: Why Trying to Ban Sex Robots is Wrong. The Conversation. 2015 Sep 17. URL: https://theconversation.com/in-defence-of-sex-machines-why-trying-to-ban-sex-robots-is-wrong-47641 [accessed 2020-06-17]
81. Eggleton J. Comment on 'I, sex robot: the health implications of the sex robot industry'. BMJ Sex Reprod Health 2018 Dec 18 epub ahead of print. [doi: 10.1136/bmjsrh-2018-200251] [Medline: 30563852]
82. Evans D. Wanting the impossible. In: Wilks Y, editor. Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues. Amsterdam, The Netherlands: John Benjamins Publishing Company; 2010:75-99.
83. Facchin F, Barbara G, Cigoli V. Sex robots: the irreplaceable value of humanity. Br Med J 2017 Aug 15;358:j3790. [doi: 10.1136/bmj.j3790] [Medline: 28811272]
84. Goldfeder M, Razin Y. Robotic marriage and the law. LSD J 2015;10(12):137-176 [FREE Full text]
85. Gutiu S. The roboticization of consent. In: Calo R, Froomkin AM, Kerr I, editors. Robot Law. North Hampton, MA, USA: Edward Elgar Publishing; 2016:186-212.
86. Hall L. Sex with robots for love free encounters. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex With Robots: Conference Proceedings. New York, USA: Springer; 2017:128-136.
87. Hauskeller M. Automatic sweethearts for transhumanists. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:11-21.
88. Herzfeld N. Religious perspectives on sex with robots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:6-14.
89. Klein WE, Lin VW. 'Sex robots' revisited. SIGCAS Comput Soc 2018 Jul 27;47(4):107-121. [doi: 10.1145/3243141.3243153]
90. Kolivand H, Ehsani RA, Tully D. Virtual sex: good, bad or ugly? In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:26-36.
91. Levy D. Why not marry a robot? In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots. New York, USA: Springer; 2017:3-13.
92. Migotti M, Wyatt N. On the very idea of sex with robots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:2-17.
93. Musiał M. Enchanting Robots: Intimacy, Magic, and Technology. New York, USA: Palgrave Macmillan; 2019.
94. Nyholm S, Frank L. From sex robots to love robots: is mutual love with a robot possible? In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:12-32.
95. Rousi R. Lying cheating robots: Robots and infidelity. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:51-64.
96. Rousi R. Me, my bot and his other (robot) woman? Keeping your robot satisfied in the age of artificial emotion. Robotics 2018 Aug 15;7(3):44. [doi: 10.3390/robotics7030044]
97. Sharkey N, van Wynsberghe A, Robbins S, Hancock E. Our Sexual Future With Robots: A Foundation for Responsible Robotics Consultation Report. Foundation for Responsible Robotics (FRR). 2017. URL: http://responsiblerobotics.org/wp-content/uploads/2017/07/FRR-Consultation-Report-Our-Sexual-Future-with-robots_Final.pdf [accessed 2020-06-10]
98. Søraa RA. Mechanical genders: how do humans gender robots? Gender Technol Dev 2017 Oct 31;21(1-2):99-115. [doi: 10.1080/09718524.2017.1385320]
99. Wennerscheid S. Posthuman desire in robotics and science fiction. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:37-50.
100. Amuda Y, Tijani I. Ethical and legal implications of sex robot: an Islamic perspective. OIDA Int J Sustain Dev 2012;3(6):19-28 [FREE Full text]
101. Bendel O. Surgical, therapeutic, nursing and sex robots in machine and information ethics. In: van Rysewyk SP, Pontier M, editors. Machine Medical Ethics. New York, USA: Springer; 2015:17-32.
102. Bendel O. Sex robots from the perspective of machine ethics. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017.
103. Beschorner T, Krause F. Dolores and robot sex: fragments of non-anthropocentric ethics. In: Choek AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:128-137.
104. Carvalho Nascimento E, da Silva E, Siqueira-Batista R. The 'use' of sex robots: a bioethical issue. Asian Bioeth Rev 2018 Aug 15;10(3):231-240. [doi: 10.1007/s41649-018-0061-0]
105. Coeckelbergh M. Personal robots, appearance, and human good: a methodological reflection on roboethics. Int J of Soc Robotics 2009 Jun 5;1(3):217-221. [doi: 10.1007/s12369-009-0026-2]
106. Di Nucci E. Sexual Rights, Disability and Sex Robots. SSRN. 2016. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2814471 [accessed 2020-06-17]
107. Di Nucci E. Sex robots and the rights of the disabled. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:5-21.
108. Frank L, Nyholm S. Robot sex and consent: is consent to sex between a robot and a human conceivable, possible, and desirable? Artif Intell Law 2017 Aug 31;25(3):305-323. [doi: 10.1007/s10506-017-9212-y]

XSL•FO
RenderX

109. Goldstein J. Was it good for you too?: the new natural law theory and the paradoxical good of sexbots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:10-36.
110. González-González C, Gil-Iranzo R, Paderewsky P. Sex with robots: analyzing the gender and ethics approaches in design. In: Abascal J, Arrue M, Gary-Vitoria N, Lozano MD, editors. Proceedings of the XX International Conference on Human Computer Interaction. New York, USA: ACM; 2019.
111. Mackenzie R. Sexbots: replacements for sex workers? Ethical constraints on the design of sentient beings for utilitarian purposes. In: Chisik Y, editor. Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference. New York, USA: ACM; 2014:1-8.
112. McArthur N. The case for sexbots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:3-18.
113. Petersen S. Is it good for them too? Ethical concern for the sexbots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:9-22.
114. Richardson K. Are sex robots as bad as killing robots? In: Inibt J, Nørskov M, Andersen SS, editors. What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016 / TRANSOR 2016. Amsterdam, The Netherlands: IOS Press; 2016:27-32.
115. Richardson K. Sex robot matters: slavery, the prostituted, and the rights of machines. IEEE Technol Soc Mag 2016 Jun;35(2):46-53. [doi: 10.1109/mts.2016.2554421]
116. Russell AC. Blurring the love lines: the legal implications of intimacy with machines. Comput Law Secur Rev 2009 Sep;25(5):455-463. [doi: 10.1016/j.clsr.2009.07.003]
117. Shen FX. Sex Robots Are Here, but Laws Aren't Keeping Up With the Ethical and Privacy Issues They Raise. The Conversation. 2019. URL: https://theconversation.com/sex-robots-are-here-but-laws-arent-keeping-up-with-the-ethical-and-privacy-issues-they-raise-109852 [accessed 2020-06-17]
118. Simmons T. Sexbots: an obloquy. Wisconsin Law Rev 2016;45(1):45-53 [FREE Full text]
119. Sparrow R. Robots, rape, and representation. Int J of Soc Robotics 2017 Jun 1;9(4):465-477. [doi: 10.1007/s12369-017-0413-z]
120. Spencer W. Digital adultery, 'meta-anon widows,' real-world divorce, and the need for a virtual sexual ethic. In: Lamers MH, Verbeek FJ, editors. Human-Robot Personal Relationships: Third International Conference. Berlin, Germany: Springer; 2011:98-107.
121. Sullins JP. Robots, love, and sex: the ethics of building a love machine. IEEE Trans Affective Comput 2012;3(4):398-409. [doi: 10.1109/t-affc.2012.31]
122. Wagner C. Sexbots: the ethical ramifications of social robotics' dark side. AI Matters 2018 Feb 16;3(4):52-58. [doi: 10.1145/3175502.3175513]
123. Welsh S. Deus Sex Machina: the Ethics of Robot Love. The Conversation. 2015. URL: https://theconversation.com/deus-sex-machina-the-ethics-of-robot-love-51049 [accessed 2020-06-17]
124. Whitby B. Do you want a robot lover?: the ethics of caring technologies. In: Lin P, Abney K, Bekey GA, editors. Robot Ethics: The Ethical and Social Implications of Robotics. Cambridge, MA: MIT Press; 2012:234-248.
125. Ziaja S. Homewrecker 2.0: an exploration of liability for heart balm torts involving AI humanoid consorts. In: Mutlu B, Bartneck C, Ham J, Evers V, Kanda T, editors. Social Robotics: Third International Conference on Social Robotics. Berlin, Germany: Springer; 2011:114-124.
126. Appel M, Marker C, Mara M. Otakuism and the appeal of sex robots. Front Psychol 2019;10:569. [doi: 10.3389/fpsyg.2019.00569] [Medline: 30984059]
127. Edirisinghe C, Cheok A. Robots and intimacies: a preliminary study of perceptions and intimacies with robots. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017:137-147.
128. Edirisinghe C, Cheok A, Khougali N. Perceptions and responsiveness to intimacy with robots: a user evaluation. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:138-157.
129. Korn O, Bieber G, Fron C. Perspectives on Social Robots: From the Historic Background to an Experts' View on Future Developments. In: Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference. 2018 Presented at: PETRA'18; June 26-29, 2018; Corfu, Greece. [doi: 10.1145/3197768.3197774]
130. Richards R, Coss C, Quinn J. Exploration of relational factors and the likelihood of a sexual robotic experience. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017:97-103.
131. Scheutz M, Arnold T. Are we ready for sex robots? In: Bartneck C, editor. Human-Robot Interaction: An Introduction. New York, USA: IEEE Press; 2016:351-358.
132. Scheutz M, Arnold T. Intimacy, bonding, and sex robots: examining empirical results and exploring ethical ramifications. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:13-17.
133. Szczuka J, Krämer N. Influences on the intention to buy a sex robot: an empirical study on influences of personality traits and personal characteristics on the intention to buy a sex robot. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017:72-83.

XSL•FO
RenderX

134. Szczuka J, Krämer N. Jealousy 4.0?: an empirical study on jealousy-related discomfort of women evoked by other women and gynoid robots. Paladyn 2018;9(1):323-336. [doi: 10.1515/pjbr-2018-0023]
135. Szczuka J, Krämer N. There's more to humanity than meets the eye: differences in gaze behavior toward women and gynoid robots. Front Psychol 2019;10:693 [FREE Full text] [doi: 10.3389/fpsyg.2019.00693] [Medline: 31068845]
136. Yulianto B, Shidarta. Philosophy of information technology: Sex robot and its ethical issues. Int J Soc Ecol Sustainable Dev 2015;6(4):67-76. [doi: 10.4018/ijsesd.2015100105]
137. Barber T. Kinky Borgs and sexy robots: the fetish, fashion and discipline of Seven of Nine. In: Geraghty L, editor. Channeling the Future: Essays on Science Fiction and Fantasy Television. Lanham, MD: Scarecrow Press; 2009:133-148.
138. Beggan JK. 2040 (Brad Armstrong, 2009). Porn Stud 2017 Sep 18;4(3):353-360. [doi: 10.1080/23268743.2017.1333019]
139. Conn V. Affective labor and technologies of gender in Wei Yahua's 'conjugal happiness in the arms of Morpheus'. In: Cheok AD, Devlin K, Levy D, editors. Love and Sex with Robots: Second International Conference. New York, USA: Springer; 2017:27-39.
140. Gevers I. Robot Love: Can We Learn from Robots about Love?. Tielt, Belgium: Lannoo Publishers; 2018.
141. Hasse C. The vitruvian robot. AI Soc 2017 Feb 28;34(1):91-93. [doi: 10.1007/s00146-017-0701-z]
142. Hauskeller M. Sex and the Posthuman Condition. London, UK: Palgrave Macmillan; 2014.
143. Hawkes R, Lacey C. 'The future of sex': intermedial desire between fembot fantasies and sexbot technologies. J Pop Cult 2019 Jan 27;52(1):98-116. [doi: 10.1111/jpcu.12748]
144. Danaher J. Robotic rape and robotic child sexual abuse: should they be criminalised? Crim Law Philos 2014 Dec 13;11(1):71-95. [doi: 10.1007/s11572-014-9362-x]
145. Danaher J. Regulating child sex robots: restriction or experimentation? Med Law Rev 2019 Nov 1;27(4):553-575. [doi: 10.1093/medlaw/fwz002] [Medline: 30938445]
146. Strikwerda L. Legal and moral implications of child sex robots. In: Danaher J, McArthur N, editors. Robot Sex: Social and Ethical Implications. Cambridge, MA: MIT Press; 2017:8-24.
147. Bendel O. SSML for sex robots. In: Cheok AD, Levy D, editors. Love and Sex with Robots: Third International Conference. New York, USA: Springer; 2018:1-11.
148. Danaher J. Building Better Sex Robots: Lessons From Feminist Pornography. PhilArchive: The Philosophy E-Print Archive. 2017. URL: https://philarchive.org/archive/DANBBSv1 [accessed 2020-06-10]
149. Summers CG. What Would Sex Robots for Women Look Like? Vice. 2016. URL: https://www.vice.com/en_uk/article/mvx44x/what-would-sex-robots-for-women-look-like [accessed 2020-06-10]
150. Golby J. We Got a Bunch of People to Draw Us Their Ideal Sex Robots. Vice. 2015. URL: https://www.vice.com/en_us/article/zngyz3/we-got-people-to-draw-their-ideal-sex-bot-205 [accessed 2020-06-10]
151. Marsh A. Love among the objectum sexuals. Electron J Hum Sex 2010;13:1 [FREE Full text]
152. Motschenbacher H. Focusing on normativity in language and sexuality studies. Crit Discourse Stud 2013 Oct 7;11(1):49-70. [doi: 10.1080/17405904.2013.836113]
153. Fren AD. Technofetishism and the uncanny desires of ASFR (alt.sex.fetish.robots). Sci-Fict Stud 2009;36:404-440 [FREE Full text]

Abbreviations

AI: artificial intelligence
BBC: British Broadcasting Corporation
CREEPER: Curbing Realistic Exploitative Electronic Pedophilic Robot
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RQ: review question

Edited by G Eysenbach; submitted 04.03.20; peer-reviewed by O Bendel; comments to author 25.03.20; revised version received 03.04.20; accepted 03.04.20; published 30.07.20.

Please cite as:
Döring N, Mohseni MR, Walter R
Design, Use, and Effects of Sex Dolls and Sex Robots: Scoping Review
J Med Internet Res 2020;22(7):e18551
URL: https://www.jmir.org/2020/7/e18551
doi:10.2196/18551
PMID:32729841

XSL•FO
RenderX

©Nicola Döring, M Rohangis Mohseni, Roberto Walter. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Computer-Controlled Virtual Humans in Patient-Facing Systems: Systematic Review and Meta-Analysis

Debaleena Chattopadhyay[1], PhD; Tengteng Ma[2], MSc; Hasti Sharifi[1], BSc; Pamela Martyn-Nemeth[3], PhD

[1]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, United States

[2]Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, United States

[3]Department of Biobehavioral Health Science, University of Illinois at Chicago, Chicago, IL, United States

Corresponding Author:
Debaleena Chattopadhyay, PhD
Department of Computer Science
University of Illinois at Chicago
851 S Morgan (M/C 152)
Room 1120 SEO
Chicago, IL, 60607
United States
Phone: 1 312413518
Email: debchatt@uic.edu

Abstract

Background: Virtual humans (VH) are computer-generated characters that appear humanlike and simulate face-to-face conversations using verbal and nonverbal cues. Unlike formless conversational agents, like smart speakers or chatbots, VH bring together the capabilities of both a conversational agent and an interactive avatar (computer-represented digital characters). Although their use in patient-facing systems has garnered substantial interest, it is unknown to what extent VH are effective in health applications.

Objective: The purpose of this review was to examine the effectiveness of VH in patient-facing systems. The design and implementation characteristics of these systems were also examined.

Methods: Electronic bibliographic databases were searched for peer-reviewed articles with relevant key terms. Studies were included in the systematic review if they designed or evaluated VH in patient-facing systems. Of the included studies, studies that used a randomized controlled trial to evaluate VH were included in the meta-analysis; they were then summarized using the PICOTS framework (population, intervention, comparison group, outcomes, time frame, setting). Summary effect sizes, using random-effects models, were calculated, and the risk of bias was assessed.

Results: Among the 8,125 unique records identified, 53 articles describing 33 unique systems, were qualitatively, systematically reviewed. Two distinct design categories emerged — simple VH and VH augmented with health sensors and trackers. Of the 53 articles, 16 (26 studies) with 44 primary and 22 secondary outcomes were included in the meta-analysis. Meta-analysis of the 44 primary outcome measures revealed a significant difference between intervention and control conditions, favoring the VH intervention (SMD = .166, 95% CI .039-.292, P=.012), but with evidence of some heterogeneity, I^2=49.3%. There were more cross-sectional (k=15) than longitudinal studies (k=11). The intervention was delivered using a personal computer in most studies (k=18), followed by a tablet (k=4), mobile kiosk (k=2), head-mounted display (k=1), and a desktop computer in a community center (k=1).

Conclusions: We offer evidence for the efficacy of VH in patient-facing systems. Considering that studies included different population and outcome types, more focused analysis is needed in the future. Future studies also need to identify what features of virtual human interventions contribute toward their effectiveness.

(J Med Internet Res 2020;22(7):e18839) doi:10.2196/18839

KEYWORDS

virtual humans; avatars; patient-facing systems; meta-analysis; conversational agents; chatbot; digital interlocutors

XSL•FO
RenderX

Introduction

Patient-facing systems are digital technologies that offer health services and engage people in their health and wellbeing [1]. These systems promote and facilitate both patient engagement with the health care system and patient empowerment in self-care [2]. Globally, one in three adults suffers from multiple chronic conditions [3]; self-management of chronic conditions has become increasingly complex, requiring sophisticated knowledge, motivation, and skill by patients and their families [4]. Engaging in self-management reduces hospitalization and improves health outcomes and quality of life [5]. Patient-facing systems can facilitate self-management of chronic conditions and fill the gap of time and distance to meet with health care providers. Data increasingly support the value of these patient-facing systems, in the form of mobile health (mHealth) applications [6-9] or sensors to monitor physiological parameters (eg, blood glucose) and behaviors (eg, activity) [10-14]. The latest technological addition to patient-facing systems is computer-controlled virtual humans (VH).

Unlike formless conversational agents, like smart speakers or chatbots, VH bring together the capabilities of both a conversational agent and an interactive avatar (computer-represented digital characters). While their humanlike physical appearance is computer-generated (ie, animated), VH are not human-controlled [15-20] nor cartoonlike [21,22]. They are controlled algorithmically, based on active or passive user input during an interaction. These algorithms can simulate key properties of human face-to-face conversation — both verbal and nonverbal (eg, gaze, emotions, head movements, and metaphoric gestures). While humanlike appearance, movements, and nonverbal behavior offer VH the added advantage of communicating empathy and building rapport, they can also bring psychological and social concerns that may not arise when interacting with formless conversational agents [23-26].

Attempts to make computer interfaces anthropomorphic are not new [27,28]. The sophistication of current digital interlocutors, however, can largely be attributed to the recent advancements in artificial intelligence and computer graphics. Recent variants of anthropomorphic interfaces include relational agents, which are agents designed to build and maintain long-term relationships with users [29], including voice-based, intelligent virtual assistants (eg, Amazon's Alexa, Microsoft's Cortana, or Apple's Siri) and text-based chatbots that run on instant messaging applications [30]. Intricate taxonomies capturing different aspects of conversational agents can be found elsewhere [26,31-33].

While intelligent virtual assistants and chatbots have gained mainstream popularity, VH applications are still in their infancy. As graphic rendering capacities progress and ubiquitous computing peripherals such as virtual reality (VR) and augmented reality head-mounted displays (HMDs) become inexpensive, VH will be increasingly adopted for everyday use. Indeed, similar to education and training [34,35], health care researchers and practitioners have already begun to explore the use of VH in health-related assessments and interventions. As VH come of age and stakeholders in health care deliberate whether to adopt this new computing technology, it is crucial that we examine how prior applications have fared in affecting health outcomes.

To our knowledge, VH in patient-facing systems have not been surveyed before. Only recently, other types of conversational agents in health have been reviewed [31,36-40]. Provoost et al [36] reported a scoping review of 54 articles (49 studies) on low-tech embodied conversational agents in clinical psychology. Laranjo et al [37] surveyed conversational agents with unconstrained natural language input capabilities in health care and included 17 articles, of which only 2 were randomized controlled trials (RCTs). Within mental health, both chatbots ([39]; 10 studies) and relational agents ([38]; 13 studies) have been reviewed. Furthermore, the current state of chatbots and embodied conversational agents as expert systems was recently surveyed ([31]; 40 articles). Kocaballi et al [40] reported a systematic review of how conversational agents can be personalized. It included 13 studies and found that personalization features were assessed for user satisfaction, not in improving health outcomes. None of these reviews included a meta-analysis nor focused on VH. The purpose of this study was to conduct a systematic review and meta-analysis of VH in patient-facing systems.

Methods

Overview

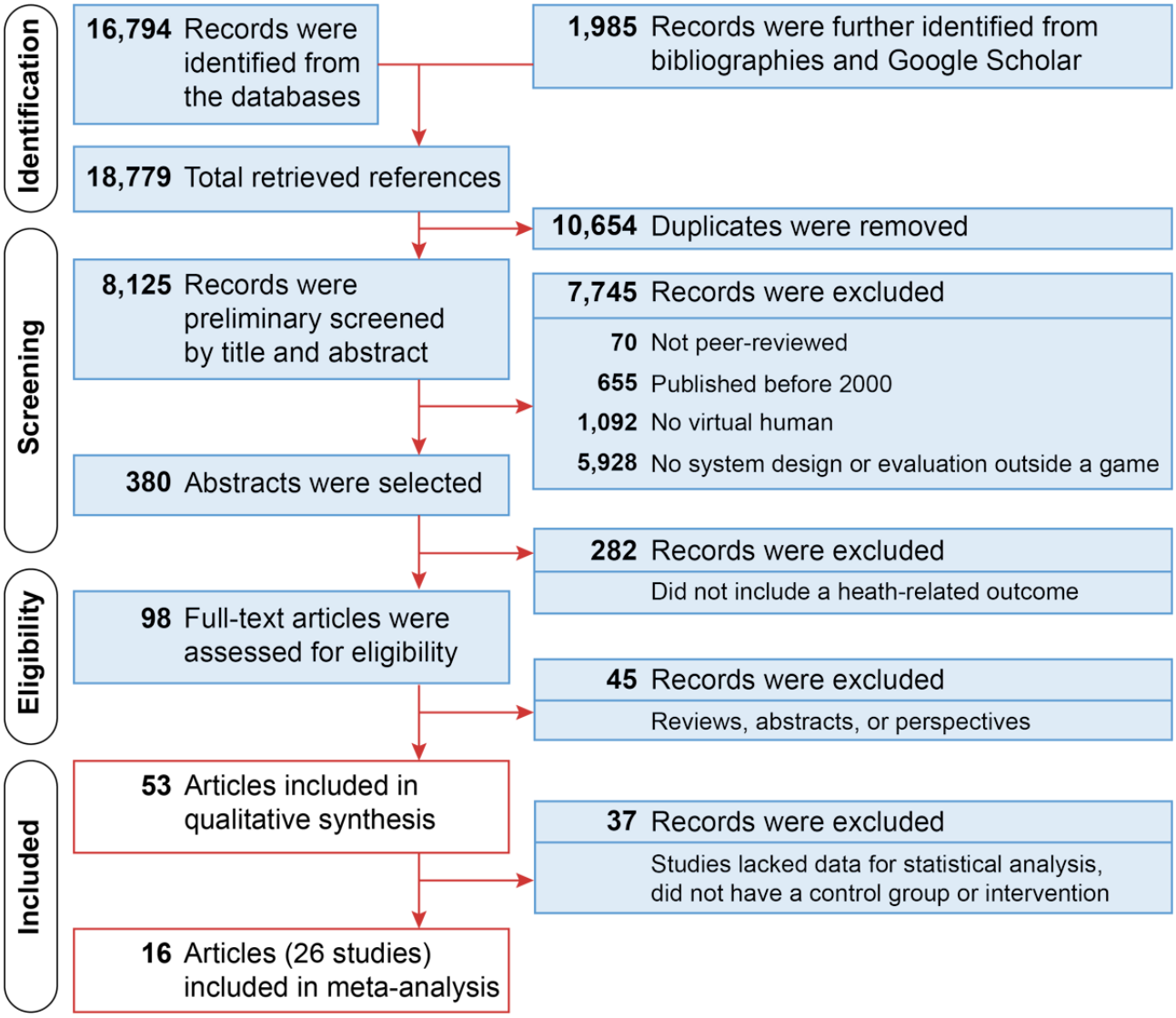
This systematic review of the English-language scholarly literature followed standard guidelines for conducting and reporting systematic reviews, including Preferred Reporting Items for Systematic Reviews and Meta-analyses [41] and guidelines from the Cochrane Diagnostic Test Accuracy Working Group [42,43].

Search Strategy

Literature searches were performed from inception to December 31, 2019 in Google Scholar and 7 online databases: MEDLINE, EMBASE, PsycINFO, CINAHL, Cochrane Central Register of Controlled Trials, PubMed, and ACM Digital Library. Search queries covered 3 domains: (1) avatars, (2) (digital) narratives, and (3) virtual assistants (for details, see Multimedia Appendix 1). Our search was limited to peer-reviewed articles published in English. Manual searches were extended to bibliographies of review articles. Figure 1 shows a summary of the literature search.

XSL•FO
RenderX

Figure 1. Summary of the literature search.



Inclusion and Exclusion Criteria

The topic of our review crosscuts two disciplines — health care and computer science — and their disciplinary priorities are divergent if not orthogonal. On the one hand, health care research prioritizes reporting and replicating empirical evidence of efficacy. On the other hand, driving innovation is a key mission of computer science research. Thus, empirical investigation of these innovative designs — particularly replication of such studies — often takes a back seat. To offer a comprehensive review of the topic at hand, we first present a qualitative, systematic review of VH in patient-facing systems. Some of those articles were then included in a quantitative meta-analysis (Figure 1).

Qualitative Systematic Review

Studies were included in the qualitative review if they met the following criterion: designed or evaluated VH for health-related outcomes in a patient-facing system. Studies were excluded if they used VH for the training and education of health care professionals or students.

Meta-Analysis

Some of the articles included in the qualitative synthesis were further included in a meta-analysis if they met the following criteria: (1) compared the effectiveness of VH in a health-related outcome in a target population against a control group with no VH; (2) studied humans of any age; and (3) reported the sample size and mean and variance of the outcome measure in control and experimental groups. Studies were excluded if they did not use a comparator that was equivalent but different from VH (eg, [44]).

Data Extraction

All records were first downloaded into an EndNote X8.2 library [45], and duplicates were removed. Titles and abstracts were then screened for inclusion and exclusion criteria. To accelerate screening, records were first collated into topics using keyword searches in the EndNote library and then reviewed for inclusion. Two investigators (the first and second authors) independently assessed the relevancy of search results and selected full text articles for further review. The second and third authors independently abstracted the key study factors into a data extraction form and then came to a consensus on which studies

XSL•FO
RenderX

met the inclusion criteria; the first author made the definitive decision when discrepancies would arise.

Data from eligible articles were extracted into a spreadsheet. For the qualitative review, data included target population, design objective, type of evaluation, principal findings, and VH characteristics. For the meta-analysis, studies were summarized using the PICOTS (population, intervention, comparison group, outcomes, time frame, setting) framework [46].

Risk of Bias and Quality Assessment

Articles included in the meta-analysis were assessed for quality and risk of bias using the latest criteria from the Cochrane Consumers and Communication Review Group [47]. Studies were categorized as low risk, unclear, or high risk for each Cochrane bias domain: selection, performance, detection, attrition, reporting, and other [47]. Publication bias was assessed using Egger's test [48] and a *P*-curve analysis [49] and graphically examined using funnel plots [50].

Statistical Analysis

All statistical analyses were performed in R using the meta [51] and metafor packages [52,53]. Standardized mean differences (SMD) were used as the effect size to quantify the effectiveness of VH. The overall effectiveness was estimated using a random-effects model, and forest plots were used to graphically present the combined effectiveness. Statistical heterogeneity among the studies was assessed by I^2, which estimates the percentage of total variation across studies due to heterogeneity rather than chance alone. We considered heterogeneity statistically significant at $P<.05$ and used random-effects models to take into account the heterogeneity among the included studies. Subgroup analysis was conducted across studies by the type of health outcome — health and wellbeing as well as attitudes toward health and wellbeing.

Results

Qualitative Systematic Review

A total of 16,794 search records were retrieved from the databases, and 1,985 additional records were identified from the bibliographies and Google Scholar. After removing duplicates, we screened the titles and abstracts of 8,125 articles; 380 articles involved a functioning VH system outside a game. Because the computing technology for creating VH did not exist prior to circa 2000, all studies published before 2000 were excluded. These 380 abstracts were further reviewed for their relevance to health, and 282 articles were excluded because the VH was involved in contexts like training, education, or demonstration. The remaining 98 articles underwent full-text review, and 53 articles met the inclusion and exclusion criteria for the qualitative systematic review.

A total of 30 health-related outcomes were identified in the 53 eligible articles, targeting 25 types of populations (Table 1) and 6 modalities of technology delivery apart from a personal computer, desktop, or laptop (Table 2). Unconstrained speech input was sparsely used in VH systems [54,55]. While most systems allowed constrained user input via a touchscreen or keyboard, some systems reacted to nonverbal conversational inputs, such as gaze [56] or proximity [57].

In the 53 eligible articles, 33 unique systems were identified. Of these, 9 systems were used for health assessments and the rest for health interventions (Table 3). Two broad design categories emerged — simple VH and VH with health sensors or trackers (Table 3). These additional trackers did not augment the interaction capabilities of VH, but provided additional data about users (eg, via a heart rhythm monitor, respiration sensor, or eye tracker).

XSL•FO
RenderX

Table 1. In the 53 eligible articles, 30 health-related outcomes and 25 target populations were identified.

Health outcome	Target population	Studies
Improve quality of life	Women with overactive bladder (OAB) symptoms	[58][a]
Self-manage chronic conditions	Individuals with chronic atrial fibrillation (heart condition)	[59-61]
	Individuals with spinal cord injury	[62]
Engage in physical activity	Older adults	[63][a] [64,65]
	Individuals with Parkinson's disease	[66]
	Inactive older adults with low socioeconomic status	[67][a]
	Healthy adults (no reported health conditions)	[68,69][a][70]
	Individuals with schizophrenia	[71]
Improve mood	Individuals with depression	[72][a] [73]
Assess auditory verbal hallucinations (AVH)	Individuals with schizophrenia	[74]
Stress management	Women	[75]
	Individuals with chronic pain and depression	[76][a] [77]
Healthy eating	Women	[75]
	Healthy adults (no reported health conditions)	[69][a]
Improve social skills	Children with autism spectrum disorders (ASD)	[44]
	Individuals with schizophrenia	[78][a]
Assess PTSD[b] symptoms	US military service members	[54][a] [55]
Assess body image disturbance (BID)	Women on diet (nonclinical)	[79]
Anxiety toward death	Older adults	[80]
Find health-related information online	Individuals with low health and computer literacy	[81][a] [82]
Explain health documents	Individuals with low health literacy	[83][a] [65,84-86]
Attitude toward regular physical activity	Healthy adults (no reported health conditions)	[87][a]
Attitude toward breastfeeding	Pregnant women in their third semester	[88][a] [89,90]
Attitude toward weight loss	Healthy adults (no reported health conditions)	[91][a]
Retention of medication knowledge	Individuals with type 2 diabetes mellitus	[92][a]
Attitudes toward prenatal testing for Down syndrome	Nulliparous women	[93]
Improve medication adherence	Individuals with schizophrenia	[71]
Assess emotion recognition	Adults with ASD	[94]
	Individuals with schizophrenia	[95][a]
	Children with ASD	[96]
Preconception risk assessment	Women	[97]
Assess the effects of social rejection	Individuals with psychotic disorder	[98]
Assess social attention	Children with ASD	[99]
Assist in deep breathing	Healthy adults (no reported health conditions)	[100]
Substance use counseling	Individuals with alcohol use disorder	[56]
	Individuals with opioid use disorder	[101]
Patient trust	Healthy adults (no reported health conditions)	[102]
Assess social anxiety disorder	Women with high social anxiety	[103]

XSL•FO
RenderX

Health outcome	Target population	Studies
Alleviate social isolation	Older adults	[57]
Understand the distinction between connective and fatty tissue in the breast	Mammography-eligible middle-aged women (40-74 years old)	[104]
Pill count adherence >80%	HIV-positive African American men who have sex with men	[105]

[a]Studies included in the meta-analysis.

[b]PTSD: post-traumatic stress disorder.

Table 2. Technology characteristics identified in the eligible studies.

Technology characteristics	Studies
Unconstrained speech input	[54,55]
Computer at a community center or school	[44,67]
Smartphone	[59-61]
Head-mounted display (HMD)	[74,78,79,98,99,103]
Virtual reality (VR) in a PC or HMD	[74,78,79,96,98,99,103]
Mobile kiosk with a computer	[63,65,84-86,88]
Tablet	[63,65,66,70,76,77,80,88]

Table 3. Two broad categories of virtual humans emerged from the 53 articles included in the qualitative review.

Type of use	Number of simple virtual humans	Number of virtual humans with health trackers
Intervention	34 [44,57,58,62-64,67-69,71,72,75-78,80-93,97,101,102,104,105]	9 [56,59-61,65,66,70,73,100]
Assessment	7 [54,74,79,94-96,98]	3 [55,99,103]

Virtual Humans for Health Interventions

Of the 34 articles that described simple VH in health-related interventions (Table 3), 27 were based on the same core technology for generating speech [106] and nonverbal behavior [107]. The typical system included a knowledge base of domain-specific top-level dialog fragments (Figure 2). These subdialogs were scripted and then reused to generate natural language speech using a hierarchical transition network, based on augmented transition networks. Augmented transition networks are mathematical structures that can model the grammar of relatively complex natural languages [108]. The generated text was then converted to synthetic speech by the Behavior Expression Animation Toolkit [107]. The Behavior Expression Animation Toolkit also generated appropriate and contextualized nonverbal behavior for the VH. The front end consisted of 3 components: (1) an animated image of the VH communicating with users using speech and gestures, (2) a dynamically updated but constrained multiple-choice menu for user input (via touch or keyboard), and (3) other content supporting the conversation as necessary, like text and images. VH in such a system talked to users with synthetic speech and presented synchronized nonverbal behavior, such as hand gestures and eyebrow raises for emphasis, looking away to signal turn-taking, and posture shifts to mark topic boundaries.

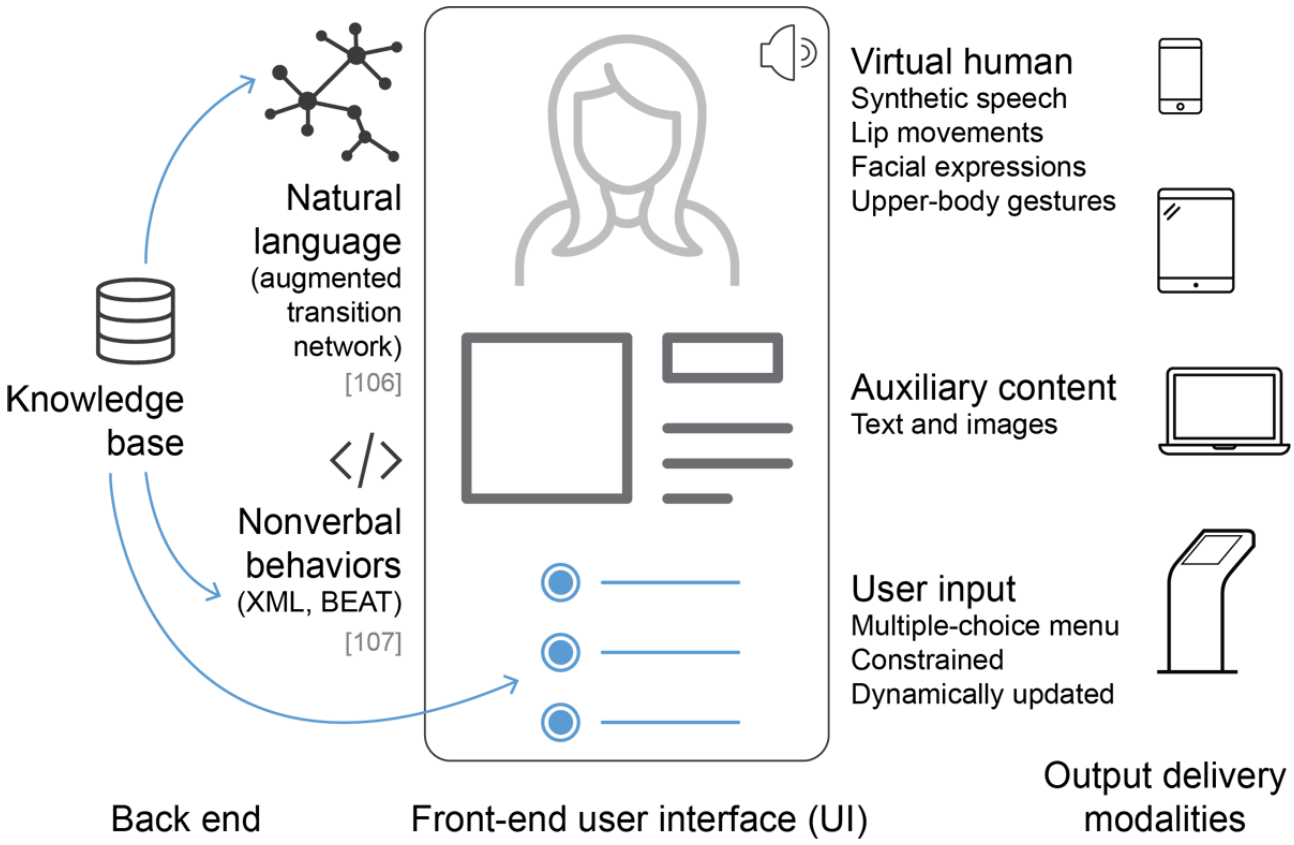
The VH interface in Figure 2 was mostly delivered via smartphones or personal computers. By tailoring the knowledge base, VH were designed for a range of domains, from managing depression [77], through end-of-life planning [80], stress management [75], and educating about breastfeeding [89], to preconception care [97].

Design approaches were varied in the rest of the studies. Dworkin et al [105] followed the same concepts as in Figure 2 but did not discuss implementation details. Two studies used a third-party system [44,72]. Another used VR and an HMD to deliver the VH intervention [78]. Some studies did not detail exactly how the conversations between humans and VH were automated [58,68,78,92], other than mentioning that nonverbal behaviors were programmed to align with verbal utterances. Two studies modeled their intervention after tutorials, with functionality, such as play, pause, repeat, and answering multiple-choice questions [58,92].

Some designs augmented VH with sensor-based tracking (Table 3), such as heart rhythm monitors for managing chronic heart failure [59-61], a pedometer for promoting walking [65,66], and a breathing monitor for assisting in meditation [100].

XSL•FO
RenderX

Figure 2. The most common structure of a simple virtual human system designed for health-related interventions. BEAT: Behavior Expression Animation Toolkit.



Virtual Humans for Health Assessments

Compared with interventions, fewer studies were found for health-related assessments (Tables 1 and 3); most of these were for people with autism spectrum disorder [94,96,99] and psychotic disorders [74,98,103]. Other studies used VH to assess PTSD symptoms [54,55], emotion recognition [95], and body image disorder [79]. Six studies used VH in a VR environment; 5 used HMDs [74,79,98,99,103], while 1 used a computer [96]. Two studies used advanced tracking systems; these tracking modules were more sophisticated than those used in health interventions. To assess social attention in high-functioning children with autism spectrum disorder, researchers used a sensor to measure head orientation and rotational motion [99]. The MultiSense framework was used for the SimSensei Kiosk [55], which tracks an array of perceptual signals, like smile intensity, gaze direction, and lack of facial expressions, and allows unconstrained speech input [54].

Virtual Human Characteristics

Overall, the physical appearances of VH were primarily created using 3-dimensional (3D) character modeling and animation software, such as the Unity3D game engine. They were designed to be racially ambiguous [64] or accordant with target users [86]. But Marcos-Pablos et al [95] generated dynamic expressions for VH (eg, anger, happiness, sadness) by first using a laser scanner to obtain a 3D face model and then using animation software. The speech of VH was either synthetically produced [76] or lip-synced to audio-recorded voice narrations [58,92]. Synthetic speech was generated using commercially available software.

Characteristics, personalities, and mannerisms of VH were manipulated to build rapport with end users [102]. Schulman et al [87] explored the use of social dialog (small talk) and persuasive arguments (about 30 turns) to change attitude toward exercise. VH argued about the advantages of regular exercise and against statements that emphasized disadvantages. Friederichs et al [68] used social dialog at the beginning of an intervention and during transitions between different parts. To persuade users, Andrade et al [58] developed a peer avatar — based on the front-face view of the participant — who delivered a tutorial about self-managing overactive bladder symptoms. Motivational interviewing techniques were used by Friederichs et al [68] to promote physical activity in adults.

Another study explored the use of personal stories available on the internet to personalize the VH's message and change health behavior [91]. These stories were indexed based on participants's personal state of change during weight loss. To promote breastfeeding, lactation education was framed from a feminist perspective [90]; a VH was designed to motivate women in their third semester to follow breastfeeding recommendations by the US Centers for Disease Control and Prevention. This was realized by introducing a feminist introductory and closing script (eg, "I like to think of myself as a fairly progressive woman, and I hope I do not offend you by some of my opinions" or "A lot of people think that breastfeeding represents a dilemma for feminism.") [90]. Participants who did not self-identify as feminists were significantly less satisfied with the feminist agent than the controls.

XSL•FO
RenderX

Middle-aged Caucasian and African American VH were designed to achieve racial concordance with users [92], but racial concordance did not significantly affect perceived similarity [86]. Researchers also investigated the effects of attire, background image, and alignment on trust. Study participants found a patient-aligned VH explaining an informed consent document more trustworthy than a medical-aligned or federal-aligned VH. Deictic gestures, such as pointing to a document with a finger or open hand, were designed to aid in document explanation [85]. When explaining a medical document, study participants with low literacy were more satisfied with such a VH than a human. Finally, a common theme across many VH designed for health-related interventions was continuity of care, or longitudinal engagement with end users.

Design Process

Of the 53 articles, 23 explicitly described their VH design process [55,58,59,68,71,73,75,78,80,82-87,89-93,95,97,104]. The most common design approaches were collaboration with domain experts, qualitative observation of similar human-human interactions, and adoption of public health and governmental guidelines. Some studies conducted focus groups and one-to-one interviews with target users [59,68,90,92,97,104]. Others informed their design decisions from prior literature or theoretical frameworks [58,59,68,71,90,91,93]. Qualitative observations were often videotaped and coded to generate conversation content [55,71,84]. VH for health-related assessments were always designed in collaboration with health care providers but not end users (Table 3).

Theory

Only a few papers explicitly mentioned adopting theoretical frameworks to ground their design of VH [56,60,68,70,91,93,100]. The chronic care model was used to design VH for people with chronic illnesses [109]; the following 3 guidelines were adopted: facilitate communication between patients and providers, make patients aware of the latest care guidelines consistent with scientific evidence, and motivate patients to manage their health [60].

Two frameworks of behavior change were used widely — the transtheoretical model (TTM) of health behavior change [110] and motivational interviewing (MI) counseling style [111]. The transtheoretical model operationalizes intentional behavior change with 3 core constructs — stages of change (precontemplation, contemplation, preparation, action, and maintenance), processes of change, and levels of change — and recommends stage-matched interventions. VH could ask questions to determine these stages and offer matched interventions, such as therapeutic alliance [56,70,91]. MI was developed to elicit people's readiness to change, explore their attitudes toward change, and transition ambivalence toward reducing resistance to behavior change. The conversation style of VH was sometimes designed based on the MI technique [56,68]. One study designed VH to mirror basic user actions [112], like inhaling and exhaling, to promote behavior change [100]. In another study [93], VH played the role of a decision coach to facilitate shared decision making [113,114]. They informed users of all available options, provided detailed information about those options, and finally assisted in choosing one of them.

Design Guidelines

Four papers explicitly offered design guidelines for VH [57,60,64,71]. To design VH that can maintain long-term social relationships with end users (ie, relational agents), Bickmore et al [64] recommended developing a model of user-agent relationship; using relational behavior like social dialog, empathy, and humor; maintaining a memory of past interactions; and providing some variability in agent behavior and overall variability in the dialog structure. When designing VH for mental health interventions that are also relational agents, special design considerations may be needed. For patients with schizophrenia, Bickmore et al [71] recommended prolonging the introduction and conclusion phases of the conversation, using clear and concrete language that focuses on real events to reinforce reality, and not relying solely on nonverbal behaviors like a gaze-away gesture. Specific populations, health outcomes, and device modalities may need additional design considerations. When addressing loneliness in older adults, Ring et al [57] recommended VH engage in social interactions, such as small talk and games, help older adults stay connected with friends and family via contemporary technologies, and alleviate mood disorders or loneliness through different types of talk therapy. When designing for smartphones, Bickmore et al [60] recommended keeping interactions short, allowing interruptions during interactions, and using constrained user input.

Meta-Analysis

A total of 26 studies (16 articles) published between 2000 and December 31, 2019 were eligible for the meta-analysis, targeting 11 types of populations and including 10 studies with healthy adults [68,69,81,83,87], 3 studies with older adults [63,67], 3 studies with women [58,88], 2 studies with individuals with schizophrenia [78,95], and 3 studies with individuals with depression [72,76].

Study Characteristics

The included studies comprised approximately 1400 participants across 13 health and wellbeing objectives. The PICOTS information [46] from the 26 studies is available in Multimedia Appendix 2. Outcomes included 44 primary outcomes and 22 secondary outcomes. Of the 26 studies, 9 focused on attitudes toward health and wellbeing [81,83,87,88,91]. There were more cross-sectional (k=15) [54,68,81,83,87,88,91,92,95] than longitudinal studies (k=11) [58,63,67,69,72,76,78,88]. Longitudinal studies ranged from 1 month to 6 months (see Multimedia Appendix 2). The VH intervention was delivered using a personal computer in most studies (k=18), followed by a tablet (k=4), mobile kiosk (k=2), HMD (k=1), and a desktop computer in a community center (k=1).

Evidence of Intervention Efficacy

As evident from Table 2 and Multimedia Appendix 2, the 26 studies eligible for meta-analysis varied in terms of outcomes, target population, timing, and intervention design. Hence, a random-effects model was used. A random-effects model does not assume that the estimated effects come from a single homogeneous population, but that true effect sizes vary from

XSL•FO
RenderX

study to study. Hedges' *g* was calculated for each reported outcome, and SMD were used as the effect size to quantify the overall evidence for VH interventions. Next, we report the meta-analysis of the 44 primary outcomes across 26 studies. A meta-analysis of all 66 outcomes (including 22 secondary outcomes) is available in Multimedia Appendix 2.

The meta-analysis of data from 26 studies (44 outcomes) revealed a significant difference between intervention and control conditions, favoring the VH intervention (SMD .166, 95% CI .039-.292, 95% prediction interval –.548 to .879, *P*=.012) but with evidence of some heterogeneity: I^2 = 49.3%, 95% CI 28.1%-64.3% (Figure 3).

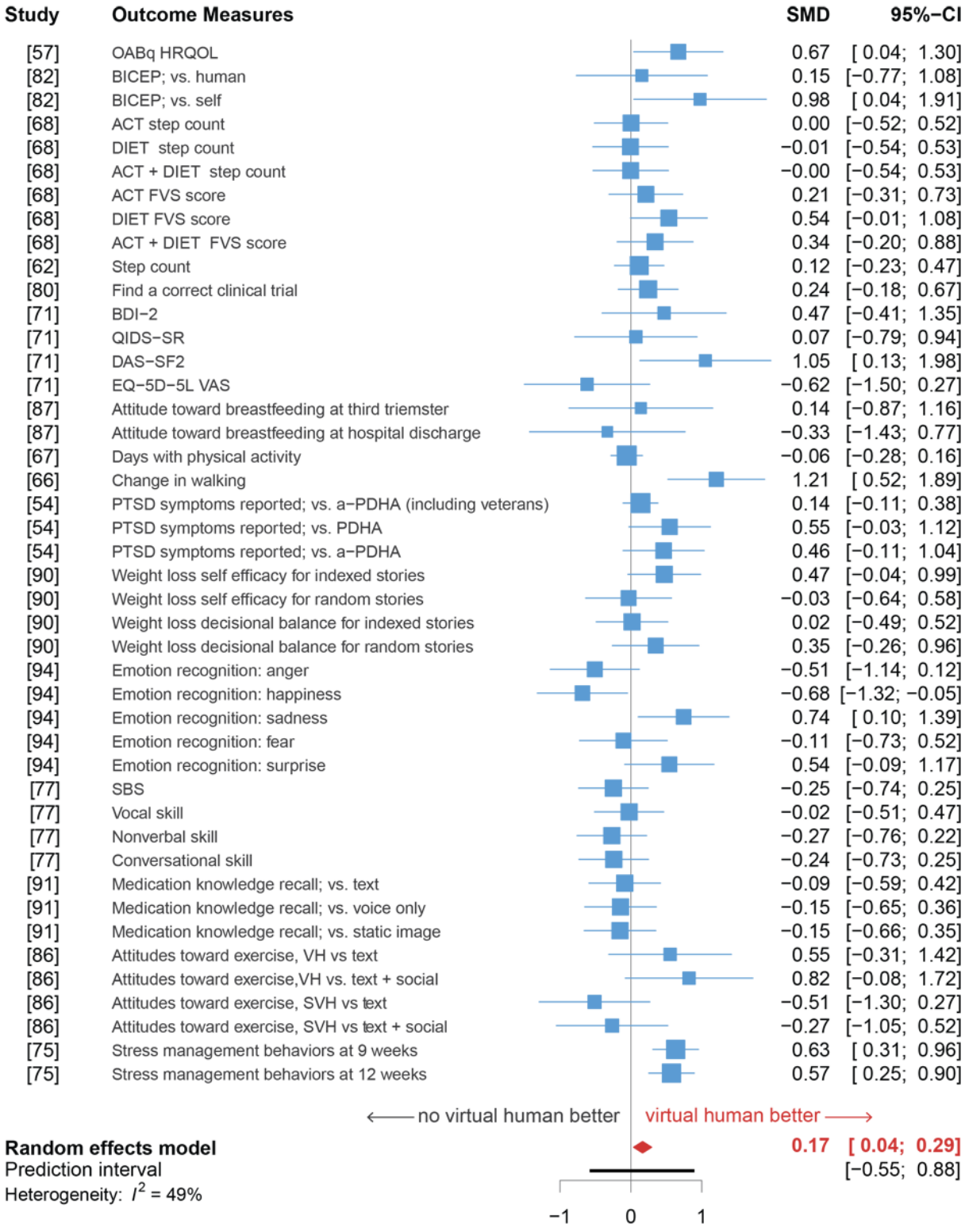
A 3-level model (level 2: different outcome measures; level 3: different studies) did not capture a significant amount of variability in the data (*P*>.05). Thus, a 2-level model was used. The between-study heterogeneity of the data was moderate: τ^2=.12, I^2=49.3%. We examined whether this heterogeneity was caused by outliers or influential cases [115,116]. No influential cases were detected, but we spotted 3 outliers [67,76,95]. While 2 of these outlier studies found extremely positive evidence about VH efficacy [67,76], 1 study found extremely negative evidence for one of its outcome measures: happiness emotion recognition by patients with schizophrenia [95]. However, the meta-analysis after removing these outliers (*k*=41, SMD .144, 95% CI .028-.260, *P*=.016; I^2=33.7%, 95% CI 2.6%-54.8%) was similar to the original analysis (Figure 3). Because these 3 studies were interesting outliers, but neither error nor influential outliers [117], we chose to retain them.

A subgroup analysis for health-related outcomes and health-related attitudes was conducted, but no significant difference was found in the overall effect between outcome types (*P*=.762). The number of studies was insufficient to conduct subgroup analyses for different population or outcome types.

To explore publication bias, a funnel plot was generated. Egger's test was not significant (*P*=.70), indicating no substantial asymmetry in the funnel plot. Furthermore, the *P*-curve analysis did not indicate publication bias (Figure 4).

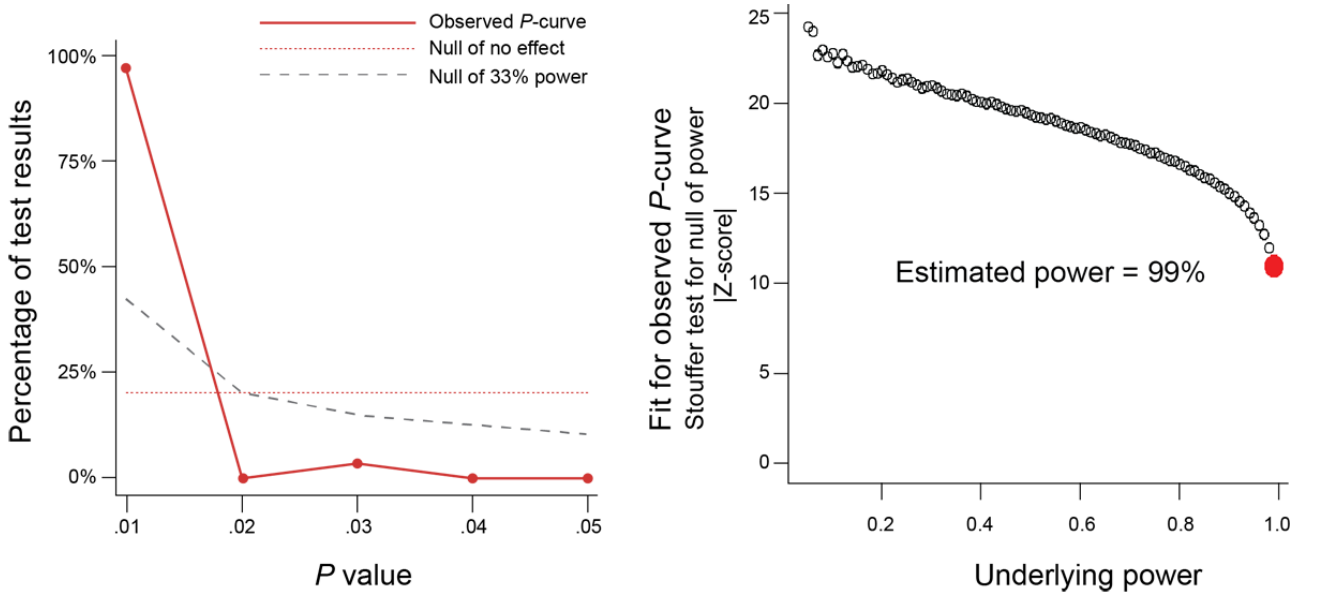
XSL•FO
RenderX

Figure 3. Forest plot of the meta-analysis of health-related virtual human interventions from 26 studies (44 primary outcomes). a-PDHA: anonymized post-deployment health assessment; ACT: physical activity; BDI-2: Beck Depression Inventory-II; BICEP: brief informed consent evaluation protocol; DAS−SF2: Dysfunctional Attitude Scale-Short Form 2; DIET: fruit and vegetable consumption; EQ−5D−5L VAS: 5-level version of the EuroQol 5D visual analogue scale; FVS: NIH/NCI Fruit and Vegetable Scan; HRQOL: health-related quality of life; OABq: overactive bladder questionnaire; PDHA: post-deployment health assessment; PTSD: post-traumatic stress disorder; QIDS−SR: Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form; SBS: social behavior scales; SMD: standardized mean difference; SVH: social virtual human.



XSL•FO
RenderX

Figure 4. The observed *P*-curve has an estimated power of 99% (left and right), significant right skewness, $P_{full} < .0001$, $P_{half} < .0001$ (left), and no significant flatness, $P_{full} > .9999$, $P_{half} > .9999$.



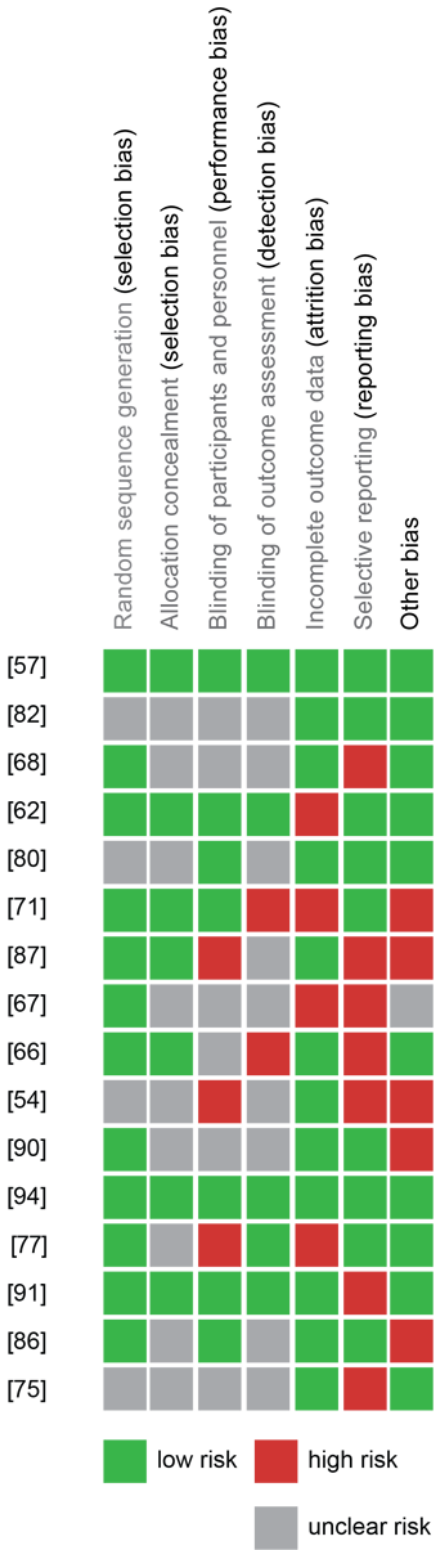
Risk of Bias in Included Studies

The quality of studies included in the meta-analysis was evaluated for risk of bias (Figures 5 and 6). Randomization was adequate in 19 studies [58,63,67-69,72,78,87,88,91,92,95] but unclear in 7 others [54,76,81,83]. Allocation was reported to be concealed in 11 studies [58,63,67,72,88,92,95], but other studies did not provide enough information to assess allocation bias. It was unclear in 15 studies whether participants and research personnel were blinded to the allocated interventions [54,68,69,76,78,81,83,87,91].

Eleven studies were successful in blinding participants and research personnel to the allocated interventions [58,63,72,81,87,92,95], while 5 studies were not [54,78,88]; the remaining studies were unclear. Blinding the assessors of outcomes was achieved in 8 studies [58,63,78,92,95], 2 studies did not blind research personnel [67,72], and 16 did not clearly report this [54,68,69,76,81,83,87,88,91]. Attrition was high (>20%) in 5 studies [63,68,72,78]. With respect to selective reporting, we identified 13 studies that did not report either descriptive statistics for nonsignificant outcomes or participant demographics [54,67-69,76,88,92]. In assessing other potential sources of bias, we identified 11 studies at a risk of bias due to a small sample size [87,88] or self-selection from provider invitations [72] and websites like Craigslist or Amazon Mechanical Turk [54,69,87,91].

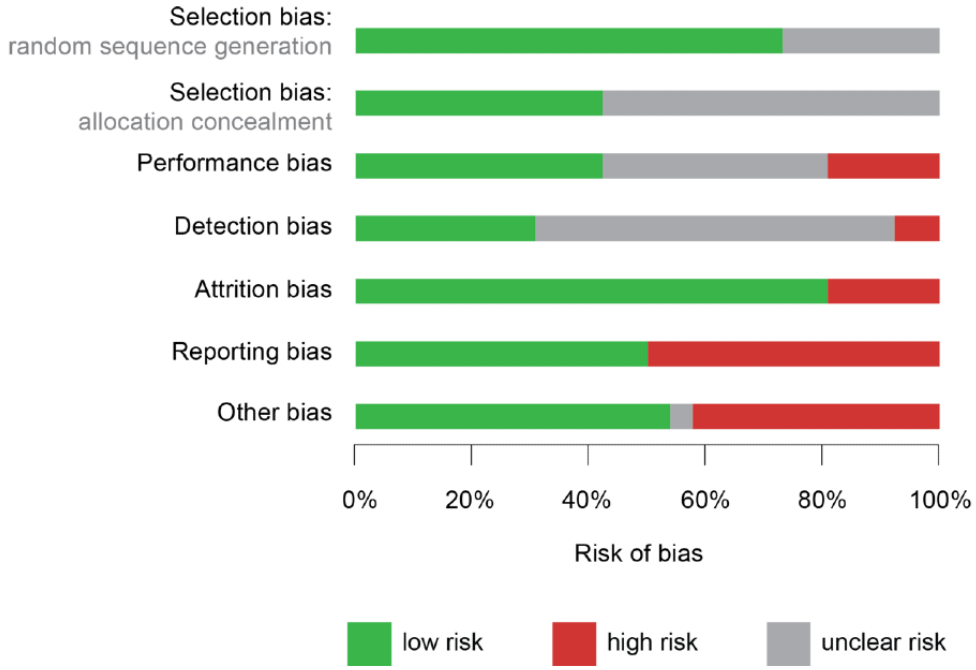
XSL•FO
RenderX

Figure 5. Summary of the authors' consensus judgment about the risk of bias for each study included in the meta-analysis, by various sources of potential bias.



XSL•FO
RenderX

Figure 6. Risk of bias presented as percentages across all 26 studies included in the meta-analysis.



Discussion

Primary Findings

This meta-analysis found VH interventions significantly more effective than other types of traditional interventions that did not include conversational agents (SMD .166, 95% CI=.039-.292, *P*=.012). This is an important finding because effective interaction modalities are needed to promote improved consumer or patient engagement in health, thereby promoting behavior change and the management of chronic conditions. The effect size may have been small because the studies involved many different population and outcome types. As more studies become available in the future, some populations and outcome types may show stronger effect sizes than others.

The efficacy of VH may also depend on the type of intervention for which they are used, such as delivering cognitive behavioral therapy vs delivering education. However, the number of current studies was insufficient to conduct such subgroup analyses. Nevertheless, the effectiveness of VH did not significantly differ between health-related outcomes and health-related attitudes.

Like prior reports [37], we found considerable heterogeneity in evaluation methods and outcome measures and a predominance of quasiexperimental study designs or usability studies over RCTs. But unlike previous reviews [36], we found most applications were fully developed, and many of them were either evaluated or soon to be evaluated in RCTs. Our findings cannot be compared directly with the previous reviews because each of the reviews used a different definition of conversational agents or avatars. Typically, earlier reviews did not discriminate between conversational agents with and without a humanlike physical appearance; therefore, they did not discriminate between whether they were able or unable to engage in nonverbal conversation. Only 5 of the 26 studies that were included in the meta-analysis compared the efficacy of VH interventions with text-based chatbots or any other type of conversational agents (Multimedia Appendix 2) [58,68,87,91,92]. Other studies compared interventions against treatment as usual, such as a therapist visit or standard educational materials. Voice-based and text-based conversational agents, without a physical humanlike appearance, were used as comparators in 2 studies [58,92] and 4 studies [68,87,91,92], respectively. Thus, it is too early to tell how much the efficacy of VH in health-related interventions can be attributed to their humanlike physical appearance.

Of course, some health applications, especially some mental health assessments, would only work with VH and could not be replaced with a voice-based or text-based conversational agent. For example, VH were used to assess the influence of unusual voices on daily activities of hallucinating patients [74] and the effect of social rejection in individuals with psychotic disorder [98]. But apart from these, when should a VH intervention be more desirable than other types of conversational agents? Engineering health-related VH interventions is far from trivial and may further cause unintended effects, such as a lack of empathy or a sense of complacency among users [23,24]. When is it worth investing in designing VH over chatbots or smart speakers? This question kept resurfacing in our analysis but remained unanswered. As more studies become available, subgroup analyses of specific interventions could answer this question.

Currently, the prevalence of VH in health applications appears to lag behind that in other areas, such as education and training [118,119] or serious games [120], which is not surprising given the concerns about patient safety [121], ethical and legal issues [122], and perception of professional counselors [123]. Technological advancements will continue to augment VH capabilities; equally important is identifying the design tradeoffs associated with those capabilities in patient-facing systems. When should a conversational agent have a humanlike appearance or any physical embodiment? Would allowing for unconstrained speech and gesture input improve intervention

XSL•FO
RenderX

effectiveness or reduce patient safety and privacy? When are nonverbal cues significant in designing a conversational agent? Does age, gender, socioeconomic status, literacy levels, or certain characteristics of target users mediate the effectiveness of VH interventions?

Finally, we found that the input and output of VH systems have evolved significantly over the last 2 decades, drawing on the most recent technological advancements. While systems in the 2000s extensively used desktops, kiosks, 2-dimensional graphics, and constrained text input [64], more recent systems were developed in 3D, were delivered in VR, sometimes used HMDs, and allowed for unconstrained input, such as speech, gestures, and facial emotions. Two key design trends were identified: (1) multimodal sensing of the user's state using computer vision algorithms and ubiquitous computing technologies, ranging from upper-body gestures to heart rate (input), and (2) striving toward high fidelity, humanlike appearance and behavior of the VH to improve presence and immersion (output).

Limitations and Future Research

The limitations of this review should be noted and can be addressed by future studies. First, not all studies on VH in patient-facing systems were included in our work. This is because they did not present sufficient quantitative information, only reported usability metrics, or did not clarify whether their avatar technology was computer-controlled or human-controlled. Including additional studies and VH designs could allow reinforcing the results reported here or provide different results. Second, some of the studies included in our meta-analysis had relatively small sample sizes (<20 participants); thus, additional caution is recommended when generalizing these results. Third, there was moderate heterogeneity among trials in the meta-analysis (τ^2=.12, I^2=49.3%.). This can be attributed to the different health outcome measures, population types, and health objectives. For example, health objectives ranged from increasing physical activity and improving mood to improving social skills. When more studies become available with the same or similar health objectives, it would be worth updating this study with new results.

Although research on conversational agents began circa 2000, their design and capabilities have changed and diverged substantially as new technologies and sensors have emerged. This change is expected to continue. Future studies are suggested to consider the difference between different types of conversational agents when synthesizing or generalizing the agents. For example, does a physical appearance or nonverbal behavior increase the effectiveness of a conversational agent? In what kind of tasks? Furthermore, there is rich literature on behavior change and health behavior change theories. However, theoretical frameworks explicating how different features of VH work together in building (or disrupting) rapport with patients is lacking. As such models emerge, future studies will need to examine those relationships between model constructs with methods such as meta-analytic structural equation models.

Conclusion

VH are conversational agents with a humanlike physical appearance; autonomy in verbal and nonverbal behavior; and speech, gaze, or gesture interaction capabilities. In patient-facing systems, they can demonstrate listening and empathy, as well as tailor to various sociocultural backgrounds, languages, and literacy levels. We surveyed the existing literature on VH in patient-facing systems — from inception to December 2019. Of the 53 articles reviewed, a meta-analysis of 26 studies with more than 1400 participants showed that VH interventions significantly improve health outcomes compared with other traditional intervention methods. But whether a physical embodiment is crucial for a conversational agent to significantly improve health-related outcomes remains to be explored, as does any effect of the VH's physical appearance, type of voice, or quality of movements.

Although not yet comparable to computer-animated films or high-end video games, the appearance and behavior of VH in health care are increasingly becoming sophisticated, with studies finding that users prefer more humanlike VHs in health care [124,125]. Elsewhere, studies continue to report the possibility of unintended negative user reactions (ie, the uncanny valley effect) when interacting with VH, owing to a mismatch in the levels of realism, either physical, behavioral, or both [23,24,126]. However, the literature on VH in patient-facing systems has not yet examined whether the uncanny valley effect affects patient perception and, in turn, health outcomes.

Acknowledgments

This article's publication was partially funded by the Research Open Access Publishing (ROAAP) Fund of the University of Illinois at Chicago (UIC), administered by the UIC library.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search details.
[PDF File (Adobe PDF File), 94 KB - jmir_v22i7e18839_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

PICOTS information.
[PDF File (Adobe PDF File), 1419 KB - jmir_v22i7e18839_app2.pdf]

References

1. Jensen RE, Gummerson SP, Chung AE. Overview of Patient-Facing Systems in Patient-Reported Outcomes Collection: Focus and Design in Cancer Care. JOP 2016 Oct;12(10):873-875. [doi: 10.1200/jop.2016.015685]
2. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
3. Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. Prev Med Rep 2018 Dec;12:284-293 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
4. Schulman-Green D, Jaser SS, Park C, Whittemore R. A metasynthesis of factors affecting self-management of chronic illness. J Adv Nurs 2016 Jul 19;72(7):1469-1489 [FREE Full text] [doi: 10.1111/jan.12902] [Medline: 26781649]
5. Riegel B, Lee CS, Albert N, Lennie T, Chung M, Song EK, et al. From Novice to Expert. Nursing Research 2011;60(2):132-138. [doi: 10.1097/nnr.0b013e31820978ec]
6. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e136 [FREE Full text] [doi: 10.2196/mhealth.6445] [Medline: 27993761]
7. Wakefield BJ, Turvey CL, Nazi KM, Holman JE, Hogan TP, Shimada SL, et al. Psychometric Properties of Patient-Facing eHealth Evaluation Measures: Systematic Review and Analysis. J Med Internet Res 2017 Oct 11;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
8. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
9. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
10. Baron KG, Duffecy J, Richardson D, Avery E, Rothschild S, Lane J. Technology Assisted Behavior Intervention to Extend Sleep Among Adults With Short Sleep Duration and Prehypertension/Stage 1 Hypertension: A Randomized Pilot Feasibility Study. J Clin Sleep Med 2019 Nov 15;15(11):1587-1597. [doi: 10.5664/jcsm.8018] [Medline: 31739848]
11. Garg SK, Shah VN, Akturk HK, Beatson C, Snell-Bergeon JK. Role of Mobile Technology to Improve Diabetes Care in Adults with Type 1 Diabetes: The Remote-T1D Study iBGStar in Type 1 Diabetes Management. Diabetes Ther 2017 Aug 29;8(4):811-819 [FREE Full text] [doi: 10.1007/s13300-017-0272-5] [Medline: 28555339]
12. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a Tablet Computer on Self-care, Quality of Life, and Knowledge. The Journal of Cardiovascular Nursing 2018;33(4):336-343. [doi: 10.1097/jcn.0000000000000462]
13. Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
14. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
15. Fox J, Bailenson J, Binney J. Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. Presence: Teleoperators and Virtual Environments 2009 Aug;18(4):294-303. [doi: 10.1162/pres.18.4.294]
16. Magnenat-Thalmann N, Kasap Z. Virtual Humans in Serious Games. 2009 Presented at: International Conference on CyberWorlds; 2009; Bradford, UK. [doi: 10.1109/cw.2009.17]
17. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: A pilot clinical trial. Schizophr Res 2018 Jul;197:176-181. [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
18. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. The Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/s2215-0366(17)30427-3]
19. Leff J, Williams G, Huckvale MA, Arbuthnot M, Leff AP. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry 2013 Jun 02;202(6):428-433. [doi: 10.1192/bjp.bp.112.124883] [Medline: 23429202]
20. Fabio RA, Caprì T, Iannizzotto G, Nucita A, Mohammadhasani N. Interactive Avatar Boosts the Performances of Children with Attention Deficit Hyperactivity Disorder in Dynamic Measures of Intelligence. Cyberpsychol Behav Soc Netw 2019 Sep 01;22(9):588-596. [doi: 10.1089/cyber.2018.0711] [Medline: 31441667]
21. Ogle AD, Graham DJ, Lucas-Thompson RG, Roberto CA. Influence of Cartoon Media Characters on Children's Attention to and Preference for Food and Beverage Products. J Acad Nutr Diet 2017 Feb;117(2):265-270.e2 [FREE Full text] [doi: 10.1016/j.jand.2016.08.012] [Medline: 27793520]

XSL•FO
RenderX

22. Sinor MZ. Comparison between conventional health promotion and use of cartoon animation in delivering oral health education. Intl J Humanities and Social Science 2011;1(3):169-174.
23. Chattopadhyay D, MacDorman KF. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. J Vis 2016 Sep 01;16(11):7 [FREE Full text] [doi: 10.1167/16.11.7] [Medline: 27611063]
24. Dai Z, MacDorman K. The doctor's digital double: how warmth, competence, and animation promote adherence intention. PeerJ Computer Science 2018;4:e168. [doi: 10.7717/peerj-cs.168]
25. Zhou MX, Mark G, Li J, Yang H. Trusting Virtual Agents. ACM Trans. Interact. Intell. Syst 2019 Apr 25;9(2-3):1-36. [doi: 10.1145/3232077]
26. Feine J, Gnewuch U, Morana S, Maedche A. A Taxonomy of Social Cues for Conversational Agents. International Journal of Human-Computer Studies 2019 Dec;132:138-161. [doi: 10.1016/j.ijhcs.2019.07.009]
27. Weizenbaum J. ELIZA---a computer program for the study of natural language communication between man and machine. Commun. ACM 1966;9(1):36-45. [doi: 10.1145/365153.365168]
28. Cassell J. Embodied conversational agents: representation and intelligence in user interfaces. AI magazine 2001;22(4):67-67. [doi: 10.7551/mitpress/2697.001.0001]
29. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Hum. Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
30. Hill J, Randolph Ford W, Farreras IG. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior 2015 Aug;49:245-250. [doi: 10.1016/j.chb.2015.02.026]
31. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
32. Gnewuch U, Morana S, Heckmann C, Maedche A. Designing Conversational Agents for Energy Feedback. 2018 Presented at: International Conference on Design Science Research in Information Systems and Technology; 2018; Chennai, India p. 18-33. [doi: 10.1007/978-3-319-91800-6_2]
33. Diederich S, Brendel AB, Kolbe LM. Designing Anthropomorphic Enterprise Conversational Agents. Bus Inf Syst Eng 2020 Mar 10;62(3):193-209. [doi: 10.1007/s12599-020-00639-y]
34. Ieronutti L, Chittaro L. Employing virtual humans for education and training in X3D/VRML worlds. Computers & Education 2007 Aug;49(1):93-109. [doi: 10.1016/j.compedu.2005.06.007]
35. Kleinsmith A, Rivera-Gutierrez D, Finney G, Cendan J, Lok B. Understanding Empathy Training with Virtual Patients. Comput Human Behav 2015 Nov 01;52:151-158 [FREE Full text] [doi: 10.1016/j.chb.2015.05.033] [Medline: 26166942]
36. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
37. Laranjo L, Dunn A, Tong H, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
38. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
39. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul 21;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
40. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
42. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. Syst Rev 2013 Oct 07;2(1):82 [FREE Full text] [doi: 10.1186/2046-4053-2-82] [Medline: 24099098]
43. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0. Hoboken, New Jersey: The Cochrane Collaboration; 2010.
44. Hopkins IM, Gower MW, Perez TA, Smith DS, Amthor FR, Casey Wimsatt F, et al. Avatar Assistant: Improving Social Skills in Students with an ASD Through a Computer-Based Intervention. J Autism Dev Disord 2011 Feb 2;41(11):1543-1555. [doi: 10.1007/s10803-011-1179-z] [Medline: 21287255]
45. EndNote®. URL: https://endnote.com/ [accessed 2020-07-08]
46. Riva JJ, Malik KMP, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. J Can Chiropr Assoc 2012 Sep;56(3):167-171 [FREE Full text] [Medline: 22997465]

XSL•FO
RenderX

47. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
48. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006 Feb 08;295(6):676-680. [doi: 10.1001/jama.295.6.676] [Medline: 16467236]
49. Simonsohn U, Nelson LD, Simmons JP. p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. Perspect Psychol Sci 2014 Nov 17;9(6):666-681. [doi: 10.1177/1745691614553988] [Medline: 26186117]
50. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
51. meta: General Package for Meta-Analysis. 2012. URL: https://cran.r-project.org/web/packages/meta/index.html [accessed 2020-07-06]
52. metafor: Meta-Analysis Package for R. URL: https://cran.r-project.org/web/packages/metafor/index.html [accessed 2020-07-07]
53. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J statistical software 2010 Aug 05;36(3):1-48.
54. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. Front. Robot. AI 2017 Oct 12;4:1. [doi: 10.3389/frobt.2017.00051]
55. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, et al. SimSensei kiosk: a virtual human interviewer for healthcare decision support. 2014 Presented at: International conference on Autonomous Agents and Multi-Agent Systems; 2014; Paris, France p. 1061-1068. [doi: 10.5555/2615731.2617415]
56. Zhou S, Bickmore T, Rubin A, Yeksigian C, Sawdy M, Simon S. User Gaze Behavior while Discussing Substance Use with a Virtual Agent. 2018 Presented at: International Conference on Intelligent Virtual Agents; 2018; Sydney NSW Australia p. 353-354. [doi: 10.1145/3267851.3267906]
57. Ring L, Barry B, Totzke K, Bickmore T. Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support. : IEEE; 2013 Presented at: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; 2013; Geneva, Switzerland. [doi: 10.1109/ACII.2013.17]
58. Andrade AD, Anam R, Karanam C, Downey P, Ruiz JG. An overactive bladder online self-management program with embedded avatars: a randomized controlled trial of efficacy. Urology 2015 Mar;85(3):561-567. [doi: 10.1016/j.urology.2014.11.017] [Medline: 25586473]
59. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The Atrial Fibrillation Health Literacy Information Technology System: Pilot Assessment. JMIR Cardio 2017 Dec 12;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
60. Bickmore T, Kimani E, Trinh H, Pusateri A, Paasche-Orlow M, Magnani J. Managing Chronic Conditions with a Smartphone-based Conversational Virtual Agent. 2018 Presented at: International Conference on Intelligent Virtual Agents; 2018; Sydney NSW Australia p. 119-124. [doi: 10.1145/3267851.3267908]
61. Kimani E, Bickmore T, Trinh H, Ring L, Paasche-Orlow M, Magnani J. A Smartphone-Based Virtual Agent for Atrial Fibrillation Education and Counseling. 2016 Presented at: International Conference on Intelligent Virtual Agents; 2016; Los Angeles, USA p. 120-127. [doi: 10.1007/978-3-319-47665-0_11]
62. Shamekhi A, Trinh H, Bickmore T, DeAngelis T, Ellis T, Houlihan B, et al. A Virtual Self-care Coach for Individuals with Spinal Cord Injury. 2016 Presented at: 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '16; 2016; Reno Nevada USA p. 327-328. [doi: 10.1145/2982142.2982199]
63. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct 03;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
64. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interacting with Computers 2005 Dec;17(6):711-735. [doi: 10.1016/j.intcom.2005.09.002]
65. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010 Aug 31;15 Suppl 2(sup2):197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
66. Ellis T, Latham NK, DeAngelis TR, Thomas CA, Saint-Hilaire M, Bickmore TW. Feasibility of a Virtual Exercise Coach to Promote Walking in Community-Dwelling Persons with Parkinson Disease. American Journal of Physical Medicine & Rehabilitation 2013;92(6):472-485. [doi: 10.1097/phm.0b013e31828cd466]
67. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013 Dec;18(12):1449-1464 [FREE Full text] [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
68. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
69. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]

XSL•FO
RenderX

70. Bickmore TW, Mauer D, Brown T. Context Awareness in a Handheld Exercise Agent. Pervasive Mob Comput 2009 Jun 01;5(3):226-235 [FREE Full text] [doi: 10.1016/j.pmcj.2008.05.004] [Medline: 20161031]
71. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: Relational agents for antipsychotic medication adherence. Interacting with Computers 2010 Jul;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]
72. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, et al. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Jul 09;22(6):348-355. [doi: 10.1177/1357633x15609793]
73. Ring L, Bickmore T, Pedrelli P. An affectively aware virtual therapist for depression counseling. 2016 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI) workshop on Computing and Mental Health; 2016; Denver Colorado USA.
74. Han K, Heo J, Seo S, Hong M, Lee JS, Shin YS, et al. The effect of simulated auditory hallucinations on daily activities in schizophrenia patients. Psychopathology 2012 Jul 31;45(6):352-360 [FREE Full text] [doi: 10.1159/000337264] [Medline: 22854179]
75. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
76. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting Group Medical Visits with Conversational Agents for Stress Management Behavior Change. Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviorspringer International Publishing; 2017 Presented at: International Conference on Persuasive Technology; 2017; Amsterdam, The Netherlands p. 55-67. [doi: 10.1007/978-3-319-55134-0_5]
77. Shamekhi A, Bickmore T, Lestoquoy A, Negash L, Gardiner P. Blissful Agents: Adjuncts to Group Medical Visits for Chronic Pain and Depression. Blissful Agents: Adjuncts to Group Medical Visits for Chronic Pain and Depression. In. Intelligent Virtual Agentspringer International Publishing; 2016 Presented at: International Conference on Intelligent Virtual Agents; 2016; Los Angeles, USA p. 433-437. [doi: 10.1007/978-3-319-47665-0_49]
78. Park K, Ku J, Choi S, Jang H, Park J, Kim SI, et al. A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. Psychiatry Res 2011 Sep 30;189(2):166-172. [doi: 10.1016/j.psychres.2011.04.003] [Medline: 21529970]
79. Mountford VA, Tchanturia K, Valmaggia L. "What Are You Thinking When You Look at Me?" A Pilot Study of the Use of Virtual Reality in Body Image. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):93-99. [doi: 10.1089/cyber.2015.0169] [Medline: 26378613]
80. Utami D, Bickmore T, Nikolopoulou A, Paasche-Orlow M. Talk About Death: End of Life Planning with a Virtual Agent. 2017 Presented at: International Conference on Intelligent Virtual Agents; 2017; Stockholm, Sweden p. 441-450. [doi: 10.1007/978-3-319-67401-8_55]
81. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
82. Utami D, Barry B, Bickmore T, Paasche-Orlow M. A conversational agent-based clinical trial search engine. 2013 Presented at: Annual Symposium on Human-Computer Interaction and Information Retrieval; 2013; Vancouver, BC, Canada.
83. Bickmore TW, Pfeifer LM, Paasche-Orlow MK. Using computer agents to explain medical documents to patients with low health literacy. Patient Educ Couns 2009 Jun;75(3):315-320 [FREE Full text] [doi: 10.1016/j.pec.2009.02.007] [Medline: 19297116]
84. Bickmore T, Pfeifer L, Jack B. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. 2009 Presented at: SIGCHI conference on human factors in computing systems; 2009; Boston MA USA p. 1265-1274. [doi: 10.1145/1518701.1518891]
85. Bickmore T, Pfeifer L, Paasche-Orlow M. Health Document Explanation by Virtual Agents. 2007 Presented at: International Workshop on Intelligent Virtual Agents; 2007; Paris, France p. 183-196. [doi: 10.1007/978-3-540-74997-4_18]
86. Zhou S, Bickmore T, Paasche-Orlow M, Jack B. Agent-User Concordance and Satisfaction with a Virtual Hospital Discharge Nurse. 2014 Presented at: International Conference on Intelligent Virtual Agents; 2014; Boston, MA, USA p. 528-541. [doi: 10.1007/978-3-319-09767-1_63]
87. Schulman D, Bickmore T. Persuading users through counseling dialogue with a conversational agent. 2009 Presented at: International Conference on Persuasive Technology; 2009; Claremont California USA p. 1-8. [doi: 10.1145/1541948.1541983]
88. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec 18;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
89. Zhang Z, Bickmore T, Mainello K, Mueller M, Foley M, Jenkins L, et al. Maintaining Continuity in Longitudinal, Multi-method Health Interventions Using Virtual Agents: The Case of Breastfeeding Promotion. 2014 Presented at: International Conference on Intelligent Virtual Agents; 2014; Boston, MA, USA p. 504-513. [doi: 10.1007/978-3-319-09767-1_61]
90. Shi L, Bickmore T, Edwards R. A Feminist Virtual Agent for Breastfeeding Promotion. 2015 Presented at: International Conference on Intelligent Virtual Agents; 2015; Delft, The Netherlands p. 461-470. [doi: 10.1007/978-3-319-21996-7_49]

XSL•FO
RenderX

91. Manuvinakurike R, Velicer WF, Bickmore TW. Automated indexing of Internet stories for health behavior change: weight loss attitude pilot study. J Med Internet Res 2014 Dec 09;16(12):e285 [FREE Full text] [doi: 10.2196/jmir.3702] [Medline: 25491243]
92. Ruiz JG, Andrade AD, Anam R, Lisigurski M, Karanam C, Sharit J. Computer-based programmed instruction did not improve the knowledge retention of medication instructions of individuals with type 2 diabetes mellitus. Diabetes Educ 2014 Oct 29;40(1):77-88. [doi: 10.1177/0145721713508825] [Medline: 24168837]
93. Zhang Z, Bickmore T. Medical Shared Decision Making with a Virtual Agent. 2018 Presented at: International Conference on Intelligent Virtual Agents; 2018; Sydney NSW Australia p. 113-118. [doi: 10.1145/3267851.3267883]
94. Cassidy S, Stenger B, Van Dongen L, Yanagisawa K, Anderson R, Wan V, et al. Expressive visual text-to-speech as an assistive technology for individuals with autism spectrum conditions. Comput Vis Image Underst 2016 Jul;148:193-200 [FREE Full text] [doi: 10.1016/j.cviu.2015.08.011] [Medline: 27375348]
95. Marcos-Pablos S, González-Pablos E, Martín-Lorenzo C, Flores LA, Gómez-García-Bermejo J, Zalama E. Corrigendum: Virtual Avatar for Emotion Recognition in Patients with Schizophrenia: A Pilot Study. Front Hum Neurosci 2016 Nov 02;10:554 [FREE Full text] [doi: 10.3389/fnhum.2016.00554] [Medline: 27807413]
96. Kim K, Rosenthal MZ, Gwaltney M, Jarrold W, Hatt N, McIntyre N, et al. A Virtual Joy-Stick Study of Emotional Responses and Social Motivation in Children with Autism Spectrum Disorder. J Autism Dev Disord 2015 Dec 17;45(12):3891-3899 [FREE Full text] [doi: 10.1007/s10803-014-2036-7] [Medline: 24435405]
97. Gardiner P, Hempstead MB, Ring L, Bickmore T, Yinusa-Nyahkoon L, Tran H, et al. Reaching Women through Health Information Technology: The Gabby Preconception Care System. Am J Health Promot 2013 Jan 02;27(3_suppl):eS11-eS20. [doi: 10.4278/ajhp.1200113-quan-18]
98. Hesse K, Schroeder PA, Scheeff J, Klingberg S, Plewnia C. Experimental variation of social stress in virtual reality - Feasibility and first results in patients with psychotic disorders. J Behav Ther Exp Psychiatry 2017 Sep;56:129-136. [doi: 10.1016/j.jbtep.2016.11.006] [Medline: 27939053]
99. Jarrold W, Mundy P, Gwaltney M, Bailenson J, Hatt N, McIntyre N, et al. Social attention in a virtual public speaking task in higher functioning children with autism. Autism Res 2013 Oct 20;6(5):393-410 [FREE Full text] [doi: 10.1002/aur.1302] [Medline: 23696132]
100. Shamekhi A, Bickmore T. Breathe Deep: A Breath-Sensitive Interactive Meditation Coach. 2018 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; 2018; New York NY USA. [doi: 10.1145/3240925.3240940]
101. Olafsson S, O'Leary T, Bickmore T. Coerced Change-talk with Conversational Agents Promotes Confidence in Behavior Change. 2019 Presented at: International Conference on Pervasive Computing Technologies for Healthcare; 2019; Trento, Italy p. 31-40. [doi: 10.1145/3329189.3329202]
102. Zhang Z, Bickmore TW, Paasche-Orlow MK. Perceived organizational affiliation and its effects on patient trust: Role modeling with embodied conversational agents. Patient Educ Couns 2017 Sep;100(9):1730-1737. [doi: 10.1016/j.pec.2017.03.017] [Medline: 28381330]
103. Wieser MJ, Pauli P, Grosseibl M, Molzow I, Mühlberger A. Virtual social interactions in social anxiety--the impact of sex, gaze, and interpersonal distance. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):547-554. [doi: 10.1089/cyber.2009.0432] [Medline: 20950179]
104. Gunn C, Maschke A, Bickmore T, Kennedy M, Hopkins MF, Fishman MD, et al. Acceptability of an Interactive Computer-Animated Agent to Promote Patient-Provider Communication About Breast Density: a Mixed Method Pilot Study. J Gen Intern Med 2020 Apr 9;35(4):1069-1077. [doi: 10.1007/s11606-019-05622-2] [Medline: 31919723]
105. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, Feasibility, and Preliminary Efficacy of a Theory-Based Relational Embodied Conversational Agent Mobile Phone Intervention to Promote HIV Medication Adherence in Young HIV-Positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
106. Bickmore T. Relational agents: Effecting change through human-computer relationships. Massachusetts, USA: Doctoral dissertation, Massachusetts Institute of Technology; 2003.
107. Cassell J, Vilhjálmsson H, Bickmore T. BEAT: the Behavior Expression Animation Toolkit. 2001 Presented at: International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH; 2001; Los Angeles, USA p. 477-486. [doi: 10.1007/978-3-662-08373-4_8]
108. Woods WA. Transition network grammars for natural language analysis. Commun. ACM 1970;13(10):591-606. [doi: 10.1145/355598.362773]
109. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 2002 Oct 16;288(15):1909-1914. [doi: 10.1001/jama.288.15.1909] [Medline: 12377092]
110. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Aug 26;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
111. Rollnick S, Miller WR. What is Motivational Interviewing? Behav. Cogn. Psychother 2009 Jun 16;23(4):325-334. [doi: 10.1017/s135246580001643x]

XSL•FO
RenderX

112. Soller A, Martínez A, Jermann P, Muehlenbrock M. From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. Intl J Artificial Intelligence in Education 2005;15(4):261-290.
113. Stacey D, Murray MA, Légaré F, Sandy D, Menard P, O'Connor A. Decision coaching to support shared decision making: a framework, evidence, and implications for nursing practice, education, and policy. Worldviews Evid Based Nurs 2008 Mar;5(1):25-35. [doi: 10.1111/j.1741-6787.2007.00108.x] [Medline: 18266768]
114. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct 23;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
115. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr 04;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
116. Baujat B, Mahé C, Pignon J, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
117. Aguinis H, Gottfredson RK, Joo H. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. Organizational Research Methods 2013 Jan 14;16(2):270-301. [doi: 10.1177/1094428112470848]
118. Gratch J, DeVault D, Lucas G. The Benefits of Virtual Humans for Teaching Negotiation. 2016 Presented at: International Conference on Intelligent Virtual Agents; 2016; Los Angeles, USA p. 283-294. [doi: 10.1007/978-3-319-47665-0_25]
119. Carnell S, Lok B, James M, Su J. Predicting Student Success in Communication Skills Learning Scenarios with Virtual Humans. 2019 Presented at: International Conference on Learning Analytics & Knowledge; 2019; Tempe, AZ, USA p. 436-440. [doi: 10.1145/3303772.3303828]
120. Feng Z, González VA, Amor R, Lovreglio R, Cabrera-Guerrero G. Immersive virtual reality serious games for evacuation training and research: A systematic literature review. Computers & Education 2018 Dec;127:252-266. [doi: 10.1016/j.compedu.2018.09.002]
121. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
122. Yellowlees PM, Holloway KM, Parish MB. Therapy in virtual environments--clinical and ethical issues. Telemed J E Health 2012 Sep;18(7):558-564. [doi: 10.1089/tmj.2011.0195] [Medline: 22823138]
123. Witt KJ, Oliver M, McNichols C. Counseling via Avatar: Professional Practice in Virtual Worlds. Int J Adv Counselling 2016 Jul 2;38(3):218-236. [doi: 10.1007/s10447-016-9269-4]
124. Parmar D, Olafsson S, Utami D, Bickmore T. Looking the part: The effect of attire and setting on perceptions of a virtual health counselor. 2018 Presented at: International Conference on Intelligent Virtual Agents; 2018; Sydney, NSW, Australia p. 301-306. [doi: 10.1145/3267851.3267915]
125. Ring L, Utami D, Bickmore T. The right agent for the job? 2014 Presented at: International Conference on Intelligent Virtual Agents; 2014; Boston, MA, USA p. 374-384. [doi: 10.1007/978-3-319-09767-1_49]
126. Meah L, Moore R. The Uncanny Valley: A Focus on Misaligned Cues. 2014 Presented at: International Conference on Social Robotics; 2014; Sydney, NSW, Australia p. 256-265. [doi: 10.1007/978-3-319-11973-1_26]

Abbreviations

3D: 3-dimensional.
a-PDHA: anonymized post-deployment health assessment.
ACT: physical activity.
ASD: autism spectrum disorders.
ATN: augmented transition network.
BDI-2: Beck Depression Inventory-II.
BEAT: Behavior Expression Animation Toolkit.
BICEP: brief informed consent evaluation protocol.
DAS-SF2: Dysfunctional Attitude Scale-Short Form 2.
DIET: fruit and vegetable consumption.
EQ−5D−5L VAS: 5-level version of the EuroQol 5D visual analogue scale.
FVS: NIH/NCI Fruit and Vegetable Scan.
HMD: head-mounted display.
HRQOL: health-related quality of life.
mHealth: mobile health.
MI: motivational interviewing.
OABq: overactive bladder questionnaire.
PDHA: post-deployment health assessment.
PTSD: post-traumatic stress disorder.
QIDS-SR: Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form.

XSL•FO
RenderX

RCT: randomized controlled trial.
SBS: social behavior scales.
SMD: standardized mean difference.
SVH: social virtual human.
VH: virtual human.
VR: virtual reality.

Edited by G Eysenbach; submitted 23.03.20; peer-reviewed by E Bellei, KM Kuo; comments to author 20.04.20; revised version received 08.05.20; accepted 20.05.20; published 30.07.20.

Please cite as:
Chattopadhyay D, Ma T, Sharifi H, Martyn-Nemeth P
Computer-Controlled Virtual Humans in Patient-Facing Systems: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(7):e18839
URL: http://www.jmir.org/2020/7/e18839/
doi:10.2196/18839
PMID:32729837

©Debaleena Chattopadhyay, Tengteng Ma, Hasti Sharifi, Pamela Martyn-Nemeth. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Technology-Enabled Knowledge Translation Strategies in Improving the Use of Research in Public Health: Systematic Review

Alison Brown[1,2,3,4], BNutrDiet; Courtney Barnes[1,2,3,4], BNutrDiet; Judith Byaruhanga[1,2,3,4], MPH; Matthew McLaughlin[1,2,3,4], BSc; Rebecca K Hodder[1,2,3,4], PhD; Debbie Booth[5], MAppSci; Nicole Nathan[1,2,3,4], PhD; Rachel Sutherland[1,2,3,4], PhD; Luke Wolfenden[1,2,3,4], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

[4]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

[5]University Library, Academic Division, University of Newcastle, Callaghan, Australia

Corresponding Author:
Alison Brown, BNutrDiet
Hunter New England Population Health
Booth Building
Longworth Avenue
Wallsend, 2287
Australia
Phone: 61 249246685
Email: alison.brown7@health.nsw.gov.au

Abstract

Background: Knowledge translation (KT) aims to facilitate the use of research evidence in decision making. Changes in technology have provided considerable opportunities for KT strategies to improve access and use of evidence in decision making by public health policy makers and practitioners. Despite this opportunity, there have been no reviews that have assessed the effects of digital technology-enabled KT (TEKT) in the field of public health.

Objective: This study aims to examine the effectiveness of digital TEKT strategies in (1) improving the capacity for evidence-based decision making by public health policy makers and practitioners, (2) changing public health policy or practice, and (3) changes in individual or population health outcomes.

Methods: A search strategy was developed to identify randomized trials assessing the effectiveness of digital TEKT strategies in public health. Any primary research study with a randomized trial design was eligible. Searches for eligible studies were undertaken in multiple electronic bibliographic databases (Medical Literature Analysis and Retrieval System Online [MEDLINE], Excerpta Medica dataBASE [EMBASE], PsycINFO, Cumulative Index to Nursing and Allied Health Literature [CINAHL], and Scopus) and the reference lists of included studies. A hand search of 2 journals (Implementation Science and Journal of Medical Internet Research) and a gray literature search were also conducted. Pairs of independent review authors screened studies, assessed the risk of bias, and extracted data from relevant studies.

Results: Of the 6819 citations screened, 8 eligible randomized trials were included in the review. The studies examined the impact of digital TEKT strategies on health professionals, including nurses, child care health consultants, physiotherapists, primary health care workers, and public health practitioners. Overall, 5 of the interventions were web-training programs. The remaining 3 interventions included simulation games, access to digital resource materials and the use of tailored messaging, and a web-based registry. The findings suggest that digital TEKT interventions may be effective in improving the knowledge of public health professionals, relative to control, and may be as effective as a face-to-face KT approach. The effectiveness of digital TEKT strategies relative to a control or other digital KT interventions on measures of health professional self-efficacy to use evidence to enhance practice behavior or behavioral intention outcomes was mixed. The evidence regarding the effects on changes to health policy or practice following exposure to digital TEKT was mixed. No trials assessed the effects on individual or population-level health outcomes.

XSL•FO
RenderX

Conclusions: This review is the first to synthesize the effectiveness of digital TEKT interventions in a public health setting. Despite its potential, relatively few trials have been undertaken to investigate the impacts of digital TEKT interventions. The findings suggest that although a digital TEKT intervention may improve knowledge, the effects of such interventions on other outcomes are equivocal.

(J Med Internet Res 2020;22(7):e17274) doi:10.2196/17274

KEYWORDS

knowledge translation; public health

Introduction

Background

Investment in public health research is intended to improve public health policy, practice, and decision making [1]. To enhance the impact of public health services, public health care decision making based on high-quality research evidence is recommended [1]. Currently, research evidence is often not utilized optimally in public health decision making [2-4], increasing the risk that public health policies or practices may be enacted that are inferior to alternative interventions, that have little evidence of benefit, or that may be harmful [2,3].

Knowledge translation (KT) is "the synthesis, dissemination, exchange, and application of knowledge in an effort to improve health services and products and strengthen the health care system" [5]. The process of KT aims to bridge the gap between research evidence generated by researchers and the use of this evidence by health professionals in their decision making regarding the adoption of health policies, practices, or programs [3,6]. Conceptual frameworks suggest that KT is a dynamic process that is inherently linked to the engagement of end users, including patients, policy makers, and health care professionals, to enhance the uptake of research via the creation, dissemination, and use of knowledge (research) [3]. An important part of the KT processes is the development of knowledge tools, products, or other strategies (*interventions*) that are accessible and interactive and meet the needs of stakeholders for informed decision making [3,4,7].

A variety of factors impede KT in public health. A 2018 critical interpretive synthesis for KT in public health in low- and middle-income countries highlighted the complex nature of creating and accessing evidence for the end user and the contextual factors, including cultural, political, and economic factors that influence the ability to inform evidence-based decision making [8]. A number of KT strategies have previously been applied to address some of these barriers. A systematic review of 5 randomized and non-randomized studies evaluating the effectiveness of public health KT [9] found that single KT strategies, such as disseminating materials to health professionals, were as effective as complex, multicomponent interventions, such as interventions with multiple face-to-face contacts, in changing the practice behavior of public health professionals [9]. However, some KT strategies, such as access to web-based registries to find evidence, did not significantly impact decision making in public health professionals [9].

Changes in technology over recent decades have provided considerable opportunities to improve access and use of evidence in decision making [4]. Technology-enabled KT (TEKT) is the incorporation of digital technology in the application of KT [10]. Digital TEKT does this by using digital technologies, for example, via the use of social media, email, internet, electronic databases, electronic prompts or reminders, web-based webinars, and training or interactive websites. Such strategies may include *push* strategies, whereby research is disseminated (eg, via social media) to target end-user audiences to increase its awareness and use of *pull* strategies that aim to increase the target end user's demand and use for research (eg, webinars to improve research literacy) [2,11].

Despite the opportunity that digital TEKT presents in facilitating KT, there are few reviews assessing the impact of digital TEKT on decision making in health. To the best of our knowledge, there have been no reviews that have assessed the effects of digital TEKT in the field of public health. This is important given the contextual differences in public health and clinical practice decision making. Nonetheless, reviews of their impact in clinical settings suggest that they can be beneficial. For example, a 2016 systematic review by De Angelis et al [12] examined the impact of information and communication technologies in the dissemination of clinical practice guidelines to health professionals. The review included 21 studies in which dissemination occurred via a variety of communication technologies, including computer software, web-based workshops, educational games, and email and assessed the skills, knowledge, intention of the health professionals, or perceived usefulness of the intervention [12]. The review found that website and computer software dissemination of practice guidelines showed little evidence in improving practice behavior [12]. Conversely, web-based workshops and emails were found to improve practice behavior by improving the skills and knowledge of clinical practice guidelines [12]. Similarly, a 2019 systematic review highlighted that the most effective KT strategies for health professionals in child health settings were those relating to web-based education and computerized prompts and reminders [13]. The existing literature highlights the promise of the use of digital TEKT; however, the effectiveness of digital TEKT in the public health setting is not known.

Objectives

In the context of existing evidence synthesis gaps for public health digital TEKT, this review aimed to examine the effectiveness of digital TEKT strategies in improving the following measures:

1. Public health policy makers or practitioners' capacity to make evidence-based health policy and practice decisions such as changes in knowledge acquisition, knowledge

XSL•FO
RenderX

retention, change in reasoning, judgment, or decision making [14].
2. Evidence-based public health policy and practice such as changes in behavior, public health policy, or practice.
3. Individual- or population-level health outcomes from the use of research in public health.

Methods

Registration

This review was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42018112715) and is reported per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [15].

Eligibility Criteria

Types of Studies

Any primary research study with a randomized trial design, including cluster randomized, was eligible.

Studies published in peer-reviewed journals, as well as gray literature publications, were included in this review. There was no restriction on the length of the study follow-up period, the language of publication, or country of origin.

Types of Participants

Studies were eligible for inclusion if they included public health end users, such as health care policy makers, health care managers, and health professionals employed by public health services. For this review, public health services were defined as aiming to promote health and well-being and prevent illness and disease [16]. Such services may be delivered by government or nongovernment organizations. Examples of public health services include health protection and health promotion of preventable diseases and illnesses such as childhood obesity, injury prevention, vaccinations, and immunizations [9,16].

Health professionals who are involved in public health may have included, but was not limited to, health practitioners; allied health professionals such as dietitians, occupational therapists, physiotherapists, speech pathologists; pharmacists; nurses; physicians; and social workers with a focus on preventative care. Studies targeting clinicians were only included if they were engaging in preventative health services. Studies that assessed the impact on researchers embedded within public health services were excluded from this review.

Types of Interventions

Any study that reported on the effectiveness of digital TEKT strategies targeting public health policy makers or practitioners was eligible for inclusion that compared the following:

1. Digital TEKT strategies targeting public health policy makers and practitioners with no intervention.
2. Two or more alternative KT strategies targeting public health policy makers and practitioners, where at least one alternative included a digital TEKT strategy only.

Digital TEKT could make use of social media, email, internet, electronic databases, electronic prompts or reminders, web-based webinars, or training or interactive websites to facilitate research use by end users. Digital TEKT strategies could employ either or both *push* and *pull* strategies. Examples of digital TEKT strategies included in this review form part of the knowledge creation component (knowledge tools and products) of the knowledge to action framework [3].

Studies in which digital TEKT was not the exclusive component of a study were excluded.

Comparison

Groups may have received any alternative KT strategy, usual care, no intervention, or a waitlist control.

Types of Outcome Measures

We included any trial that included the assessment of the effects of digital TEKT strategies on the following:

1. Measures of public health end users' capacity to make evidence-informed health policy and practice decisions. Any cognitive measures of end users' capacity for evidence-informed decisions were included, such as measures of knowledge acquisition, knowledge retention, change in reasoning, judgment, or decision making [14]. In addition, measures of change in intention, attitude, and self-efficacy were included. Such data could be collected via surveys of policy makers and practitioners, completion of performance tasks, observations, or other measures.
2. Measures of evidence-based public health policy and practice, including changes in the behavior of policy makers or practitioners in decision making, in actual public health policy, or practice. Measures could include surveys of policy makers and practitioners, practice reviews, and assessment of core competencies or observations.
3. Measures of individual-, community-, or population-level health outcomes. These could include measures of the presence or absence of disease, health condition, or behavioral risk factors (eg, tobacco use) at the level of an individual, collected via patient surveys, use of medical records, objective measures of behavior or disease, or any other method to determine health outcomes; or at a population level, for example, the use of population-level surveys or disease surveillance systems or change in health service use [14].

Information Sources and Search Methods

A comprehensive search was developed in consultation with an information specialist (DB) conducted for peer-reviewed articles in electronic databases.

Electronic Searches

The following electronic databases were searched from inception to October 5, 2018: Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica dataBASE (EMBASE), PsycINFO, Cumulative Index of Nursing and Allied Health Literature (CINAHL), and SCOPUS.

We developed the search strategy in MEDLINE and adapted the search for each database (Multimedia Appendix 1). Search filters used in other systematic reviews for KT strategies [9,17,18] and digital dissemination or intervention [12] were

XSL•FO
RenderX

adapted for use in this review. Additional search filters were developed to capture the outcomes included in the review [14].

Searching for Other Resources

The reference lists of all included studies were searched for additional relevant studies. A hand search of studies published between December 2016 and October 2018 was conducted in the *Implementation Science* and *Journal of Medical Internet Research (JMIR)* journals due to their relevance to the aims of the systematic review. Hand searches were conducted for 2 years, consistent with the practices of other Cochrane Reviews [19,20].

We also conducted a gray literature search using Google, with the search terms "knowledge translation" AND ("digital dissemination OR digital intervention") AND "public health" and reviewed the first 200 results against the eligibility criteria.

Selection of Studies

All studies obtained in the literature search were deduplicated. The titles and abstracts of the identified studies were independently screened for eligibility by 2 reviewers (from a pool of 4 reviewers: AB, CB, JB, and MM) and full texts of all relevant or unclear studies were obtained and reviewed against the inclusion criteria using Covidence. Any ambiguity in the inclusion of a study was resolved by discussion or by a third reviewer to reach consensus. Review authors were not blinded to author or journal information.

Data Collection and Analysis

Data Extraction and Management

Two review authors (AB and LW) independently extracted information from the included studies using a data extraction form developed and piloted by the research team. Any discrepancies regarding data extraction between the review authors were resolved via discussion. An attempt to contact study authors occurred if there was insufficient data in the included studies. The following information was extracted: general information (author name, title, date of publication, and country) methods (study design, setting, duration, sample size, and number of experimental conditions), participants (total number and participant characteristics), types of intervention (characteristics of the digital TEKT intervention and comparison components including the type of strategy and modality [website, web-based training, and emails]), type of outcome measures and results (outcomes aligned to the review inclusion criteria, data collection procedure, effect size, and summary data for each intervention group), and information to allow assessment of risk of study bias. Additionally, we extracted information regarding the intervention for each element recommended by the Template for Intervention Description and Replication (TIDieR) checklist. The TIDieR checklist is a 12-item checklist used to report characteristics of interventions in a structured manner. It includes information on why the study was conducted, how the study was conducted, materials used, where it was conducted, frequency of the intervention, and any tailoring or modifications [21]. Such information was not planned in the review protocol and was not used in formal study synthesis but has been included to enhance characterization of the trials.

Assessment of Risk of Bias

Two reviewers (AB and RH) independently assessed the risk of bias using the *Cochrane Handbook for Systematic Review of Interventions Risk of bias 2.0* tool [22]. Each included study was assessed for the following risks of bias (and rated as *high*, *low*, or *some concerns*): randomization, deviations from the intended intervention, missing outcome data, measurement of the outcome, and free of selective outcome reporting. The overall risk of bias for each study was determined using guidance from the *Cochrane Handbook for Systematic Review of Interventions Risk of bias 2.0* tool [22]. Discrepancies in rating the risk of bias between reviewers were resolved by discussion.

Data Synthesis

A random effects meta-analysis was planned in the protocol to provide a quantitative assessment of the effects of interventions. However, meta-analysis could not be conducted as outcomes reported were not able to be pooled due to considerable heterogeneity across the included studies. As such, the results are described narratively by synthesizing study findings by outcome. Within the outcome category, synthesis was then undertaken by first synthesizing the findings of studies comparing digital TEKT versus the control group and then by synthesizing the effects of studies comparing digital TEKT strategies versus an alternative intervention.

Results

Description of the Included Studies

A total of 6819 citations were screened for eligibility (Multimedia Appendix 2). The full text of 46 studies was obtained to determine eligibility against the review criteria. Of these 46 studies, 8 were included in the review [23-30]. The primary reasons for exclusion of ineligible studies were as follows: non–randomized controlled trials (non-RCTs; n=17), nondigital studies (n=6), not specifically KT (n=6), protocol only (n=4), based on the clinical setting (n=4), or had no digital TEKT outcome (n=1).

A summary of the study characteristics and outcomes of the 8 included studies is provided in Table 1. All 8 studies were RCTs, of which 4 were cluster RCTs [25,27,29,30]. A total of 3 studies were conducted in the United States [23-25], 2 in the Netherlands [26,28], 2 in Canada [27,30], and the remaining study was conducted in China [29].

Health professionals recruited in the included studies comprised nurses [23,26,30], nurse practitioners [23], physicians [23], child care health consultants [24], physiotherapists [30], primary health care workers [29], and public health practitioners [30]. A further study recruited program managers or coordinators of public health departments [27] and the remaining study recruited professionals employed in schools, community agencies, and policy-making bodies [25].

There was variability in the digital TEKT interventions tested. Overall, 5 of the interventions were web-based training programs

XSL•FO
RenderX

aimed at improving nutrition and physical activity knowledge in children [24], utilizing smoking cessation guidelines [26], body positivity in children and adolescents [30], physical activity levels in patients with cardiovascular risk factors [28], and improving basic public health services knowledge [29]. The remaining 3 interventions included simulation games to improve communication with patients with mental health disorders [23], access to digital resource materials on substance abuse prevention programs [25], and the use of tailored messaging in comparison with knowledge brokering or a web-based registry concerning healthy body weight promotion in children [27]. Of the 8 included studies, 6 reported using evidence-based information to develop the intervention or as part of the program [23-27,30], including the use of systematic reviews [27], and previous programs and studies [23-26,30]. A full description of the intervention consistent with the TIDieR checklist is provided in Multimedia Appendix 3 [23-30].

The majority of comparison groups for the included studies were control groups that had no access to the digital TEKT strategies (n=3) [23,24,26] or waitlist control group (n=2) [28,30]. A study that aimed at providing information on substance abuse prevention programs compared groups accessing information via CD-ROM or the internet to those who accessed information via printed materials [25]. A further study in public health department personnel compared its most interactive KT strategy (access to knowledge broker, web-based registry, and tailored messages) and its moderate interactive KT strategy (web-based registry and tailored messages) with the least interactive KT strategy of access to a web-based registry only [27]. The study in primary health care workers compared its face-to-face education group with its comparison group that accessed information via websites only, in an effort to improve cardiovascular risk management [29].

The majority of studies reported on fidelity with the intervention to some extent; however, studies were inconsistent in how fidelity was measured, which included time to complete intervention and percent use of program [23,24,27,28,30]. Of the 2 studies that reported on time to complete, the mean time ranged from 124 to 4998 min [23,24] and the percent usage of the program ranged from 45% to 100% in the 3 studies that reported this [27,28,30].

The most commonly reported KT outcome measure was a change in knowledge, as reported in 4 of the included studies [23,24,26,29,30]. Of the 8 included studies, 3 reported the outcome measure of intention to change behavior [23,25,28] and 2 reported self-efficacy for identifying evidence and confidence to improve the health professional's practice behavior [25,30]. Only 1 study reported the impact of digital TEKT strategies on measures of attitude regarding health professional practice behavior to improve physical activity in cardiovascular patients and on the health professional's perceived behavior control [28]. Changes to health policy, practice, or decision making were reported by 3 studies [26-28]. No study reported outcomes relating to individual-, community-, or population-level health outcomes.

The majority of outcome measures within the included studies (n=7) were evaluated using telephone, written, or web surveys [23,25-30], with the remaining study using a knowledge test to measure its outcome [24]. Of the 8 included studies, 2 studies reported using validated measures of trial outcomes [23,30].

XSL•FO
RenderX

Table 1. Study characteristics and key findings.

Study (reference, country, study design)	Population: type of participant, sample size, mean age and gender of participants, industry experience	Intervention description; duration of intervention; comparator	Outcomes and measures	Key findings
Albright et al [23]; United States; RCT[a]	227 nurses, nurse practitioners, and physicians; 81.9% female; 65.6% nurses; Industry experience: mean 10.89 (SD 11.01) years; Age not reported	*Description*: At risk in primary care web-based simulation *role-playing* game. Provides learners opportunities to practice role playing with emotionally responsive virtual patients that are experiencing mental health disorders; *Duration*: Simulation takes 1-1.5 hours to complete. *Comparator*: control	*Outcome*: knowledge and skills; *Measure*: 6-item Gatekeeper Behavior Scale web-based questionnaire; *Outcome*: likelihood to screen and manage mental health issues; *Measure*: single item, 4-point Likert-type scale web-based questionnaire	The score for the treatment group, postsimulation (mean 3.40, SD 0.89) was significantly higher than the control group at presimulation (mean 2.91, SD 0.69), *P*<.001; Likelihood of engaging in screening behavior for the treatment group (mean 3.27, SD 0.74) was significantly higher than the control group (mean 2.90, SD 0.87), *P*<.01
Benjamin et al [24]; United States; RCT	51 CCHCs[b]; Control (n=17): 6.9 years old; 94% female; 88% nursing degree. Web-based (n=17): 41.9 years old; 100% female; 94% nursing degree. In-person (n=16): 39.8 years old; 100% female; 87% nursing degree. Industry experience not reported	*Description*: web-based (group 1) and in-person trained (group 2) CCHCs. Each training included 4 modules: intervention overview, introduction to childhood overweight, nutrition and physical activity, and providing consultation to child care centers. In-person training and web-based training were designed to be similar in both content and structure. *Duration*: training took 3 hours; *Comparator*: control	*Outcome*: nutrition knowledge related to childhood overweight. *Measure*: 28 multiple choice questions (childhood overweight=4, nutrition for children=10, physical activity for children=8, and nutrition and physical activity for adults=6) with 2-5 possible response options	Participants from the web-based trained group (difference in pre/post score=16.18) did not perform better than the in-person trained group (difference in pre/post score=16.53). Both training groups improved significantly more than controls (difference in pre/post score=1.89; *P*<.001 for each group)
de Ruijter et al [26]; the Netherlands; RCT	269 PNs[c] across the Netherlands. Mean 47.3 years old; 97.8% female. PN counseling experience was mean 5.6 years	*Description*: Guideline adherence to smoking cessation counseling. Computer-tailored, web-based program relating to smoking cessation. Consisted of web-based modules, tailored advice, forum, and smoking cessation counseling materials. *Duration*: 6 months to access and use the program. *Comparator:* control group engaged in normal smoking cessation counseling practices	*Outcome*: adherence to STIMEDIC[d] guidelines. *Measure*: questions on guideline adherence concerned the 9 evidence-based counseling steps, as described in the STIMEDIC guideline. PNs adherence at baseline was assessed by asking PNs to self-report their adherence to each guideline step during complete smoking cessation trajectories of their last 10 patients (range 0-10). Additionally, during the trial period, guideline adherence was assessed by asking PNs to self-report their adherence to each guideline step after every consultation with a smoking patient using the counseling checklist	Overall intervention effect not reported. Significant interaction between groups based on the average years of counseling experience (*P*=.045)

XSL•FO
RenderX

Study (reference, country, study design)	Population: type of participant, sample size, mean age and gender of participants, industry experience	Intervention description; duration of intervention; comparator	Outcomes and measures	Key findings
Di Noia et al [25]; United States; RCT	188 school, community agency, and policy-making professionals; 68.6% females; 25% 30-39 years, 23% 40-49 years, 19% 50-59 years (mean age not reported); 48% some graduate school, 22.3% college, 10.6% some college. Industry experience not reported	*Description*: illustrative dissemination materials for 3 youth-oriented substance abuse prevention programs. Materials for each program were tailored for each setting (school, community agencies, and policy makers) and disseminated by: Group 1: accessed resource materials via CD-ROM (n=64); Group 2: accessed resource materials via the internet (n=69). *Duration*: 2 years. *Comparator*: resource materials accessed via printed pamphlets (n=55)	*Outcome*: self-efficacy. *Measure*[e]: via survey assessing professionals' self-efficacy for identifying and obtaining prevention programs to serve the needs of youth; confidence in ability to recommend programs to their constituents. *Outcome*: intention to apply prevention program materials; likelihood of their future applications of materials disseminated in the trial; likelihood of request program materials, implementing a prevention program and recommending programs to their constituents. *Measure*: via survey	No significant differences between groups for self-efficacy for the ability to identifying programs or recommend programs. Significant between channel effects in ability to obtain programs for pamphlet (mean 1.37, SD 0.93) versus internet (mean 0.87, SD 0.79) and pamphlet versus CD-ROM (mean 0.94, SD 0.84) at $P<.05$ at follow-up; No significant differences between groups for likelihood of requesting program or likelihood of implementing program. Significant difference between CD-ROM (mean 1.41, SD 1.13) and pamphlet (mean 1.55, SD 1.13) for likelihood of recommending program at $P<.05$ and significant difference between CD-ROM and internet (mean 1.06, SD 1.05) for likelihood of recommending program at $P<.05$ at follow-up
Dobbins et al [27]; Canada; RCT	108 public health departments with program managers and/or coordinators and/or program directors responsible for making program decisions related to healthy body weight promotion in children; 35% frontline staff; 26% manager; 47% nursing discipline; Mean 5 years in current position; Mean years in a public health role=13. Age and gender not reported	*Description*: the 3 interventions included access to a web-based registry of research evidence, tailored messaging, and a knowledge broker. Moderate interactive intervention (digital TEKT[f] strategy): tailored targeted messages plus access to a health evidence repository (TM). Most interactive intervention: access to a knowledge broker, tailored targeted messages plus access to health evidence repository (KB). Least interactive intervention: access to health evidence repository (HE). *Duration*: program implemented over 1 year	*Outcome*: public health policies and programs; *Measure*: This measure was derived as the sum of actual strategies, policies, and/or interventions for health body weight promotion in children being implemented by the health department. Participants were asked whether the public health policies and programs were being implemented by their health department (yes/no); *Outcome*: global evidence-based decision making. *Measure*: in a telephone-administered survey, participants were asked to report on the extent to which research evidence was considered in a recent program planning decision (previous 12 months) related to healthy body weight promotion	TM group improved significantly from baseline to follow-up in comparison to the HE and KB groups that showed no significant change ($P<.01$); Intervention had no significant effect on global evidence-based decision making ($P<.45$), although all groups improved to some extent (HE group: 0.74; TM group: −0.42; KB group: −0.09)

XSL•FO
RenderX

Study (reference, country, study design)	Population: type of participant, sample size, mean age and gender of participants, industry experience	Intervention description; duration of intervention; comparator	Outcomes and measures	Key findings
McVey et al [30]; Canada; RCT	89 public health practitioners (100% female) from 2 Canadian provinces; Public health participants: public health nurses (n=62) and nutritionists (n=27), with average number of years 12.72; 84.4% identified as white. Age not reported	*Description*: the student body Promoting Health at Any Size web-based program; 6 learning modules: (1) media and peer pressure (2) healthy eating, (3) active living, (4) teasing, (5) adult role models, and (6) school climate including case studies, background information, additional resources, and classroom activities. *Duration*: the intervention group had 60-day access to the web-based intervention. *Comparator*: waitlist control	*Outcome*: knowledge of the physical changes associated with puberty, facts concerning restrictive dieting, peer and adult influences, and the influence of the media on weight loss. *Measure*: assessed via a survey using true-false questions and Likert scales; *Outcome*: efficacy to fight weight bias. *Measure*: 6-item subscale used to assess self-efficacy expectations for fighting weight bias in their schools. On the basis of a 4-point Likert scale (Cronbach α=.44)	Physical changes associated with puberty: there was no significant interaction effect, $F_{1,77}$=.486, *P*=.488. Facts concerning restrictive dieting: there were no significant interactions or time effects found for any of the items that tapped knowledge about dieting. Peer or adult influences: there were no significant interactions or time effects. Influence of the media on weight loss: there were no significant interactions or time effects; There was a significant interaction effect found for the variable efficacy to fight weight bias, $F_{1,77}$=10.81, *P*=.002. Participants in the intervention group only reported significant improvements in efficacy scores between baseline and the postintervention periods, *P*<.001
Sassen et al [28]; the Netherlands; RCT	69 health care professionals with at least a bachelor's degree in nursing or physiotherapy and who had regular consultations with patients with CV[g] risk factors and low levels of PA. Control group: 78% female, mean 39.7 years old; 68% bachelor's degree, mean 9.58 years professional experience. Intervention group: 69% female, mean 38.6 years old; 79% bachelor's degree, mean 9.76 years professional experience	*Description*: web-based intervention to increase health care professionals' intention and encouraging behavior toward patient self-management, following CV risk management guidelines. Website contained modules to help the health professionals improve their professional behavior, support the health professional, improve patients' intention, and risk reduction. The website also included a forum directed at health professionals to share experiences with other professionals. *Duration*: not reported. *Comparator*: waitlist control	*Outcome*: intention to encourage CV patients to become physically active. *Measure*: self-assessed through a questionnaire (3 items on intention); *Outcome*: attitude to encourage PA[h] in CV patients. *Measure*: self-assessed through a questionnaire of a series of 8 questions regarding the usefulness of assessing patients' motivation, pros and cons of PA, teaching patients : resisting social pressure, teaching specific PA skills, teaching patients how to handle barriers, formulating PA goals, teaching patients to handle relapses, and helping patients understand the relationship between health problems and PA; *Outcome*: perceived behavior control. *Measure*: self-assessed through a questionnaire (23 items on behavior outcomes); *Outcome*: behavior change in encouraging CV patients to PA. *Measure*: assess via 2 items by asking whether professionals encourage CV patients to increase PA and how often do they encourage CV patients to become physically active	No significant differences in both the intervention and the control groups between baseline (mean: 6.25, SD: 1.00 and mean: 5.87, SD: 1.15) and follow-up (mean: 6.06, SD: 1.11 and mean: 6.02, SD: 091) for intention No significant differences in both the intervention and the control groups between baseline (mean: 6.30SD: 0.44 and men: 6.23, SD: 0.69) and follow-up for attitude (mean: 6.30 SD:0.56 and mean: 6.31SD: 0.68); Significant difference in perceived behavior control between baseline and follow-up for the intervention group (t_{26}=−2.954, *P*<.001, effect size=0.50) and a significant increase for the control group (t_{19}=−2.651, *P*=.02, effect size=0.54). No significant difference between intervention and control group; No significant differences in both the intervention and the control groups between baseline (mean 4.54, SD: 1.02 and mean: 4.83 SD: 0.69) and follow-up for behavior (mean: 4.63, SD: 0.85 and mean:4.79, SD: 0.82)

XSL•FO
RenderX

Study (reference, country, study design)	Population: type of participant, sample size, mean age and gender of participants, industry experience	Intervention description; duration of intervention; comparator	Outcomes and measures	Key findings
Zhan et al [29]; China; RCT	1237 primary health care workers. Blended learning group (n=569): Mean 41.67 years old, 48.9% female, 9.6% technical secondary school or below; Pure web-based learning group (n=563): mean 41.98 years old, 43.2% female, 77.3% technical secondary school or below	*Description*: the blended learning (intervention) and pure web-based learning (control) groups had the same course materials to improve basic public health services knowledge. Participants in the blended learning group studied PowerPoint-based theoretical materials, received the handouts of case study materials for self-studying and attended 1-day (8 hour) face-to-face case study training. *Duration*: overall study period was 5 weeks. *Comparator*: control (pure web-based learning group–digital TEKT strategy); received via a web-based platform: Microsoft PowerPoint; case studies consisted of 3 video sessions, and 2 discussion forums were developed on the training platform	*Outcome*: knowledge for course module components. *Measure*: a total of 3 knowledge MCQ[i] tests were developed, consisting of a 10-item MCQ test in course module 1, a 15-item MCQ test in course module 2, and a 20-item MCQ test in course module 3	Baseline knowledge scores of the 3 course modules between experimental and control group were similar. Higher gains in the experimental group than in the control group; module 1: adjusted mean difference=4.92, *P*<.001; module 2: adjusted mean difference=3.67, *P*=.004; module 3: adjusted mean difference=4.63, *P*<.001

[a]RCT: randomized controlled trial.

[b]CCHC: child care health consultants.

[c]PN: practice nurses.

[d]STIMEDIC: A registered trademark that stands for smoking cessation (SMR) in health care.

[e]Lower scores indicate more favorable ratings.

[f]TEKT: technology-enabled knowledge translation.

[g]CV: cardiovascular.

[h]PA: physical activity.

[i]MCQ: multiple choice questions.

Risk of Bias

Figures 1 and 2 summarize the risk of bias for each of the studies.

Of the 8 studies, 7 were assessed as having an overall high risk of bias [23,24,26-30]. Only 1 study was considered to have a high risk of bias for the randomization process [27]. For deviations from intended interventions, 3 studies were assessed as having a high risk of bias [24,29,30], with 1 study having some concerns [27]. Missing outcome data resulted in 3 studies having a high risk of bias [24,26,28], with measurement of the outcome highlighting 2 studies with a high risk of bias [23,26]. All 8 studies were classified as having some concerns in relation to the selection of the reported results.

XSL•FO
RenderX

Figure 1. Risk of bias summary.

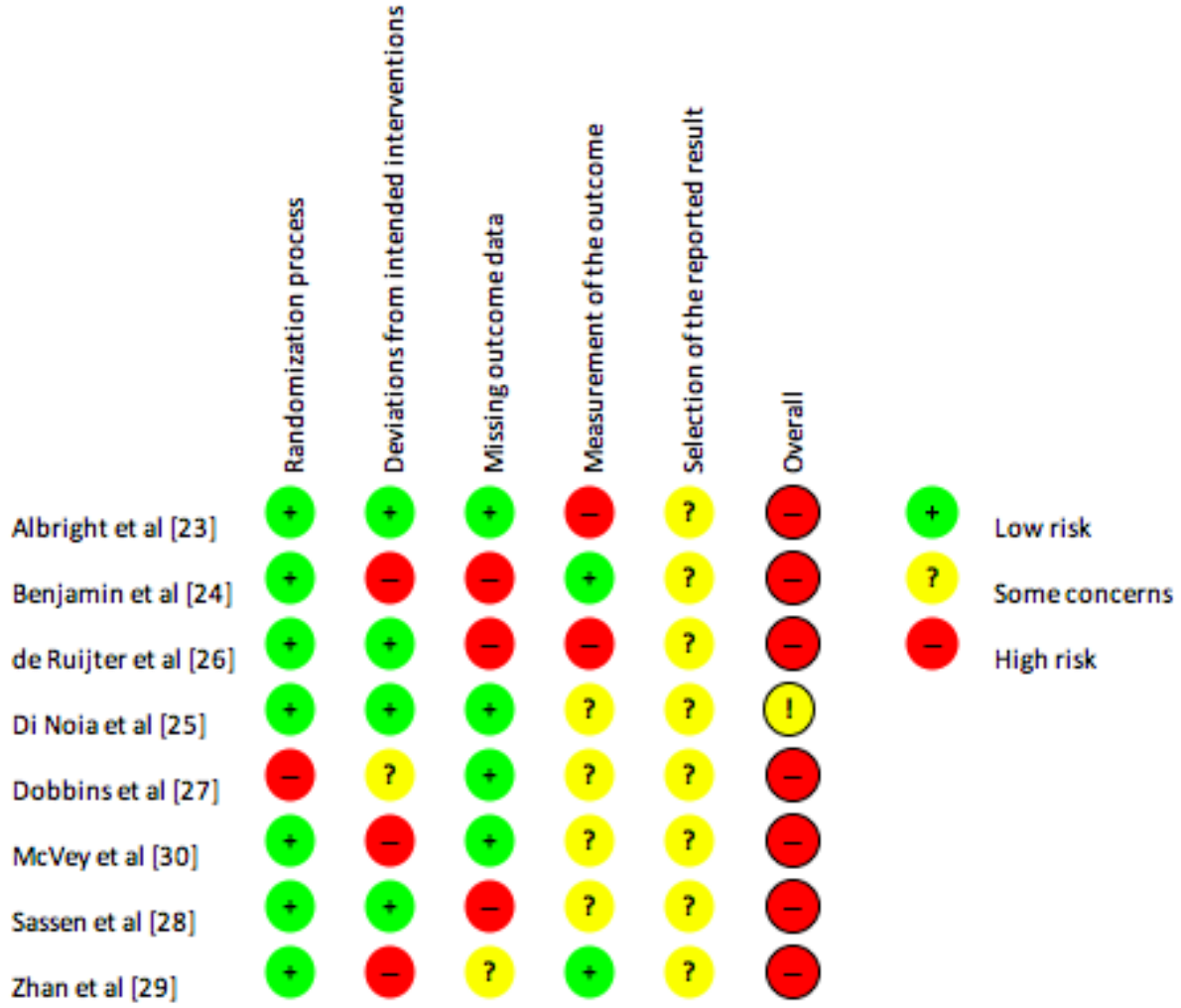
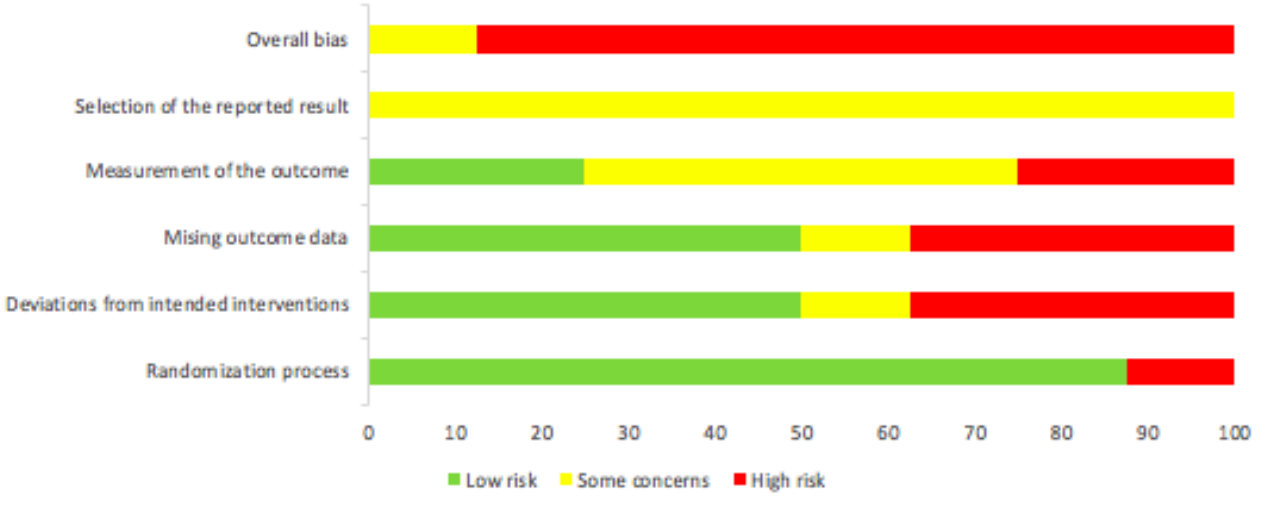


Figure 2. Risk of bias graph.



Key Findings

A summary of the outcomes and key findings is provided in Table 1.

XSL•FO
RenderX

Measures of Public Health End Users' Capacity to Make Evidence-Informed Decisions

Knowledge Change

A total of 4 trials examined the effects of digital TEKT strategies on knowledge outcomes [23,24,29,30]. Of these, 3 compared the effects of digital TEKT interventions with a control, each of which reported significant improvements in knowledge favoring the intervention [23,24,29]. A study using a web-based simulation game reported significant improvements in knowledge and skill scores postintervention to enhance the assessment of mental health disorders (P<.001) [23]. Similarly, a study in child care health consultants reported significant improvement in nutrition and physical activity knowledge among those receiving web-based training relative to those in a control that did not provide training (P<.001) [24]. A study in primary health care workers in rural China compared the provision of web-based learning only to improve knowledge regarding basic public health services in rural China with a blended learning group provided with a combination of digital and face-to-face training. The study found greater improvements in knowledge of public health services in the blended learning group (P<.001) [29]. Finally, a study allocating public health professionals to receive a web-based program to prevent eating disorder behavior in children or a waitlist control reported no significant differences across any knowledge outcomes [30].

Only 1 trial compared the effects of a digital TEKT intervention with an alternate KT intervention [24]. The study of child care health consultants compared nutrition knowledge among those receiving face-to-face training versus those receiving web-based training [24]. The study found no differences between groups in terms of knowledge outcomes.

Self-Efficacy

Overall, 2 trials assessed self-efficacy for identifying evidence and confidence to improve the health professional's practice behavior as a digital TEKT outcome, the findings of which were mixed [25,30]. The first, a study by Mcvey et al [30], found that public health professionals who had accessed the web-based program had significant improvements in self-efficacy scores compared with controls (P=.002). The second study compared alternate KT interventions and found no difference in professionals from schools, community agencies, and policy-making bodies, confidence to recommend suitable programs, or self-efficacy for identifying and recommending relevant programs between those allocated to receive substance abuse prevention program resources via CD-ROM or websites compared with those receiving such information via pamphlets [25]. However, there were between-group differences in self-efficacy to obtain youth prevention programs, with those in the internet group and the CD-ROM group reporting significant improvement when compared with the pamphlet group (P<.05) [25]. Bonferroni post hoc analyses highlighted favorable differences for the internet group (P<.05).

Intention

In all 3 studies that measured behavioral intentions reported mixed effects of digital TEKT interventions on these outcomes [23,25,28] and 2 studies compared digital TEKT interventions versus no intervention control. The first, a study by Albright et al [23], found that participants allocated to the web-based simulation game reported a higher likelihood of assessing and screening patients for mental health disorders relative to the control group (P<.01). However, Sassen et al [28] reported no significant differences between the website-trained group and the control group in the intention of health professionals to encourage cardiovascular patients to become physically active [28].

The study by Di Noia et al [25] compared the effects of different KT interventions for a substance abuse prevention program. The study found that participants receiving dissemination materials via the internet had significant improvements in their likelihood of recommending programs compared with the CD-ROM group (P<.05). The CD-ROM group, however, showed significant improvements compared with the pamphlet group (P<.05) [25]. There were no differences between any of the 3 groups in their likelihood of requesting programs or the likelihood of implementing the program [25].

Attitude

Only 1 study reported attitude as an outcome measure and reported that there were no differences between the website-trained group and the control group in their attitude to encourage cardiovascular patients to become physically active [28].

Perceived Behavior Control

Perceived behavior control to assess health professionals' perceived skills and knowledge in encouraging physical activity in cardiovascular patients was reported by Sassen et al [28], who found no significant difference between the website-trained group and the control group [28].

Measures of Evidence-Based Public Health Policy and Practice

Changes to Health Policy, Practice or Decision Making

Of the 8 included studies, 3 reported changes to health policy, practice, or decision making as an outcome measure [26-28]. A web-based study to improve the behavior of health professionals who were nurses or physiotherapists in following cardiovascular risk management guidelines reported no significant differences between groups in encouraging cardiovascular patients to increase physical activity levels (P=.68) [28].

A study by de Ruijter et al [26] in practice nurses reported on the outcome of smoking cessation guideline adherence and found that there was a significant difference between groups as counseling experience increased guideline adherence; however, the overall intervention effect was not reported.

The study by Dobbins et al [27] in public health departments reported outcome measures relating to a change in public health policies and programs and a change in global evidence-informed decision making. In relation to public health policies and programs, the group that received tailored messages via email plus access to a health repository improved significantly in the number of public health policies and programs implemented in comparison with both the group that received only access to the

XSL•FO
RenderX

health repository and the group that received access to the health repository, tailored messages, and a knowledge broker ($P<.01$) [27]. The study reported no significant differences between groups in relation to global evidence-informed decision making [27].

Measures of Individual-, Community-, or Population-Level Health Outcomes

No included studies reported on individual-, community-, or population-level health outcomes.

Discussion

Principal Findings

To the best of our knowledge, this review is the first to report on the effectiveness of digital TEKT strategies in improving public health end users' capacity to make evidence-informed health policy and practice decisions. Overall, the findings from the 8 included studies suggest that digital TEKT interventions may be effective at improving public health professionals' knowledge, and may be as effective at improving knowledge as a face-to-face KT approach. The effectiveness of digital TEKT strategies relative to control or other KT interventions for self-efficacy or behavioral intention outcomes and changes to health policy or practice were mixed. Such findings offer little guidance for those interested in utilizing digital TEKT strategies to promote the transfer of knowledge to improve public health and demonstrate a considerable need for further research in this field.

The small number of trials identified in this review examined the impact of the digital TEKT intervention on a narrow range of potential outcomes, namely knowledge, self-efficacy, and behavioral intentions. Other outcomes suggested by evaluation frameworks are important in assessing digital TEKT interventions, including behavioral and population-level health outcomes; however, these were rarely reported [14]. These findings are consistent with other systematic reviews of KT and digital dissemination strategies. For example, a systematic review published in 2012 on the effects of KT on public health identified only 5 included studies and reported outcomes relating to knowledge change and change in practice only [9]. Similarly, a systematic review in 2016 including 21 studies on the effectiveness of digital dissemination of clinical practice guidelines included outcomes such as usability of the technology (perceived usefulness and perceived ease of use) and practice behavior (using the theoretical domains framework including knowledge, skills, beliefs, and motivations), with knowledge and skills being the most frequently reported outcome and limited studies reporting other behavior change outcomes [12].

A 2019 systematic review on KT strategies in the clinical child health setting included 48 studies and reported outcomes relating to health care professionals or patient-reported outcomes [13]. Although the review included some digital TEKT studies, only 9 RCTs had a positive effect on clinician-related or patient outcomes [13]. This recent systematic review highlights the continued paucity of research on digital TEKT broadly. The limited reporting of behavioral and population-level health outcomes and the differences in measures used to report other outcomes highlight the challenges of conducting and synthesizing the effects of KT research. Although a variety of tools have been suggested to be useful in assessing the impact of digital KT strategies, including altmetric scores, measures of engagement and process indicators, and use of consistent evaluation metrics in reporting KT outcomes, would assist in appropriately assessing the effectiveness of digital TEKT and comparing these across trials. Core outcome sets are currently being used in several specific health areas to determine a standardized set of outcomes that should be measured and reported [31]. However, there are currently no core outcome sets developed for KT [32]. To address the issue of inconsistent evaluation outcomes in KT research, it may be necessary to develop KT-specific core outcome sets.

For knowledge outcomes, the findings of this review are broadly comparable with other systematic reviews of the effects of KT interventions in other health professions. For example, a 2012 review of public health KT strategies suggested that KT strategies were effective in improving knowledge outcomes, with 2 of 3 included studies reporting significant improvements in measures of knowledge acquisition [9]. Similarly, a 2016 systematic review found improvements in knowledge following web-based workshops, emails, educational web-based games, and multifaceted KT interventions, but not interventions using websites or computer software [12]. Consistent with the findings of this review, other systematic reviews have also reported mixed effects of KT on measures of behavioral intentions [12]. The findings of the review on other outcomes are difficult to contextualize, given the limited number of trials reporting these outcomes. Collectively these findings suggest that digital TEKT strategies may influence precursors for behavior change, such as knowledge and intention, but are yet to demonstrate impacts on public health policy and community outcomes. This is an important evidence gap for researchers to remedy, as public health policy and community impacts represent the ultimate goal of TEKT interventions. Indeed, such outcomes are required for TEKT strategies to yield improvements in public health decision making. The conduct of rigorous trials of digital TEKT strategies that include policy or community health outcomes, however, may represent a considerable challenge for researchers given the limited resources available to conduct such large trials with long periods of follow-up required for KT to occur [3]. Embedding such trials into public health services, which have an interest in KT for health service improvement, and have existing infrastructure and access to routinely collected data on community health improvement may represent a means of addressing the logistical challenges of undertaking such trials [33].

Strengths and Limitations

There are a number of strengths to this review. A comprehensive search strategy was used in consultation with an information specialist, using multiple electronic databases, hand search of reference lists, and gray literature searches. A systematic approach was utilized to review the current literature on digital TEKT in public health, including the use of pairs of reviewers to double-screen studies for inclusion, extract data, and determine the quality of the studies included. However, there were some limitations that must be considered when interpreting

XSL•FO
RenderX

the findings of the review. All but 1 study was assessed as having an overall high risk of bias, particularly due to the measurement of the outcome and selection of the reported results. As most studies had follow-up periods <12 months, it is difficult to determine the effects of the intervention over a longer period of time, which is particularly important when considering behavior and policy change. There was also a high level of heterogeneity due to differing health professional populations, interventions, and outcome measures, making it difficult to draw comparisons between studies and limiting the ability to conduct a meta-analysis. The KT search terms used in this review may have missed relevant KT studies, given the high variability of terms that are used in health to describe KT, a common limitation described in the literature [2,3].

Conclusions

This review addresses an important knowledge gap for digital TEKT in the public health setting and is the first to synthesize the effectiveness of digital TEKT interventions. Although the review has highlighted potential improvements in knowledge using digital TEKT strategies, it remains unclear whether digital TEKT strategies improve other behavioral and population health outcomes. Currently, there are limited studies assessing digital TEKT in the public health setting, with a limited range of outcomes to assess their effectiveness appropriately. As such, the findings of the review provide limited guidance to assist in the development of effective digital TEKT strategies. A recent scoping review conducted in 2020 summarized the most relevant KT frameworks for use by researchers, policy makers, and clinicians in the health care setting [34]. In the absence of sufficient evidence on the effectiveness of TEKT, the use of relevant KT frameworks is suggested to appropriately disseminate research findings to influence evidence-informed health policy and practice decisions.

Acknowledgments

This work was supported by infrastructure support from Hunter New England Population Health. RH was supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship and a Prevention Research Support Program Fellowship funded by the New South Wales Ministry of Health. NN was supported by an NHMRC Translating Research Into Practice (TRIP) Fellowship and a Hunter New England Clinical Research Fellowship. RS was supported by an NHMRC TRIP Fellowship. LW was supported by an NHMRC Career Development Fellowship and a Heart Foundation Future Leader Fellowship and a Hunter New England Clinical Research Fellowship.

Authors' Contributions

AB, RS, NN, and LW conceived and designed the systematic review. DB, AB, and LW developed the search strategy and DB conducted the search. AB, CB, JB, and MM screened studies for inclusion. AB and RH completed the risk of bias assessment. AB and LW completed the data extraction. AB drafted the manuscript, with all coauthors contributing to drafts of the paper. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medline Search terms.
[DOCX File , 24 KB - jmir_v22i7e17274_app1.docx]

Multimedia Appendix 2
PRISMA Flow Diagram.
[DOCX File , 42 KB - jmir_v22i7e17274_app2.docx]

Multimedia Appendix 3
Template for Intervention Description and Replication checklist for all included studies.
[DOCX File , 35 KB - jmir_v22i7e17274_app3.docx]

References

1. Orton L, Lloyd-Williams F, Taylor-Robinson D, O'Flaherty M, Capewell S. The use of research evidence in public health decision making processes: systematic review. PLoS One 2011;6(7):e21704 [FREE Full text] [doi: 10.1371/journal.pone.0021704] [Medline: 21818262]
2. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7:50 [FREE Full text] [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]
3. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]

XSL•FO
RenderX

4. Syed MA. Knowledge translation facilitating co-creation of evidence in public health. BMJ Evid Based Med 2019 Mar;24(1):15-19. [doi: 10.1136/bmjebm-2018-111017] [Medline: 30297379]
5. About Us. Canadian Institutes of Health Research - CIHR. URL: http://www.cihr-irsc.gc.ca/e/29418.html [accessed 2018-06-29]
6. Poot CC, van der Kleij RM, Brakema EA, Vermond D, Williams S, Cragg L, et al. From research to evidence-informed decision making: a systematic approach. J Public Health (Oxf) 2018 Mar 1;40(suppl_1):i3-i12 [FREE Full text] [doi: 10.1093/pubmed/fdx153] [Medline: 29538728]
7. Damarell RA, Tieman JJ. How do clinicians learn about knowledge translation? An investigation of current web-based learning opportunities. JMIR Med Educ 2017 Jul 13;3(2):e12 [FREE Full text] [doi: 10.2196/mededu.7825] [Medline: 28705788]
8. Malla C, Aylward P, Ward P. Knowledge translation for public health in low- and middle- income countries: a critical interpretive synthesis. Glob Health Res Policy 2018;3:29 [FREE Full text] [doi: 10.1186/s41256-018-0084-9] [Medline: 30377666]
9. LaRocca R, Yost J, Dobbins M, Ciliska D, Butt M. The effectiveness of knowledge translation strategies used in public health: a systematic review. BMC Public Health 2012 Sep 7;12:751 [FREE Full text] [doi: 10.1186/1471-2458-12-751] [Medline: 22958371]
10. Ho K, Jarvis-Selinger S, Lauscher H. Technology Enabled Knowledge Translation for eHealth. New York, USA: Springer; 2012.
11. Wilson PM, Petticrew M, Calnan MW, Nazareth I. Disseminating research findings: what should researchers do? A systematic scoping review of conceptual frameworks. Implement Sci 2010 Nov 22;5:91 [FREE Full text] [doi: 10.1186/1748-5908-5-91] [Medline: 21092164]
12. de Angelis G, Davies B, King J, McEwan J, Cavallo S, Loew L, et al. Information and communication technologies for the dissemination of clinical practice guidelines to health professionals: a systematic review. JMIR Med Educ 2016 Nov 30;2(2):e16 [FREE Full text] [doi: 10.2196/mededu.6288] [Medline: 27903488]
13. Campbell A, Louie-Poon S, Slater L, Scott SD. Knowledge translation strategies used by healthcare professionals in child health settings: an updated systematic review. J Pediatr Nurs 2019;47:114-120. [doi: 10.1016/j.pedn.2019.04.026] [Medline: 31108324]
14. Ho K, Bloch R, Gondocz T, Laprise R, Perrier L, Ryan D, et al. Technology-enabled knowledge translation: frameworks to promote research and practice. J Contin Educ Health Prof 2004;24(2):90-99. [doi: 10.1002/chp.1340240206] [Medline: 15279134]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Macintyre CR. Public health and health reform in Australia. Med J Aust 2011 Jan 3;194(1):38-40. [Medline: 21449867]
17. Albrecht L, Archibald M, Snelgrove-Clarke E, Scott SD. Systematic review of knowledge translation strategies to promote research uptake in child health settings. J Pediatr Nurs 2016;31(3):235-254. [doi: 10.1016/j.pedn.2015.12.002] [Medline: 26786910]
18. Scott SD, Albrecht L, O'Leary K, Ball GD, Hartling L, Hofmeyer A, et al. Systematic review of knowledge translation strategies in the allied health professions. Implement Sci 2012 Jul 25;7:70 [FREE Full text] [doi: 10.1186/1748-5908-7-70] [Medline: 22831550]
19. Hodder RK, O'Brien KM, Stacey FG, Tzelepis F, Wyse RJ, Bartlem KM, et al. Interventions for increasing fruit and vegetable consumption in children aged five years and under. Cochrane Database Syst Rev 2019 Nov 7;2019(11):-. [doi: 10.1002/14651858.CD008552.pub6] [Medline: 31697869]
20. Wolfenden L, Barnes C, Jones J, Finch M, Wyse RJ, Kingsland M, et al. Strategies to improve the implementation of healthy eating, physical activity and obesity prevention policies, practices or programmes within childcare services. Cochrane Database Syst Rev 2020 Feb 10;2:CD011779. [doi: 10.1002/14651858.CD011779.pub3] [Medline: 32036618]
21. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
22. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
23. Albright G, Bryan C, Adam C, McMillan J, Shockley K. Using virtual patient simulations to prepare primary health care professionals to conduct substance use and mental health screening and brief intervention. J Am Psychiatr Nurses Assoc 2018;24(3):247-259. [doi: 10.1177/1078390317719321] [Medline: 28754067]
24. Benjamin SE, Tate DF, Bangdiwala SI, Neelon BH, Ammerman AS, Dodds JM, et al. Preparing child care health consultants to address childhood overweight: a randomized controlled trial comparing web to in-person training. Matern Child Health J 2008 Sep;12(5):662-669. [doi: 10.1007/s10995-007-0277-1] [Medline: 17713847]

XSL•FO
RenderX

25. di Noia J, Schwinn TM, Dastur ZA, Schinke SP. The relative efficacy of pamphlets, CD-ROM, and the internet for disseminating adolescent drug abuse prevention programs: an exploratory study. Prev Med 2003 Dec;37(6 Pt 1):646-653 [FREE Full text] [doi: 10.1016/j.ypmed.2003.09.009] [Medline: 14636798]
26. de Ruijter D, Candel M, Smit ES, de Vries H, Hoving C. The effectiveness of a computer-tailored e-learning program for practice nurses to improve their adherence to smoking cessation counseling guidelines: randomized controlled trial. J Med Internet Res 2018 May 22;20(5):e193 [FREE Full text] [doi: 10.2196/jmir.9276] [Medline: 29789278]
27. Dobbins M, Hanna SE, Ciliska D, Manske S, Cameron R, Mercer SL, et al. A randomized controlled trial evaluating the impact of knowledge translation and exchange strategies. Implement Sci 2009 Sep 23;4:61 [FREE Full text] [doi: 10.1186/1748-5908-4-61] [Medline: 19775439]
28. Sassen B, Kok G, Schepers J, Vanhees L. Supporting health care professionals to improve the processes of shared decision making and self-management in a web-based intervention: randomized controlled trial. J Med Internet Res 2014 Oct 21;16(10):e211 [FREE Full text] [doi: 10.2196/jmir.3170] [Medline: 25337988]
29. Zhan X, Zhang Z, Sun F, Liu Q, Peng W, Zhang H, et al. Effects of improving primary health care workers' knowledge about public health services in rural China: a comparative study of blended learning and pure e-learning. J Med Internet Res 2017 May 1;19(5):e116 [FREE Full text] [doi: 10.2196/jmir.6453] [Medline: 28461286]
30. McVey G, Gusella J, Tweed S, Ferrari M. A controlled evaluation of web-based training for teachers and public health practitioners on the prevention of eating disorders. Eat Disord 2009;17(1):1-26. [doi: 10.1080/10640260802570064] [Medline: 19105058]
31. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials 2007 Nov 26;8:39 [FREE Full text] [doi: 10.1186/1745-6215-8-39] [Medline: 18039365]
32. Advanced Search. The COMET Initiative. URL: http://www.cometinitiative.org/Studies [accessed 2020-02-28]
33. Wolfenden L, Yoong SL, Williams CM, Grimshaw J, Durrheim DN, Gillham K, et al. Embedding researchers in health service organizations improves research translation and health service performance: the Australian Hunter New England Population Health example. J Clin Epidemiol 2017 May;85:3-11 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.03.007] [Medline: 28341367]
34. Esmail R, Hanson HM, Holroyd-Leduc J, Brown S, Strifler L, Straus SE, et al. A scoping review of full-spectrum knowledge translation theories, models, and frameworks. Implement Sci 2020 Feb 14;15(1):11 [FREE Full text] [doi: 10.1186/s13012-020-0964-5] [Medline: 32059738]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
EMBASE: Excerpta Medica dataBASE
JMIR: Journal of Medical Internet Research
KT: knowledge translation
MEDLINE: Medical Literature Analysis and Retrieval System Online
NHMRC: National Health and Medical Research Council
RCT: randomized controlled trial
TEKT: technology-enabled knowledge translation
TIDieR: Template for Intervention Description and Replication
TRIP: Translating Research Into Practice

Edited by G Eysenbach; submitted 02.12.19; peer-reviewed by J Hu, K Reuter, L Albrecht; comments to author 10.01.20; revised version received 12.03.20; accepted 23.03.20; published 31.07.20.

Please cite as:
Brown A, Barnes C, Byaruhanga J, McLaughlin M, Hodder RK, Booth D, Nathan N, Sutherland R, Wolfenden L
Effectiveness of Technology-Enabled Knowledge Translation Strategies in Improving the Use of Research in Public Health: Systematic Review
J Med Internet Res 2020;22(7):e17274
URL: http://www.jmir.org/2020/7/e17274/
doi:10.2196/17274
PMID:32735231

©Alison Brown, Courtney Barnes, Judith Byaruhanga, Matthew McLaughlin, Rebecca K Hodder, Debbie Booth, Nicole Nathan, Rachel Sutherland, Luke Wolfenden. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Behavior Change Interventions for Younger Children With Chronic Health Conditions: Systematic Review

Amberly Brigden[1], BSc, MSc; Emma Anderson[1], BA, MSc, DHealthPsych; Catherine Linney[1], BSc, MA; Richard Morris[2], BSc, MSc, PhD; Roxanne Parslow[1], BSc, PhD; Teona Serafimova[1], MBBS, MSc; Lucie Smith[1]; Emily Briggs[1], BSc; Maria Loades[1,3], DClinPsych; Esther Crawley[1], BA, BM BCh, PhD

[1]Centre for Academic Child Health, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[2]Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]Department of Psychology, University of Bath, Bath, United Kingdom

Corresponding Author:
Amberly Brigden, BSc, MSc
Centre for Academic Child Health, Bristol Medical School
University of Bristol
1-5 Whiteladies Road
Bristol, BS8 1NU
United Kingdom
Phone: 44 0117 42 83080
Email: amberly.brigden@bristol.ac.uk

Abstract

Background: The prevalence of chronic health conditions in childhood is increasing, and behavioral interventions can support the management of these conditions. Compared with face-to-face treatment, the use of digital interventions may be more cost-effective, appealing, and accessible, but there has been inadequate attention to their use with younger populations (children aged 5-12 years).

Objective: This systematic review aims to (1) identify effective digital interventions, (2) report the characteristics of promising interventions, and (3) describe the user's experience of the digital intervention.

Methods: A total of 4 databases were searched (Excerpta Medica Database [EMBASE], PsycINFO, Medical Literature Analysis and Retrieval System Online [MEDLINE], and the Cochrane Library) between January 2014 and January 2019. The inclusion criteria for studies were as follows: (1) children aged between 5 and 12 years, (2) interventions for behavior change, (3) randomized controlled trials, (4) digital interventions, and (5) chronic health conditions. Two researchers independently double reviewed papers to assess eligibility, extract data, and assess quality.

Results: Searches run in the databases identified 2643 papers. We identified 17 eligible interventions. The most promising interventions (having a beneficial effect and low risk of bias) were 3 targeting overweight or obesity, using exergaming or social media, and 2 for anxiety, using web-based cognitive behavioral therapy (CBT). Characteristics of promising interventions included gaming features, therapist support, and parental involvement. Most were purely behavioral interventions (rather than CBT or third wave), typically using the behavior change techniques (BCTs) *feedback and monitoring*, *shaping knowledge*, *repetition and substitution*, and *reward*. Three papers included qualitative data on the user's experience. We developed the following themes: *parental involvement*, *connection with a health professional is important for engagement*, *technological affordances and barriers*, and *child-centered design*.

Conclusions: Of the 17 eligible interventions, digital interventions for anxiety and overweight or obesity had the greatest promise. Using qualitative methods during digital intervention development and evaluation may lead to more meaningful, usable, feasible, and engaging interventions, especially for this underresearched younger population. The following characteristics could be considered when developing digital interventions for younger children: involvement of parents, gaming features, additional therapist support, behavioral (rather than cognitive) approaches, and particular BCTs (*feedback and monitoring*, *shaping knowledge*, *repetition and substitution*, and *reward*). This review suggests a model for improving the conceptualization and reporting of behavioral interventions involving children and parents.

(J Med Internet Res 2020;22(7):e16924) doi:10.2196/16924

XSL•FO
RenderX

KEYWORDS

mobile phone; mHealth; mobile health; eHealth; electronic health; digital health; behavior; pediatrics; chronic illness; systematic review

Introduction

Background

The prevalence of chronic health conditions in childhood is increasing [1,2]. Chronic health conditions are defined as "any physical, emotional, or mental condition that prevented him or her from attending school regularly, doing regular school work, or doing usual childhood activities or that required frequent attention or treatment from a doctor or other health professional, regular use of any medication, or use of special equipment"[3].

Behavioral interventions can support the treatment and management of chronic health conditions and can be effective in improving symptom management, reducing physical disability, and improving mental health [4-6]. These outcomes are particularly important in childhood because they have implications for children across their lifespan [7-13]. Behavioral or behavior change interventions are sets of techniques that aim to change health behaviors [14]. For children with long-term health conditions, these interventions typically focus on adherence to medical treatment, education about the medical condition, and improving aspects of medical care [15]. A specific example is the management of diabetes via behavioral intervention; glycemic control can be improved by encouraging behaviors of blood glucose monitoring, selection of healthy food choices, attendance at routine clinical appointments, and adherence to insulin therapy or other medications [16]. Improving the management of chronic health conditions at an early age can lead to immediate health improvements, but it also lays the foundations for health across the lifespan of the patient [7]. As such, it is important that younger children (and their families) are supported to improve understanding of their condition and develop self-management skills [17].

Digital interventions can deliver behavior change interventions using mobile phones, smartphones, portable computers, desktop computers, the internet, wearable technology, and television [18]. This is an emerging and rapidly developing field of research, and the potential advantages include increased cost-effectiveness, anonymity for users, appeal to younger people, and the ability for recipients to access interventions anywhere and at their own pace [19-22]. There is a growing body of evidence to suggest that digital interventions are potentially effective for adults and adolescents with chronic health conditions; they have beneficial effects on improving knowledge, self-management, self-care, quality of life, medication use, symptom control, and health service utilization [23-30]. However, there are some potential disadvantages that may affect the uptake, attrition, and efficacy of interventions. Some individuals may not be able to access the intervention because of technical issues, illiteracy, or the cost involved in obtaining the devices. Negative attitudes toward technology may also create barriers to use, and this includes concerns about data security. A lack of strong therapeutic relationships may discourage users and reduce engagement and efficacy of interventions. These potential disadvantages [31,32] should be carefully considered when planning and designing digital interventions. Furthermore, there are limitations with the evidence base for digital interventions, with systematic reviews highlighting the need for clearer reporting and higher quality research [23-29].

Despite the increasing availability of digital interventions and a growing body of evidence for adults and adolescents, there has been inadequate attention to designing and delivering these interventions to children. Children have different developmental characteristics and needs, and the developmental stage of children should be considered when designing interventions [33,34]. To the best of our knowledge, there are no systematic reviews that specifically investigate digital interventions for the management of chronic health conditions in children (aged <13 years). Furthermore, reviews investigating digital interventions for young people with chronic health conditions typically do not include children aged below 10 years [35], or, if they do, only a minority of the interventions included in the reviews include children aged below 13 years [15,36-38], recognizing that there are "fewer interventions targeting…the extreme pediatric age ranges of early childhood and emerging adulthood" [16]. The reviews spanning childhood and adolescence note important differences between these age groups. Three separate reviews of internet and computer-based cognitive behavioral therapy (CBT) for mental health problems found different treatment effects for older and younger children. The reviews found some positive effects for adolescents, but concluded that treatment effects were smaller or more uncertain for younger children [36,37,39]. Similarly, a review of electronic health interventions for young people with long-term physical conditions concluded that effectiveness was uncertain at this time, especially in children aged <10 years [15]. One review acknowledged, "we could not take the developmental stage of the patients…into account. As evidence is mounting, this issue should be addressed in future trials" [17].

Therefore, this review aimed to explore digital interventions for the management of chronic health conditions in children aged between 5 and 12 years.

Behavior change interventions are often complex [40], which can pose a challenge when synthesizing the effects of these interventions [41]. Advances in behavioral science have provided taxonomies and coding systems that help identify specific characteristics or *active ingredients* associated with effective interventions [42]. This includes the behavior change techniques (BCTs) taxonomy [43], which presents 93 discrete BCTs, "observable, replicable and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior"[43]. In addition to understanding what is being delivered (BCTs), it is important to understand how the content is delivered; this can be categorized using the mode of delivery taxonomy [18]. Identifying the theoretical underpinnings is possible with a coding frame [44]. Using these

XSL•FO
RenderX

BCTs, mode of delivery and theory taxonomies in systematic reviews may result in more optimal evidence syntheses and health care practice recommendations [41].

Objectives

This systematic review aimed to investigate digital interventions for the management of chronic health conditions in children aged between 5 and 12 years. We used an inclusive definition of chronic health conditions that included both physical and mental health. Conceptually, behavioral interventions for physical and mental health conditions are the same; they are designed to change the child's behavior to improve the clinical outcome. Furthermore, there is a strong overlap between physical and mental conditions; comorbidity of physical and mental health conditions is common [45], and many conditions involve both mental and physical health issues (eg, chronic fatigue syndrome or myalgic encephalomyelitis, pain, and obesity), thus developing integrated approaches toward mental and physical health is increasingly becoming a priority [46]. In this review, we aimed to answer the following questions: (1) Which of these interventions are effective in promoting behavior change for the management of the chronic health condition? (2) What are the characteristics of effective interventions, considering the following: recipients, what is being delivered (BCTs), how this content is being delivered (the mode of delivery), the theoretical basis, and the modality of the intervention? and (3) What are the users' experiences of the digital intervention?

Methods

Registration

The review was prospectively registered in the Prospective Register of Systematic Reviews (PROSPERO) database.

Search Strategy

We carried out a systematic search of relevant databases: Excerpta Medica Database (EMBASE), PsycINFO, Medical Literature Analysis and Retrieval System Online (MEDLINE), and the Cochrane Library (January 2019). The search strategy included keywords and Medical Subject Headings (MeSH) for (1) children aged between 5 and 12 years, (2) behavior change, (3) randomized controlled trials (RCTs), (4) digital interventions, and (5) chronic health conditions (we used a mixture of generic terms, ie, "Chronic disease," and also search-specific terms, informed by the most common chronic illness in childhood; Multimedia Appendix 1) [47].

Screening

To be included in this review, studies had to fulfill the following criteria:

1. Include children aged between 5 and 12 years (this review aimed to examine digital interventions for children in the developmental stages of middle childhood).
2. Include children with a chronic health condition, excluding those with developmental delays.
3. Investigate a digital intervention to promote behavior change. Digital interventions included those delivered via internet (static or interactive websites, automated emails, or web-based apps), personal computers (PCs; eg, PC videogames), social media, mobile phones (automated phone calls or short text messages), or smartphones (mobile websites or smartphone apps). These may be stand-alone interventions or guided (eg, therapist supported).
4. Compare the digital intervention with any comparator.
5. Have an RCT study design (RCTs are considered the gold standard for judging the benefits of treatments [48], and including RCTs only allowed us to focus on the interventions most likely to be adopted into clinical care).
6. Published in peer-reviewed journals and available in English.
7. Published between 2014 and January 2019. We chose a 5-year time frame because of the rapid pace of digital interventions [49], indicating that older interventions were not likely to be relevant.

Titles and abstracts (stage 1) and full-text papers (stage 2) were independently double screened against the inclusion and exclusion criteria using the data management platform Rayyan (stage 1) and Covidence (stage 2). AB screened all papers, and CL, LS, and EB were responsible for the independent second screening. Reasons for exclusion were recorded at stage 2. Discrepancies at both stages were discussed and resolved in meetings by the reviewers. Papers were tracked using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [50].

Data Extraction and Synthesis

For data extraction, papers were reviewed independently by 2 researchers and conflicts were resolved in regular meetings (AB reviewed all papers, and CL, LS, and EB were responsible for an independent second review). Two researchers independently coded BCTs (EA and AB, a health psychologist and health psychology trainee, respectively). We extracted information that allowed us to answer the 2 primary research questions, as described in Table 1. If the full text did not contain the information needed, we made 2 attempts to contact the authors by email.

Due to the clinical and methodological heterogeneity, we synthesized data using narrative synthesis [51,52] to answer our research questions. We reported effectiveness based on whether interventions were deemed as *very promising*, *quite promising*, *possibly promising*, *nonpromising*, or *unable to assess effectiveness*, determined by change in the outcomes and the quality of the science (Table 1 defines these categories).

XSL•FO
RenderX

Table 1. Data extraction.

Data extraction category	Details extracted

Population	• Age: the age range of the population, at the time of entry into the study. Ages were then grouped by UK school *key stage* categories [53]:
	  • 5-7 years, corresponding to key stage 1.
	  • 8-11 years, corresponding to key stage 2.
	  • 12 years, corresponding to key stage 3.
	• Chronic health condition: the chronic health condition that the intervention was designed to target.

Overview of intervention	• Overview of aims: the overview of the aims of the RCT[a]. • Overview of intervention: an overview of the digital component of the intervention and, if applicable, other key components. • Overview of comparator: an overview of the comparator arm or arms.

Aim 1: effectiveness	• Overview of *promise*: *promise* was based on the beneficial effects of the intervention and the quality of the study (risk of bias). To determine the beneficial effects of the intervention, we looked at the CIs of the mean difference from baseline to follow-up between the intervention and control group, considering the behavioral outcomes and the primary outcome (or outcomes). We developed 5 categories:
	  • Interventions were deemed *very promising* where there were beneficial effects of the intervention on both the primary outcome *and* at least one behavior change outcome, and the evidence was judged as having a low risk of bias.
	  • Interventions were deemed *quite promising* where there were beneficial effects of the intervention on the primary outcome *or* at least one behavior change outcome (but not both), and the evidence was judged as having a low risk of bias, or some concerns.
	  • Interventions were deemed *possibly promising* where there were beneficial effects of the intervention on the primary outcome and or behavior change outcome (or outcomes), *but* the study was deemed to have a high risk of bias.
	  • Interventions were deemed *nonpromising* where there were *no* beneficial effects of the interventions on either the primary outcome or behavior change outcomes.
	  • Interventions were put in the category *unable to assess effectiveness* where there were *no* effectiveness data available for the primary outcome or behavior change outcomes, for example, if the paper was a pilot or feasibility RCT.
	• The direction and size of the effect [51] was extracted for behavioral and primary outcomes, and the following was reported:
	  • Summary of the effect of the intervention compared with the control.
	  • Statistic comparing the change in the intervention group and control group from baseline to final follow-up. Where sufficient information was available comprising either SDs and numbers of participants, or SEs, we calculated the net mean difference (difference in mean changes), with 95% CI and *P* value. Where possible, this was interpreted in the context of the authors' judgment of clinically significant effects.
	  • If available, we reported the adjusted mean difference (adjusted for baseline measures) as this is the accepted best method.
	• Outcome measure: all behavioral outcomes were extracted, as exploring the effect of the intervention on behavior change was the primary aim of this review. The primary outcome was also extracted as this is the main determinant of whether the study is considered a *success* or a *failure* [54]. For each outcome we extracted:
	  • the behavior and or primary outcome
	  • how this was measured
	  • the final time point.
	• Adverse events: health interventions carry some risk of harm. Systematic reviews should minimize bias toward favoring an intervention by assessing adverse effects alongside beneficial effects [55]. Data on adverse events associated with the intervention were extracted.

XSL•FO
RenderX

Data extraction category	Details extracted

Aim 2: characteristics of promising interventions	• The following data were extracted from very promising, quite promising, and possibly promising interventions:
	  • Recipients: whether the intervention was delivered directly to the child, via a parent-proxy or both.
	  • Intervention techniques: intervention techniques refer to what is being delivered, the content or *active ingredients* of an intervention. The behavior change techniques taxonomy provides a standardized method of classifying intervention content [43]. This taxonomy consists of 93 behavior change techniques, in 16 groupings. We coded interventions using the 16 groupings.
	  • We coded whether each BCT identified was delivered to the parent or the child and whether it featured in the digital or human component.
	• Digital mode of delivery: intervention mode of delivery refers to how the content is delivered. We categorized mode of delivery, based on elements of the mode of delivery Taxonomy [18]:
	  • *Tailored or generic:* tailored interventions change the content of the text, images, recommendation, messages, etc based on the individual needs of the user.
	  • *Guided or stand-alone:* guided interventions are delivered with some form of support by a professional or coach, either with human contact or electronic means (eg, email or telephone calls).
	  • *Interactive techniques*: these include play, arts, story, and or game-based techniques.
	• Theoretical basis: whether a named theory of behavior or behavior change was explicitly mentioned in the Abstract, Introduction, or Methods section [44].
	• Modality: the intervention modality, coded as either a first, second, or third wave intervention. *First wave* interventions are purely behavioral, based on the theory that all behaviors are learned (through classical and operant conditioning) [56], and that maladaptive behaviors can be changed using principles such as reinforcement, modeling, graded tasks and habit formation [43]. *Second wave* refers to cognitive behavioral interventions, based on the principle that thoughts, feelings, physical sensations and actions are interconnected; individuals are supported to identify negative or unhelpful patterns in their cognitions, emotions, behaviors, physical sensations and supported to adopt more adaptive patterns [57]. *Third wave* interventions are characterized by techniques such as metacognition, acceptance, mindfulness, compassion and spirituality [56].

Aim 3: The users' experience of the digital intervention	• Qualitative analysis: two researchers independently reviewed all eligible papers and identified those that included qualitative data about the users' experience of the digital intervention. Qualitative data were extracted, compared, and summarized into themes.

[a]RCT: randomized controlled trial.

Quality Assessments

As all studies in this review were RCTs, the Cochrane risk of bias tool for randomized trials (RoB 2.0) [58] was used to assess the scientific quality of each study. Two researchers reviewed each paper, and the researchers then compared their quality assessment and resolved conflicts (AB reviewed all papers, and ML, LS, and EB were responsible for an independent second review). Following this, each paper was given a score of either low risk of bias, some concerns, or high risk of bias. Where available, we reviewed trial registries and published protocols. If needed, we also requested further information from the authors, including statistical analysis plans.

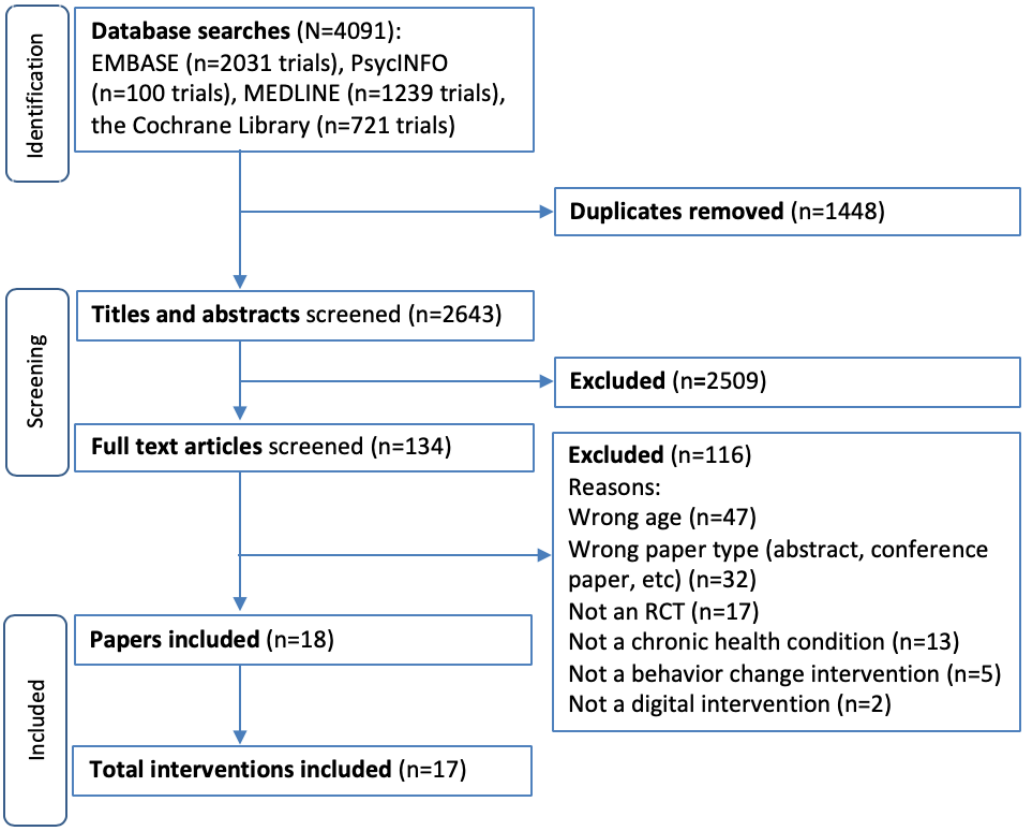
Results

Literature Search and Selection of Studies

After deduplication, 2643 papers were identified from the database searches, of which 18 papers were identified as eligible for inclusion. Two of these papers reported on the same intervention; therefore, we identified 17 digital interventions for the management of chronic health conditions in children aged between 5 and 12 years. Figure 1 displays the PRISMA diagram.

XSL•FO
RenderX

Figure 1. PRISMA flow diagram. RCT: randomized controlled trial.



Population Characteristics

The digital interventions targeted a range of chronic health conditions, including overweight or obesity (n=7), anxiety and preoperative anxiety (n=3), cerebral palsy (n=3), attention-deficit/hyperactivity disorder (ADHD; n=1), type 1 diabetes (n=1), asthma (n=1), and social-emotional problems (n=1). All the interventions included children of key stage 2 age (8-11 years), 13 included children of key stage 3 age (12 years), and 9 included children of key stage 1 age (5-7 years).

Aim 1: Which Digital Interventions Are Effective in Promoting Behavior Change for the Management of the Chronic Health Condition?

Table 2 details the characteristics of the population, intervention, and outcome data (Multimedia Appendix 2). This is presented by their potential effectiveness, based on outcomes and quality assessment (Multimedia Appendix 3).

No interventions were identified as *very promising*.

A total of 5 interventions were identified as *quite promising*; 3 of these were interventions targeting overweight or obesity. An intervention involving exergaming used Kinect and Xbox with additional components of Fitbit step count monitoring and parent-child telehealth sessions with a fitness coach. Compared with the control group, the intervention group showed an increase of 11.4 min of moderate-to-vigorous physical activity (MVPA) per day (95% CI 2.25- 20.55) at 6 months. However, there was no clear evidence of effect on the primary outcome; the reduction in BMI Z-scores was −0.08 (95% CI −0.16 to 0.003) at 6 months, which did not reliably meet the authors' threshold for change (−0.09). An exergaming plus a family-based pediatric weight management program led to an increase of 8.0 (95% CI 0.5-15.4) min of MVPA per day at 4 months, with the trial powered to detect a change of 4.0 to 4.6 min of MVPA per day (MVPA was the primary and behavior change outcome) [54]. A further trial involved 4 training units (2 face-to-face and 2 via Facebook) plus weekly support through a parental WhatsApp group for 12 weeks. This led to a decrease in BMI Z-scores of 0.14 (95% CI −0.28 to −0.003) at 6 months, with the trial powered to detect a decrease of 0.24 (no behavior change outcome was available). The third and fourth *quite promising* interventions were both internet-delivered CBT for children with anxiety disorders, both offering completely web-based modules for parents and children in conjunction with web-based therapist contact [60,63,64]. Both led to an improvement in anxiety as assessed by the clinician severity rating, equating to an estimated change of −077 (95% CI −1.15 to −0.40) at 3 months [60] and −1.16 (95% CI −1.55 to −0.77) at 10 weeks [63].

XSL•FO
RenderX

Table 2. Data on population, interventions, and effectiveness of behavior change outcomes and primary outcomes, grouped by intervention promise.

Categories and reference	Condition; age	Overview of intervention/recipients	Recipients	Behavior change outcome	Primary outcome (if different)
Quite promising					
Ahmad et al (2018) [59]	Overweight or obesity; 8-11 years	Digital component: 2 training units delivered weekly via Facebook. Weekly 1-hour sessions using a parents' WhatsApp group that lasted for 12 weeks. In the WhatsApp group, the researchers posted key information and skills, responded to parent queries, and provided feedback on the adiposity progress of the children. Parents were encouraged to interact with the group WhatsApp group; additional component: 2 half-day face-to-face training sessions	Digital component: parent only; face-to-face component: parent and child	Healthy lifestyle behaviors, children's eating, physical activity, and screen time: effectiveness data not reported	BMI Z-score: the intervention group had a reduced BMI Z-score compared with the control. Net mean difference[a]=−0.14 (95% CI −0.278 to −0.003; *P*=.05)
Jolstedt et al (2018) [60]	Anxiety; 8-12 years	Digital component: ICBT[b]. A web-based program with 12 modules delivered over 12 weeks, consisting of texts, films, illustrations, and exercises and focused mainly on exposure therapy. Limited weekly asynchronous support from a clinician to encourage families to engage in the program. Twelve parent-directed modules, covering parental behaviors, which can maintain anxiety and how to best support their child. Parents helped children with the child-directed modules	Digital component: parent and child; face-to-face component: parent and child	None reported	Anxiety (CSR)[c]: Participants allocated to ICBT showed improvements on the CSR compared with the control. Net mean difference=0.79 (95% CI 0.42-1.16; *P*=.002); the estimated between-group effect size at 12-weeks post-treatment 0.77 (95% CI 0.40-1.15)
Staiano et al (2018) [61]	Overweight or obesity; 10-12 years	Digital component: participants were provided a Kinect and Xbox 360 gaming console, 4 exergames, and a Fitbit Zip to wear during the 24-week period. Steps per day were wirelessly uploaded and reviewed by the fitness coach; additional component: telehealth, consisting of the child and parent meeting with a fitness coach over video chat, on a weekly basis for the first 6 weeks and biweekly thereafter	Digital component: child only; face-to-face component: parent and child	Physical activity: the intervention showed an increase in MVPA[d] compared with the control group; estimated mean difference 11.4 min of MVPA per day (95% CI 2.25-20.55); Dietary habits: there was no evidence of effect	BMI Z-score: there was no clear evidence of effect
Trost et al (2018) [62]	Overweight or obesity; 8-12 years	Digital component: participants were provided a game console and motion capture device (Xbox and Kinect; Microsoft Corporation) and 2 active sports games. No explicit advice or goals were given to any study participant regarding the use of their active gaming tool; additional component: a comprehensive family-based pediatric weight management program	Digital component: child only; face-to-face component: parent and child	Physical activity: the intervention group exhibited a greater increase in MVPA compared with the control group. Net mean difference=8 min of MVPA per day (SE 3.8; 95% CI 0.5-15.4; *P*=.04)	The primary outcome was physical activity (see the behavioral outcomes column)

XSL•FO
RenderX

Categories and reference	Condition; age	Overview of intervention/recipients	Recipients	Behavior change outcome	Primary outcome (if different)
Vigerland et al (2016) [63] and Vigerland et al (2017) [64]	Anxiety; 8-12 years	Digital component: a treatment platform with 11 modules, including reading material, films, animations, illustrations, and exercises. A combined parent-child intervention. Seven parent-directed modules containing information and instructions on how to help their child. Four child-directed modules. Participants had web-based contact with an assigned psychologist or CBT therapist through written messages and written feedback. Three scheduled telephone calls plus additional telephone calls if needed (to increase motivation or problem solve)	Digital component: parent and child; face-to-face component: parent and child	None reported	Anxiety (CSR): the intervention group had a larger improvement on the CSR. Net mean difference=−1.16 (95% CI −0.77 to −1.55)
Possibly promising					
Bul et al (2016) [65]	ADHD; 8-12 years	Digital component: a serious web-based adventure game (Plan-It Commander) developed by health care professionals, researchers, and game experts in collaboration with parents and children with ADHD. A web-based mission-guided game in which principles of behavior therapy and game-based learning were combined. The missions addressed specific skills addressing time management, planning and organizing, and prosocial behavior. Players could access a closed social community (*Space Club*) to stimulate prosocial behavior (eg, helping other players and giving compliments); additional component: treatment as usual	Digital component: child only; face-to-face component: N/A[e]	Time management: the intervention arm showed greater improvements in parent-reported time management skills compared with the control arm. Estimated net mean difference of parent-reported time management=5.98 (95% CI 1.32-10.64) and teacher time management=5.46 (95% CI 1.71-9.20); planning and organizing skills, social skills: no clear evidence of effect	The 3 primary outcomes were: parent-reported time management, parent-reported planning and organization, and parent-reported social skills (see the behavioral outcomes column)
Hsieh et al (2018) [66]	Cerebral palsy; 5-10 years	Digital component: a PC[f] gaming platform. The participants stood in front of the platform and viewed a monitor that displayed 1 of a series of simulated tasks, such as hitting ground rats. The gaming platform handle was loaded, 0.5 to 2.5 lb. This PC gaming platform provided trunk movements in 3 directions: horizontal, vertical, and multidirectional trunk movements	Digital component: child only; face-to-face component: child only	None reported	Measures of postural balance: there was clear evidence of effect of the intervention for 2 of the 7 measures

XSL•FO
RenderX

Categories and reference	Condition; age	Overview of intervention/recipients	Recipients	Behavior change outcome	Primary outcome (if different)
Wantanakorn et al (2018) [67]	Preoperative anxiety; 5-12 years	Digital component: *Children-Friendly Hospital*, a tablet app for pediatric patients who need bone marrow aspiration procedures. It was used to provide medical information: participants started with the cartoon about the procedure then played matching games and practiced the breathing exercise game to reduce anxiety; additional component: usual bone marrow aspiration procedures	Digital component: child only; face-to-face component: N/A	N/A	Preoperative anxiety: anxiety was lower in the intervention group versus the control group. Estimated difference means=−7.71 (95% CI −14.27 to −1.15)
Nonpromising					
Armstrong et al (2017) [68]	Overweight or obesity; 5-12 years	Digital component: daily mobile text messages, based on motivational interviewing, for 12 weeks. Initial texts encouraged parents to set a health behavior goal. In a reply text, the investigators reinforced the most evidence-based goals for BMI reduction (sugar-sweetened beverage reduction, increased physical activity, eating meals at home, and increased vegetable consumption). Subsequent daily texts prompted parents to self-monitor adherence to the goal. Each week for 12 weeks, parents can choose a new goal or continue working on the present goal; additional component: standard care, including monthly lifestyle counseling visits by a physician and dietician	Digital component: parent only; face-to-face component: parent only	Child nutrition habits, activity habits and screen time: no clear evidence of effect of intervention	BMI Z-score: no clear evidence of effect of intervention
Christison et al (2016) [69]	Overweight or obesity; 8-12 years	Digital component: The Exergaming for Health Program is a community-based, multifaceted pediatric weight management program including 1 hour of weekly group exergaming; additional component: classroom curriculum	Digital component: child only; face-to-face component: parent and child	Activity levels, sedentary screen time, and diet: no clear evidence of effect of intervention	BMI Z-score: no clear evidence of effect of intervention
Sanchez et al (2017) [70]	Social-emotional problems; 7-11 years	Digital component: a single-player story-based digital game that requires children to apply specific social-emotional skills to solve social problems encountered in the game. For example, approaching an individual that appears easy to talk to, joining a group at a game in progress, and approaching small groups that appear less easy to talk to. Within each scenario, the player had to make behavioral choices and individualized feedback to choices was given	Digital component: child only; face-to-face component: N/A	Bullying perpetration: no clear evidence of effect of intervention	No primary outcome specified
Unable to assess effectiveness					

XSL•FO
RenderX

Categories and reference	Condition; age	Overview of intervention/recipients	Recipients	Behavior change outcome	Primary outcome (if different)
Burckhardt et al (2018) [71]	Type 1 diabetes; 8-12 years	Digital component: the Dexcom G5O` Mobile continuous glucose monitoring system transmitted glucose levels via Bluetooth to a mobile device that generated alerts. This information could be shared via the *cloud* with up to 5 individuals who could remotely monitor the continuous glucose reading in real time along with the possibility to use individualized alerts. Participants were able to see their sensor glucose levels in real time	Digital component: child only; face-to-face component: N/A	Pilot study (n=14)	No primary outcome specified
Fiks et al (2015) [72]	Asthma; 6-12 years	Digital component: the features of MyAsthma include identification of parents' concerns and goals for asthma treatment; monthly tracking of symptoms, medication side effects, and progress toward goals; educational content; and access to the child's asthma care plan. Parents were encouraged with email reminders to complete monthly portal surveys with input from their affected child. In response to these surveys, families and clinicians received guideline-based decision support that directed them to speak to one another if asthma was not well controlled or if there were side effects, or to continue current therapy	Digital component: parent only; face-to-face component: parent only	Only acceptability or feasibility data	Only acceptability or feasibility data
Hamilton-Shield et al (2014) [73]	Overweight or obesity; 5-11 years	Digital component: Mandolean teaches patients how to eat and recognize satiety. The patient puts a measured portion of food determined by a therapist on the Mandolean (scales and computer), which records and displays, in real-time graphics, the removal of food from the plate as the patient eats. This is compared with a preset eating line and deviation from the training line by eating too quickly or slowly elicits a spoken request from Mandolean to slow down or eat faster. The patient rates their level of satiety, which appears as a dot on screen yielding a *development of satiety* curve allowing comparison of the development of fullness with a *normal* fullness curve again preset on screen. Additional component: standard care comprising of dietary and activity advice by trained practice nurses	Digital component: child only; face-to-face component: parent and child	None of the pilot trial's objectives were met; thus there were no full trial effectiveness results	BMI Z-score: none of the pilot trial's objectives were met; thus there were no full trial results

XSL•FO
RenderX

Categories and reference	Condition; age	Overview of intervention/recipients	Recipients	Behavior change outcome	Primary outcome (if different)
Kassee et al (2017) [74]	Cerebral palsy; 7-12 years	Digital component: a Nintendo Wii U system, 1 Wii MotionPlus remote controller, 1 Wii Nunchuck, and the Wii Sports Resort game to be played at home. Games were designed to promote higher upper-limb activity. Participants were instructed to play games using their affected hand for at least 40 min each day, 5 days a week for 6 weeks (30 days); additional component: parents supervised and recorded sessions and were asked to encourage the child to use their spastic hand as much as possible	Digital component: child only; face-to-face component: parent and child	Pilot study (n=6)	Pilot study (n=6)
Preston et al (2016) [75]	Cerebral palsy; 5-12 years	Digital component: computer-assisted arm rehabilitation gaming used at the child's home. Parents were asked to encourage their children to use the gaming technology for 30 min a day; additional component: a visit at week 3 to offer encouragement and to check the gaming technology system. Usual follow-up treatment	Digital component: child only; face-to-face component: parent and child	None reported	Pilot study (n=15)
Price et al (2015) [76]	Overweight or obesity; 6-12 years	Digital component: text messages to parents to reinforce telephone health behavior coaching. Text messages to promote behavioral self-monitoring and skills training, focused on behaviors, including limiting fast food and eating fruits and vegetables in place of high-calorie snacks. At the time of a well child care visit, digital alerts were sent to pediatricians designed to identify children with a BMI ≥95th percentile. These contained information on how to monitor and support the child; additional component: Well child visit	Digital component: parent only; face-to-face component: parent only	Only acceptability or feasibility data	Only acceptability or feasibility data

[a]Net mean difference indicates the difference in mean change between the intervention and the control arms of the study.

[b]ICBT: internet-delivered cognitive behavioral therapy.

[c]CSR: clinician severity rating.

[d]MVPA: moderate-to-vigorous physical activity.

[e]N/A: not applicable.

[f]PC: personal computer.

In all, 3 interventions were identified as *possibly promising.* A PC game led to improved balance control in children with cerebral palsy on 2 of the 7 measures of balance at 3 months [66]. An internet-based serious game for ADHD led to an improvement in parent- and teacher-rated time management skills at 5 months, but no evidence of improvement on parent- and teacher-rated planning and organization skills or social skills. A tablet app that included an educational animated video, along with games for distraction and to encourage relaxation or breathing exercises for preoperative anxiety [67], led to reduced anxiety scores on the modified Yale Preoperative Anxiety Scale of −7.71 (95% CI −14.27 to −1.15) immediately after the intervention. Although there was evidence of an effect, these studies were limited in scientific quality. There was a lack of transparency around randomization processes, a combination of nonblinded participants, and the use of self-report measures, and none of these trials were prospectively registered.

XSL•FO
RenderX

Three interventions showed no promise; 2 of these were targeting overweight and obesity, 1 was the exergaming plus classroom curriculum, and the other was motivational interviewing delivered via one-way text messaging [68,69]. Neither lead to improvements in behavioral outcomes (screen time, physical activity, and diet) or the primary outcome (BMI Z-scores) at 6 and 3 months, respectively. The other intervention was a video game for social-emotional problems [70], which did not lead to changes in bullying perpetration behavior at 9 weeks.

Six interventions were pilot studies, and they only reported acceptability or feasibility data [72,73,76] or involved small sample sizes (6, 15, and 14) that were not powered to determine effectiveness [71,74,75]. Of these studies, 3 reported that there were no further plans for investigation [72-74] and 1 reported that a larger, fully powered trial was planned for the future [75]. Information on the remaining 2 studies is unknown [71,76].

Adverse events for each study are reported in Table 3. Three studies reported adverse events; these were not severe and or there were similar numbers in the intervention and control arms. Four studies monitored adverse events and reported that no adverse events occurred during the trial. Most studies (n=10) failed to capture adverse events.

Table 3. Summary of adverse events.

References	Details of adverse events
Ahmad et al (2018) [59]	"No adverse events or unintended adverse consequences of the intervention were reported by the participants."
Armstrong et al (2017) [68]	"We observed no adverse events associated with participation in the text message intervention."
Bul et al (2016) [65]	"There were 10 adverse events that could be related to the intervention... All adverse events were of mild (n=5) or moderate (n=5) severity… Examples of adverse events were pain in the fingers, irritability, and headache. An adverse event was a reason to discontinue the study for only one known participant. This participant did not want to play the game anymore because he could not concentrate during his school activities. Sounds reminded him of the game and this consequently distracted and frustrated him. No serious adverse events were reported."
Burckhardt et al (2018) [71]	Did not capture adverse events.
Christison et al (2016) [69]	Did not capture adverse events.
Fiks et al (2015) [72]	Did not capture adverse events.
Hamilton-Shield et al (2014) [73]	"There were no adverse events regarded as serious, unexpected or suspected to be related to the study treatment"
Hsieh et al (2018) [66]	Methods: "no adverse effects were expected in participants in the intervention group." No further details of adverse effects were reported.
Jolstedt et al (2018) [60]	"No severe adverse events were reported in either group... The number of adverse events was similar between the groups." Total reported adverse events: ICBT[a] 17 (26%), ICDP (active control) 16 (25%).
Kassee et al (2017) [74]	Did not capture adverse events.
Preston et al (2016) [75]	"No adverse events were reported."
Price et al (2015) [76]	Did not capture adverse events.
Sanchez et al (2017) [70]	Did not capture adverse events.
Staiano et al (2018) [61]	"Among those randomized to the intervention group, two children reported an injury during gameplay (bruise to the ankle or wrist)." "Two adverse events (bruising) were reported in the GameSquad trial, which is similar to prior exergaming studies reporting minor bruises, hand lacerations and back pain ..."
Trost et al (2018) [62]	Did not capture adverse events.
Vigerland et al (2016) [63] and Vigerland et al (2017) [64]	Did not capture adverse events.
Wantanakorn et al (2018) [67]	Did not capture adverse events.

[a]ICBT: internet-delivered cognitive behavioral therapy.

Aim 2: What Are the Characteristics (Active Ingredients) of Effective Interventions?

We considered the 8 interventions that were classified as promising, quite promising, and possibly promising to represent *promising interventions*.

Recipients

A total of 7 of the 8 interventions had a digital component for the child, and all the interventions involved the child in some capacity (either digital or human component). In all, 5 of the 8 interventions involved the parent in some capacity (either digital or human component).

What Is Being Delivered: BCTs

Table 4 provides the definitions of the BCTs, and Table 5 provides a summary of the BCTs used in promising interventions.

All the promising interventions used more than one BCT. Digital components for the child typically included techniques coded into the following BCT categories: *feedback and monitoring*,

XSL•FO
RenderX

shaping knowledge, *repetition and substitution*, and *reward and threat* (we note that none used *threat*, but this is the overarching BCT taxonomy category label). Digital components for the parent typically included *goals and planning*, *social support*, and *natural consequences*.

The most promising interventions were for overweight or obesity (3 studies) and anxiety (2 studies). All 3 of the promising overweight or obesity interventions included a face-to-face component for both the parent and the child. Two interventions included a digital component for the child, both using the BCT repetition and substitution. Only 1 intervention had a digital component for the parent.

Both promising anxiety interventions included digital and face-face elements, all of which involved both the child and the parent. Both interventions used the following BCTs in the digital component: goals and planning (child and parent components), shaping knowledge (child and parent components), feedback and monitoring (parent component), and associations (child component).

We acknowledge that there may have been more BCTs included in the intervention; however, we were unable to code these as they were not explicitly reported in the paper. Furthermore, it was often unclear as to whether the BCT was delivered to the parent or the child and by what means it was planned to take effect. In some cases, we believe that the BCTs were directed at the parent, with the parent then eliciting behavior change in the child. However, none of the papers addressed this level of complexity; they did not describe this mechanism of change nor did they include a parent behavior change outcome measure.

Table 4. Definitions of behavior change techniques.

BCT[a] categories[b]	Definitions
Goals and planning	Includes setting and reviewing goals defined in terms of the behavior (eg, physical activity) or the outcome (eg, weight loss); problem-solving to overcome barriers and or increase facilitators; and detailed action planning of the behavior, considering the context, frequency, duration, and intensity of the behavior
Feedback and monitoring	Includes observing or recording the behavior or the outcome either by the recipient (self-monitoring) or by others; feedback on the performance of the behavior or the outcome
Shaping knowledge	Includes advising how to perform the behavior, the factors that reliably predict performance of the behavior, alternatives to unhealthy behaviors, and how to carry out behavioral experiments
Repetition and substitution	Includes practicing the behavior in a context or at a time when the performance may not be necessary to increase habit and skill; setting easy-to-perform tasks, making them increasingly difficult, but achievable, until the behavior is performed
Reward and threat	Includes using material (eg, money and vouchers) or social (eg, praise) incentives and rewards for the behavior or outcome; informing that future punishment or removal of reward will be a consequence of performance of an unwanted behavior
Social support	Includes advising, arranging, or providing social support (eg, from friends, relatives, colleagues, "buddies," or staff) for practical and or emotional reasons
Natural consequences	Includes providing information (eg, written, verbal, visual) about the health, social, emotional, or environmental consequences of performing the behavior; using methods to emphasize the consequences
Associations	Includes introducing environmental or social stimulus to prompt or cue behavior; reducing situations in which unwanted behavior can be rewarded; systematic confrontation with a feared stimulus to reduce the response to a later encounter; and presenting a neutral stimulus jointly with a stimulus that already elicits the behavior repeatedly until the neutral stimulus elicits that behavior

[a]BCT: behavioral change techniques.

[b]The study by Michie et al [43] provides a full description of all BCTs.

XSL•FO
RenderX

Table 5. Characteristics of promising interventions.

Characteristics	Child recipient–digital component (n=7), n (%)	Child recipient–human component (n=6), n (%)	Parent or caregiver–digital component (n=3), n (%)	Parent or caregiver–human component (n=5), n (%)
Digital mode of delivery				
Tailored	1 (14)	N/A[a]	1 (33)	N/A
Guided	5 (71)[b]	N/A	3 (100)[b]	N/A
Gaming features	5 (71)[b]	N/A	2 (67)[b]	N/A
Behavior change technique				
Goals and planning	3 (43)	3 (50)[b]	3 (100)[b]	4 (8)[b]
Feedback and monitoring	4 (57)[b]	4 (67)[b]	0 (0)	4 (80)[b]
Social support	2 (29)	3 (5)[b]	3 (100)[b]	4 (80)[b]
Shaping knowledge	4 (57)[b]	3 (50)[b]	1 (33)	3 (60)[b]
Natural consequences	0 (0)	0 (0)	2 (67)[b]	1 (20)
Comparison of behavior	2 (29)	2 (33)	1 (33)	1 (20)
Associations	2 (29)	1 (17)	1 (33)	1 (20)
Repetition and substitution	5 (71)[b]	0 (0)	0 (0)	0 (0)
Comparison of outcomes	0 (0)	1 (17)	1 (33)	1 (20)
Reward and threat	6 (86)[b]	2 (33)	1 (33)	1 (20)
Regulation	1 (14)	0 (0)	0 (0)	0 (0)
Antecedents	1 (14)	2 (33)	1 (33)	2 (40)
Identity	0 (0)	0 (0)	1 (33)	1 (20)
Scheduled consequences	0 (0)	0 (0)	0 (0)	0 (0)
Self-belief	0 (0)	1 (17)	1 (33)	2 (40)
Covert learning	0 (0)	0 (0)	0 (0)	0 (0)

[a]N/A: not applicable.

[b]≥50% of interventions using the characteristic.

How Is the Content Delivered: Mode of Delivery

A total of 5 of the 7 interventions with child digital components used gaming features. All the parent digital components and 5 of the child digital components were guided. In all, 3 digital interventions involving parents and 1 digital intervention for the child were tailored.

Theoretical Basis

Half of these papers reported the use of theory in the intervention: social cognitive (n=2) and CBT (n=2).

Modality

A total of 6 of the 8 interventions were first wave (purely behavioral) interventions, and 2 were second wave (cognitive-behavioral) interventions. There were no third wave interventions.

Aim 3: What Are the Users' Experience of the Digital Intervention?

Only 3 of the studies included qualitative data on users' experiences and views of the intervention [68,72,73]. One study evaluated the family experience in a preceding pilot study [77]. A table of the raw qualitative data and themes are available (Multimedia Appendix 4).

Themes

Parental Involvement

Parents talked about the interventions improving their knowledge ("made me more aware") and shaping their behavior, which in turn led to the child's behavior change ("it does make me stop him and sit him down and make him eat the breakfast"). Some commented on the problems of parent-led interventions and how a health professional, who is external to the parent-child relationship, is important to encourage the child's behavior

XSL•FO
RenderX

change ("I think some kids will listen to their doctor better than their parents").

Connection With a Health Professional Is Important for Engagement

Digital interventions were seen to facilitate *convenient* communication with a health care professional. There was a desire to share information between parents and clinicians ("It should go back somehow to the paediatrician") to increase families' motivation to engage with interventions. The involvement of a health professional was also viewed as important in engaging the child ("I think some kids will listen to their doctor better than their parents").

Technological Affordances and Barriers

Parents commented on the technologies being quick, easy, and possible to integrate into everyday life. However, others commented on practical challenges such as the cost, lack of familiarity, and difficultly to use. Users commented on the fixed nature of the technology, which meant that it was not personalized to their individual preferences or needs ("but I really want to focus on these") and did not deliver content with ongoing relevance that would maintain engagement over time ("I think enthusiasm's gone off").

Child-Centered Design

Children commented on some of the interventions being enjoyable ("I like the electronic stuff"). However, in other cases, the material was not understood by the child ("It's really confusing" and "I don't know how much [child] actually understands"), it was not acceptable to children ("boring" and "annoying"), and they expressed a wish for features such as personalization in the design.

Discussion

Principal Findings

To the best of our knowledge, this is the first review to identify effective digital interventions for younger children, report the characteristics of promising interventions, and describe the user's experience of digital interventions. Of the 17 eligible interventions, we only identified 5 that had a beneficial effect and had a low risk of bias; 3 targeted overweight or obesity, using exergaming or social media with additional human support, and 2 targeted anxiety, using web-based CBT with therapist support.

Characteristics of promising digital interventions included gaming features in the child digital component and having additional therapist support (guided digital interventions). Digital components for the child typically used the BCTs [43] *feedback and monitoring*, *shaping knowledge*, *repetition and substitution*, and *reward*. Most were purely behavioral interventions (first wave), with only a quarter using CBT (second wave) and none using third wave approaches; half of the interventions had a theoretical basis (social cognitive theory and CBT). Over 60% involved the parents in the intervention.

Only 3 papers used qualitative methods to explore the users' experience of digital intervention. These studies reported the affordances of digital interventions, including ease of use, integration into daily life, and the ability to enhance communication with a health professional. However, a lack of personalization, technical problems, and cost issues posed challenges to families. The qualitative data indicated how the content (eg, language and concepts) and design could be improved for younger users.

Strengths and Limitations

We included a range of chronic health conditions, which enabled us to review a larger number of interventions and identify patterns or commonalties of promising interventions. Spanning health conditions makes these findings relevant to a wide audience of researchers working in the field of digital interventions. We focused on RCTs because they have the strongest study design and are most likely to be adopted in clinical care [78]. This review focused on the outcomes that were most important to our research question (behavioral outcomes) and most important for that particular study (the primary outcome). It was outside the scope of this paper to review all the possible outcomes, such as health status or symptoms of the disease, quality of life, and knowledge.

Guidance was followed on how to report effectiveness in narrative reviews [51]. We extracted a common statistic to show the size and direction of effect, and where possible, we placed results in the context of clinically meaningful change [79]. Strengths of narrative synthesis include richer exploration of more complex questions, exploring both effectiveness (aim 1) and what "might explain differences in direction and size of effect... how and why interventions have or do not have an effect" (aim 2) [51]. We increased the rigor of presenting characteristics of interventions by using established coding systems and taxonomies for BCTs [43], modality [56], mode of delivery [18], ages [53], and population type [80]. We also considered parental and child components separately, which is important for this younger population.

A limitation of this review is that we only included RCTs. Although observational studies and nonrandomized trials could have provided additional information on the characteristics and effectiveness of digital interventions for this population, we excluded these study designs as they have a greater potential for risk of bias [81]. Although we believe that our search strategy (which included the terms "Randomized Controlled Trial," "Trial," and "Clinical Study") was broad enough to identify different RCT designs, it is possible that we may not have identified some designs specifically used in the evaluation of digital interventions, such as micro randomized trials. We also restricted our search to papers published after 2015. We chose this strategy as digital health is a rapidly changing field, and recently conducted studies are likely to be the most relevant. We excluded studies that included our target age group (5-12 years) but also included older and younger children (eg, 5-18 years). Although it is possible that these studies could have been stratified by age, it was not feasible to contact authors to request these stratified data. As expected, the broad scope of this review led to heterogeneity across studies (in terms of population, intervention, and outcome), meaning formal meta-analysis was not possible; therefore, we selected the most appropriate method,

XSL•FO
RenderX

narrative synthesis. Although potential limitations to narrative synthesis include a lack of transparency and reproducibility and being subject to author interpretation [52], we mitigated this by prospectively registering our protocol, with specified outcomes, and following narrative synthesis guidelines [51]. To identify the characteristics of effective interventions, we reviewed both *quite promising* and *possibly promising* interventions and acknowledged that the *possible promising* interventions were of poorer scientific quality. Due to the small number of qualitative studies, we did not conduct full meta-ethnography [82] to synthesize qualitative data, and we did not undertake critical appraisal. However, to increase the transparency of our qualitative summary, we reported the raw data from the papers along with the themes developed by us.

Implications for Developing, Evaluating, and Implementing Digital Interventions for Children With a Chronic Health Condition

Clinical Implications

This review identified promising exergaming and social media interventions for children with obesity or overweight and web-based CBT for children with anxiety. There is potential for these to be implemented in clinical practice with further surveillance, monitoring, and long-term follow-up [40]. These findings are consistent with a previous systematic review that concluded that digital game-based interventions should be considered as methods to promote physical activity among children, but that there is a need for further, high - quality research that provides more sound evidence about clinical practice and health promotion [83]. This study extends a previous meta-analyses investigating digital interventions for children with anxiety, which concluded that the quality of studies was low (lack of blinding, use of subjective outcome measures, waiting list comparison groups, and relatively small samples) and that the effect is uncertain for younger children [36]. Our review updates this work, identifying 2 interventions with promise. These trials had sample sizes of 131 and 93, and both were prospectively registered trials with prespecified primary outcomes; 1 trial used a blinded outcome assessor for the primary outcome and an active control.

Implications for Developing and Evaluating Interventions

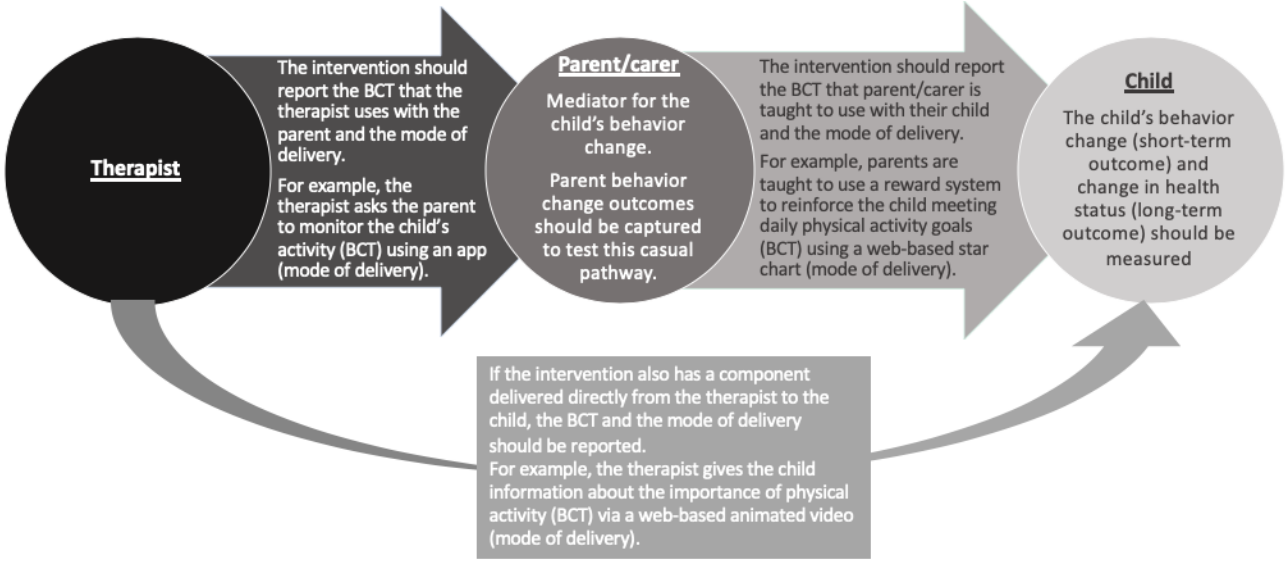
This work highlights characteristics that may be beneficial when developing digital interventions for younger populations. The finding that purely behavioral interventions (first wave, not including cognitive components) are common in promising interventions is consistent with developmental theory; children tend to be limited to concrete thought [57]. There were fewer CBT (second wave) interventions, possibly because elements of CBT require abstract thinking, which may be beyond the cognitive abilities of children aged <8 years [57]. Similarly, third wave interventions also include abstract concepts such as metacognition. The lack of third wave approaches may also be explained by the fact that this is a relatively new approach for children. As such, concrete interventions focused on behavioral recommendations may be more appropriate [84]. Caregivers are commonly involved in promising interventions. This is also consistent with developmental theory, which highlights the important role of caregivers in structuring the child's environment and shaping the child's behavior [84,85]. Gaming features have been used in many promising interventions. Digital games can be adapted to the developmental level and can effectively engage younger users in medical education and treatment, as they are typically more visually oriented, involve appealing exploration, and are perceived as fun [17]. Consistent with the literature, guided interventions were common in promising interventions and have been identified as a moderating factor that can influence therapeutic outcomes and engagement [86,87].

Guidelines encourage standardized reporting of interventions to ensure transparency and reproducibility [43,88]. On the basis of our findings, we have developed recommendations for increasing the clarity of interventions with parental involvement. Interventions with both a child and a parent recipient have a complex model of behavior change; it is likely that the therapist aims to shape the behavior of the parent, with the expectation that the parent will change the behavior of the child. Studies in this review failed to explicitly differentiate the BCTs used by the therapist for parental behavior change and the behavior techniques used by the parent for child behavior change. Furthermore, none of the studies in this review captured a parental behavior change outcome measure, when this may be on the causal pathway to the child's behavior change. This recommendation is consistent with guidelines on process evaluation; outcome measures should be used to test the causal mechanism of the intervention. Figure 2 illustrates the relationship between therapist, parent, and child, detailing our recommendations for how these interventions could be conceptualized and reported.

The low number of promising interventions demonstrates the need to better understand the perspective of those receiving interventions. Few studies have conducted qualitative research to explore the user's experiences. Qualitative methods, such as the person-based approach [89], base the development and evaluation of digital interventions on an in-depth understanding of the perspectives of the people who will use the intervention. This can lead to interventions that are more meaningful, usable, feasible, and engaging in improving uptake and adherence and maximizing effectiveness [89].

XSL•FO
RenderX

Figure 2. Conceptualising and reporting interventions involving both the parent/caregiver and the child. BCT: behavior change techniques.



Conclusions

Of the 17 interventions, we only identified 5 with promise (those with a beneficial effect and low risk of bias). Using qualitative methods during digital intervention development and evaluation may lead to more meaningful, usable, feasible, and engaging interventions, especially for this under-researched younger population. Promising interventions were exergaming and social media for obesity or overweight and a web-based CBT platform for anxiety. We identified characteristics that could be considered when developing digital interventions for younger children: involvement of parents, gaming features, additional therapist support, behavioral (rather than cognitive) approaches, and particular BCTs (*feedback and monitoring*, *shaping knowledge*, *repetition and substitution*, and *reward*). We suggest a model for improving the conceptualization and reporting of behavioral interventions involving children and parents.

Acknowledgments

AB is funded by a National Institute for Health Research (NIHR), (NIHR Doctoral Research Fellowship, DRF-DRF-2017-10-169) for this research project. ML is funded by a National Institute for Health Research (NIHR), (NIHR Doctoral Research Fellowship, DRF-2016-09-021) for this research project. EC was funded by a National Institute for Health Research (NIHR), (NIHR Senior Research Fellowship, SRF-2013-06-013) for this research project. This publication presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search Strategy.
[DOCX File , 22 KB - jmir_v22i7e16924_app1.docx]

Multimedia Appendix 2
Full data extraction table.
[DOCX File , 41 KB - jmir_v22i7e16924_app2.docx]

Multimedia Appendix 3
Summary of risk of bias assessment.
[DOCX File , 32 KB - jmir_v22i7e16924_app3.docx]

Multimedia Appendix 4
The users experience and views on the digital intervention; raw qualitative data and themes.
[DOCX File , 18 KB - jmir_v22i7e16924_app4.docx]

XSL•FO
RenderX

References

1. Lightfoot J, Mukherjee S, Sloper P. Supporting pupils with special health needs in mainstream schools: policy and practice. Child Soc 2001 Apr;15(2):57-69. [doi: 10.1002/chi.603]
2. Perrin JM, Anderson LE, van Cleave J. The rise in chronic conditions among infants, children, and youth can be met with continued health system innovations. Health Aff (Millwood) 2014 Dec;33(12):2099-2105. [doi: 10.1377/hlthaff.2014.0832] [Medline: 25489027]
3. Mokkink LB, van der Lee JH, Grootenhuis MA, Offringa M, Heymans HSA, Dutch National Consensus Committee Chronic Diseases and Health Conditions in Childhood. Defining chronic diseases and health conditions in childhood (0-18 years of age): national consensus in the Netherlands. Eur J Pediatr 2008 Dec;167(12):1441-1447. [doi: 10.1007/s00431-008-0697-y] [Medline: 18340463]
4. Eccleston C, Fisher E, Law E, Bartlett J, Palermo T. Psychological interventions for parents of children and adolescents with chronic illness. Cochrane Database Syst Rev 2015 Apr 15(4):CD009660 [FREE Full text] [doi: 10.1002/14651858.CD009660.pub3] [Medline: 25874881]
5. Bennett S, Shafran R, Coughtrey A, Walker S, Heyman I. Psychological interventions for mental health disorders in children with chronic physical illness: a systematic review. Arch Dis Child 2015 Apr;100(4):308-316 [FREE Full text] [doi: 10.1136/archdischild-2014-307474] [Medline: 25784736]
6. Depression in Children and Young People: Identification and Management. The National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/guidance/ng134 [accessed 2020-07-01]
7. Turner-Cobb J. Child Health Psychology: A Biopsychosocial Perspective. London, UK: SAGE Publications; 2013.
8. Marmot M, Bell R. Fair society, healthy lives. Public Health 2012 Sep;126(Suppl 1):S4-10. [doi: 10.1016/j.puhe.2012.05.014] [Medline: 22784581]
9. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. Ann Epidemiol 2006 Feb;16(2):91-104. [doi: 10.1016/j.annepidem.2005.06.053] [Medline: 16257232]
10. Kessler RC, McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, et al. Childhood adversities and adult psychopathology in the WHO world mental health surveys. Br J Psychiatry 2010 Nov;197(5):378-385 [FREE Full text] [doi: 10.1192/bjp.bp.110.080499] [Medline: 21037215]
11. Wadsworth M, Kuh D. Childhood influences on adult health: a review of recent work from the British 1946 national birth cohort study, the MRC national survey of health and development. Paediatr Perinat Epidemiol 1997 Jan;11(1):2-20. [doi: 10.1046/j.1365-3016.1997.d01-7.x] [Medline: 9018723]
12. Wright CM, Parker L, Lamont D, Craft AW. Implications of childhood obesity for adult health: findings from thousand families cohort study. Br Med J 2001 Dec 1;323(7324):1280-1284 [FREE Full text] [doi: 10.1136/bmj.323.7324.1280] [Medline: 11731390]
13. Langford R, Bonell C, Jones H, Pouliou T, Murphy S, Waters E, et al. The World Health Organization's health promoting schools framework: a cochrane systematic review and meta-analysis. BMC Public Health 2015 Feb 12;15:130 [FREE Full text] [doi: 10.1186/s12889-015-1360-y] [Medline: 25886385]
14. Michie S, West R, Sheals K, Godinho C. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Transl Behav Med 2018 Mar 1;8(2):212-224 [FREE Full text] [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
15. Thabrew H, Stasiak K, Hetrick S, Wong S, Huss J, Merry S. E-Health interventions for anxiety and depression in children and adolescents with long-term physical conditions. Cochrane Database Syst Rev 2018 Aug 15;8:CD012489 [FREE Full text] [doi: 10.1002/14651858.CD012489.pub2] [Medline: 30110718]
16. Hilliard ME, Powell PW, Anderson BJ. Evidence-based behavioral interventions to promote diabetes management in children, adolescents, and families. Am Psychol 2016 Oct;71(7):590-601 [FREE Full text] [doi: 10.1037/a0040359] [Medline: 27690487]
17. Charlier N, Zupancic N, Fieuws S, Denhaerynck K, Zaman B, Moons P. Serious games for improving knowledge and self-management in young people with chronic conditions: a systematic review and meta-analysis. J Am Med Inform Assoc 2016 Jan;23(1):230-239. [doi: 10.1093/jamia/ocv100] [Medline: 26186934]
18. Carey R, Jenkins E, Williams P, Evans F, Horan M, Johnston M, et al. A taxonomy of modes of delivery of behaviour change interventions: development and evaluation. Eur Health Psychol 2017:-.
19. Infantino A, Donovan CL, March S. A randomized controlled trial of an audio-based treatment program for child anxiety disorders. Behav Res Ther 2016 Apr;79:35-45. [doi: 10.1016/j.brat.2016.02.007] [Medline: 26950257]
20. Silva C, Fassnacht DB, Ali K, Gonçalves S, Conceição E, Vaz A, et al. Promoting health behaviour in Portuguese children via short message service: the efficacy of a text-messaging programme. J Health Psychol 2015 Jun;20(6):806-815. [doi: 10.1177/1359105315577301] [Medline: 26032797]
21. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. J Clin Psychol 2004 Mar;60(3):269-283. [doi: 10.1002/jclp.10263] [Medline: 14981791]

XSL•FO
RenderX

22. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
23. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, Mackenzie AM, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51 [FREE Full text] [doi: 10.2196/jmir.2814] [Medline: 24550161]
24. Wang J, Wang Y, Wei C, Yao N, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
25. McDermott MS, While AE. Maximizing the healthcare environment: a systematic review exploring the potential of computer technology to promote self-management of chronic illness in healthcare settings. Patient Educ Couns 2013 Jul;92(1):13-22. [doi: 10.1016/j.pec.2013.02.014] [Medline: 23566427]
26. Jones KR, Lekhak N, Kaewluang N. Using mobile phones and short message service to deliver self-management interventions for chronic conditions: a meta-review. Worldviews Evid Based Nurs 2014 Apr;11(2):81-88. [doi: 10.1111/wvn.12030] [Medline: 24597522]
27. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
28. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
29. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
30. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009 Jun;34(5):495-510. [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
31. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
32. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
33. Pai A, McGrady M. Systematic review and meta-analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. J Pediatr Psychol 2014 Sep;39(8):918-931 [FREE Full text] [doi: 10.1093/jpepsy/jsu038] [Medline: 24952359]
34. Bogosian A, van Vliet LM, Craig G, Fraser LK, Turner-Cobb JM. Minding the gap and moving forward in children's long-term conditions: a vital role for health psychology. Br J Health Psychol 2016 Sep;21(3):493-498. [doi: 10.1111/bjhp.12193] [Medline: 27028293]
35. Tang W, Zhang L, Ai Y, Li Z. Efficacy of internet-delivered cognitive-behavioral therapy for the management of chronic pain in children and adolescents: a systematic review and meta-analysis. Medicine (Baltimore) 2018 Sep;97(36):e12061 [FREE Full text] [doi: 10.1097/MD.0000000000012061] [Medline: 30200086]
36. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
37. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
38. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
39. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
40. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new medical research council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
41. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. Soc Sci Med 2010 May;70(10):1618-1625. [doi: 10.1016/j.socscimed.2010.01.039] [Medline: 20207464]

XSL•FO
RenderX

42. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
43. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
44. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
45. Merikangas KR, Calkins ME, Burstein M, He J, Chiavacci R, Lateef T, et al. Comorbidity of physical and mental disorders in the neurodevelopmental genomics cohort study. Pediatrics 2015 Apr;135(4):e927-e938 [FREE Full text] [doi: 10.1542/peds.2014-1444] [Medline: 25755242]
46. Das P, Naylor C, Majeed A. Bringing together physical and mental health within primary care: a new frontier for integrated care. J R Soc Med 2016 Oct;109(10):364-366 [FREE Full text] [doi: 10.1177/0141076816665270] [Medline: 27729592]
47. Torpy JM, Campbell A, Glass RM. JAMA patient page. Chronic diseases of children. J Am Med Assoc 2010 Feb 17;303(7):682. [doi: 10.1001/jama.303.7.682] [Medline: 20159879]
48. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. Br Med J 2000 Jul 29;321(7256):255-256 [FREE Full text] [doi: 10.1136/bmj.321.7256.255] [Medline: 10915111]
49. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
50. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
51. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers N, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme. Lancaster University. 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2020-07-01]
52. Campbell M, Katikireddi SV, Sowden A, McKenzie JE, Thomson H. Improving conduct and reporting of narrative synthesis of quantitative data (ICONS-QUANT): protocol for a mixed methods study to develop a reporting guideline. BMJ Open 2018 Feb 26;8(2):e020064. [doi: 10.1136/bmjopen-2017-020064]
53. The National Curriculum in England: Key Stages 1 and 2 Framework Document. Government of UK Developer Documentation. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf2013 [accessed 2019-02-01]
54. Andrade C. The primary outcome measure and its importance in clinical trials. J Clin Psychiatry 2015 Oct;76(10):e1320-e1323 [FREE Full text] [doi: 10.4088/JCP.15f10377] [Medline: 26528658]
55. Loke YK, Price D, Herxheimer A, Cochrane Adverse Effects Methods Group. Systematic reviews of adverse effects: framework for a structured approach. BMC Med Res Methodol 2007 Jul 5;7:32 [FREE Full text] [doi: 10.1186/1471-2288-7-32] [Medline: 17615054]
56. Kahl KG, Winter L, Schweiger U. The third wave of cognitive behavioural therapies: what is new and what is effective? Curr Opin Psychiatry 2012 Nov;25(6):522-528. [doi: 10.1097/YCO.0b013e328358e531] [Medline: 22992547]
57. Grave J, Blissett J. Is cognitive behavior therapy developmentally appropriate for young children? A critical review of the evidence. Clin Psychol Rev 2004 Aug;24(4):399-420. [doi: 10.1016/j.cpr.2004.03.002] [Medline: 15245828]
58. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
59. Ahmad N, Shariff ZM, Mukhtar F, Lye M. Family-based intervention using face-to-face sessions and social media to improve Malay primary school children's adiposity: a randomized controlled field trial of the Malaysian REDUCE programme. Nutr J 2018 Aug 2;17(1):74 [FREE Full text] [doi: 10.1186/s12937-018-0379-1] [Medline: 30071855]
60. Jolstedt M, Wahlund T, Lenhard F, Ljótsson B, Mataix-Cols D, Nord M, et al. Efficacy and cost-effectiveness of therapist-guided internet cognitive behavioural therapy for paediatric anxiety disorders: a single-centre, single-blind, randomised controlled trial. Lancet Child Adolesc Health 2018 Nov;2(11):792-801. [doi: 10.1016/S2352-4642(18)30275-X] [Medline: 30241993]
61. Staiano AE, Beyl RA, Guan W, Hendrick CA, Hsia DS, Newton RL. Home-based exergaming among children with overweight and obesity: a randomized clinical trial. Pediatr Obes 2018 Nov;13(11):724-733 [FREE Full text] [doi: 10.1111/ijpo.12438] [Medline: 30027607]
62. Trost SG, Sundal D, Foster GD, Lent MR, Vojta D. Effects of a pediatric weight management program with and without active video games a randomized trial. JAMA Pediatr 2014 May;168(5):407-413. [doi: 10.1001/jamapediatrics.2013.3436] [Medline: 24589566]

XSL•FO
RenderX

63. Vigerland S, Ljótsson B, Thulin U, Öst LG, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: a randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
64. Vigerland S, Serlachius E, Thulin U, Andersson G, Larsson J, Ljótsson B. Long-term outcomes and predictors of internet-delivered cognitive behavioral therapy for childhood anxiety disorders. Behav Res Ther 2017 Mar;90:67-75. [doi: 10.1016/j.brat.2016.12.008] [Medline: 28012300]
65. Bul KC, Kato PM, van der Oord S, Danckaerts M, Vreeke LJ, Willems A, et al. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: a randomized controlled trial. J Med Internet Res 2016 Feb 16;18(2):e26 [FREE Full text] [doi: 10.2196/jmir.5173] [Medline: 26883052]
66. Hsieh H. Effects of a gaming platform on balance training for children with cerebral palsy: erratum. Pediatr Phys Ther 2019 Jan;31(1):129. [doi: 10.1097/PEP.0000000000000582] [Medline: 30557292]
67. Wantanakorn P, Harintajinda S, Chuthapisith J, Anurathapan U, Rattanatamrong P. A new mobile application to reduce anxiety in pediatric patients before bone marrow aspiration procedures. Hosp Pediatr 2018 Oct;8(10):643-650 [FREE Full text] [doi: 10.1542/hpeds.2018-0073] [Medline: 30213798]
68. Armstrong S, Mendelsohn A, Bennett G, Taveras EM, Kimberg A, Kemper AR. Texting motivational interviewing: a randomized controlled trial of motivational interviewing text messages designed to augment childhood obesity treatment. Child Obes 2018 Jan;14(1):4-10. [doi: 10.1089/chi.2017.0089] [Medline: 29019418]
69. Christison AL, Evans TA, Bleess BB, Wang H, Aldag JC, Binns HJ. Exergaming for health: a randomized study of community-based exergaming curriculum in pediatric weight management. Games Health J 2016 Dec;5(6):413-421. [doi: 10.1089/g4h.2015.0097] [Medline: 27911621]
70. Sanchez R, Brown E, Kocher K, DeRosier M. Improving children's mental health with a digital social skills development game: a randomized controlled efficacy trial of adventures aboard the SS Grin. Games Health J 2017 Feb;6(1):19-27. [doi: 10.1089/g4h.2015.0108] [Medline: 28051877]
71. Burckhardt M, Chetty T, Smith GJ, Adolfsson P, de Bock M, Jones TW, et al. Use of continuous glucose monitoring trends to facilitate exercise in children with type 1 diabetes. Diabetes Technol Ther 2019 Jan;21(1):51-55. [doi: 10.1089/dia.2018.0292] [Medline: 30620642]
72. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
73. Hamilton-Shield J, Goodred J, Powell L, Thorn J, Banks J, Hollinghurst S, et al. Changing eating behaviours to treat childhood obesity in the community using Mandolean: the community mandolean randomised controlled trial (ComMando)--a pilot study. Health Technol Assess 2014 Jul;18(47):i-xxiii, 1 [FREE Full text] [doi: 10.3310/hta18470] [Medline: 25043221]
74. Kassee C, Hunt C, Holmes MW, Lloyd M. Home-based Nintendo Wii training to improve upper-limb function in children ages 7 to 12 with spastic hemiplegic cerebral palsy. J Pediatr Rehabil Med 2017 May 17;10(2):145-154. [doi: 10.3233/PRM-170439] [Medline: 28582885]
75. Preston N, Weightman A, Gallagher J, Levesley M, Mon-Williams M, Clarke M, et al. A pilot single-blind multicentre randomized controlled trial to evaluate the potential benefits of computer-assisted arm rehabilitation gaming technology on the arm function of children with spastic cerebral palsy. Clin Rehabil 2016 Oct;30(10):1004-1015. [doi: 10.1177/0269215515604699] [Medline: 26370148]
76. Price S, Ferisin S, Sharifi M, Steinberg D, Bennett G, Wolin KY, et al. Development and implementation of an interactive text messaging campaign to support behavior change in a childhood obesity randomized controlled trial. J Health Commun 2015;20(7):843-850. [doi: 10.1080/10810730.2015.1018582] [Medline: 25996181]
77. Sharifi M, Dryden EM, Horan CM, Price S, Marshall R, Hacker K, et al. Leveraging text messaging and mobile technology to support pediatric obesity-related behavior change: a qualitative study using parent focus groups and interviews. J Med Internet Res 2013 Dec 6;15(12):e272 [FREE Full text] [doi: 10.2196/jmir.2780] [Medline: 24317406]
78. Hoffmann T, English T, Glasziou P. Reporting of interventions in randomised trials: an audit of journal instructions to authors. Trials 2014 Jan 14;15:20 [FREE Full text] [doi: 10.1186/1745-6215-15-20] [Medline: 24422788]
79. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989 Dec;10(4):407-415. [doi: 10.1016/0197-2456(89)90005-6] [Medline: 2691207]
80. Mrazek PJ, Haggerty RJ. Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. Washington, DC: National Academy Press; 1994.
81. Higgins JP, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. Res Synth Methods 2013 Mar;4(1):12-25. [doi: 10.1002/jrsm.1056] [Medline: 26053536]
82. Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R. Using meta ethnography to synthesise qualitative research: a worked example. J Health Serv Res Policy 2002 Oct;7(4):209-215. [doi: 10.1258/135581902320432732] [Medline: 12425780]

XSL•FO
RenderX

83. Pakarinen A, Parisod H, Smed J, Salanterä S. Health game interventions to enhance physical activity self-efficacy of children: a quantitative systematic review. J Adv Nurs 2017 Apr;73(4):794-811. [doi: 10.1111/jan.13160] [Medline: 27688056]
84. Erickson SJ, Gerstle M, Feldstein SW. Brief interventions and motivational interviewing with children, adolescents, and their parents in pediatric health care settings: a review. Arch Pediatr Adolesc Med 2005 Dec;159(12):1173-1180. [doi: 10.1001/archpedi.159.12.1173] [Medline: 16330743]
85. Bronfenbrenner U. The Ecology of Human Development: Experiments by Nature and Design. Cambridge, MA: Harvard University Press; 1979.
86. Rickwood D, Bradford S. The role of self-help in the treatment of mild anxiety disorders in young people: an evidence-based review. Psychol Res Behav Manag 2012;5:25-36 [FREE Full text] [doi: 10.2147/PRBM.S23357] [Medline: 22427736]
87. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
88. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
89. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
BCT: behavior change technique
CBT: cognitive behavioral therapy
MVPA: moderate-to-vigorous physical activity
NIHR: National Institute for Health Research
PC: personal computer
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by D Estrin, J Brož, V Rocío, H Jin; comments to author 10.03.20; revised version received 30.04.20; accepted 20.05.20; published 31.07.20.

Please cite as:
Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, Smith L, Briggs E, Loades M, Crawley E
Digital Behavior Change Interventions for Younger Children With Chronic Health Conditions: Systematic Review
J Med Internet Res 2020;22(7):e16924
URL: http://www.jmir.org/2020/7/e16924/
doi:10.2196/16924
PMID:32735227

©Amberly Brigden, Emma Anderson, Catherine Linney, Richard Morris, Roxanne Parslow, Teona Serafimova, Lucie Smith, Emily Briggs, Maria Loades, Esther Crawley. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Artificial Intelligence and Health Technology Assessment: Anticipating a New Level of Complexity

Hassane Alami[1,2,3], MScPH, MScHP, PhD; Pascale Lehoux[1,2], MSc, PhD; Yannick Auclair[3], MSc, PhD; Michèle de Guise[3], MD, IMHL, FRCPC; Marie-Pierre Gagnon[4,5], MSc, PhD; James Shaw[6,7], PT, PhD; Denis Roy[3], MD, MPH, MSc, FRCPC; Richard Fleet[4,8,9], MD, PhD, CCFP-EM; Mohamed Ali Ag Ahmed[10], MD, MPH, PhD; Jean-Paul Fortin[4,11], MD, MPH, MBA, FRCPC

[1]Public Health Research Center, Université de Montréal, Montreal, QC, Canada
[2]Department of Health Management, Evaluation and Policy, École de santé publique de l'Université de Montréal, Montreal, QC, Canada
[3]Institut national d'excellence en santé et services sociaux, Montréal, QC, Canada
[4]Research Center on Healthcare and Services in Primary Care, Université Laval, Quebec, QC, Canada
[5]Faculty of Nursing Science, Université Laval, Quebec, QC, Canada
[6]Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada
[7]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada
[8]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada
[9]Research Chair in Emergency Medicine, Université Laval - CHAU Hôtel-Dieu de Lévis, Lévis, QC, Canada
[10]Research Chair on Chronic Diseases in Primary Care, Université de Sherbrooke, Chicoutimi, QC, Canada
[11]Department of Social and Preventive Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

Corresponding Author:
Hassane Alami, MScPH, MScHP, PhD
Institut national d'excellence en santé et services sociaux
2021, Avenue Union
Montréal, QC, H3A 2S9
Canada
Phone: 1 514 873 2563 ext 24404
Email: hassane.alami@umontreal.ca

Abstract

Artificial intelligence (AI) is seen as a strategic lever to improve access, quality, and efficiency of care and services and to build learning and value-based health systems. Many studies have examined the technical performance of AI within an experimental context. These studies provide limited insights into the issues that its use in a real-world context of care and services raises. To help decision makers address these issues in a systemic and holistic manner, this viewpoint paper relies on the health technology assessment core model to contrast the expectations of the health sector toward the use of AI with the risks that should be mitigated for its responsible deployment. The analysis adopts the perspective of payers (ie, health system organizations and agencies) because of their central role in regulating, financing, and reimbursing novel technologies. This paper suggests that AI-based systems should be seen as a health system transformation lever, rather than a discrete set of technological devices. Their use could bring significant changes and impacts at several levels: technological, clinical, human and cognitive (patient and clinician), professional and organizational, economic, legal, and ethical. The assessment of AI's value proposition should thus go beyond technical performance and cost logic by performing a holistic analysis of its value in a real-world context of care and services. To guide AI development, generate knowledge, and draw lessons that can be translated into action, the right political, regulatory, organizational, clinical, and technological conditions for innovation should be created as a first step.

(J Med Internet Res 2020;22(7):e17707) doi:10.2196/17707

KEYWORDS

artificial intelligence; health technology assessment; eHealth; health care; medical device; patient; health services

XSL•FO
RenderX

Introduction

Artificial intelligence (AI) raises many expectations in all sectors of society. There is no universally agreed upon definition of what AI encompasses. Generically, it refers to a branch of informatics that develops systems that—through their ability to *learn*—imitate the characteristics associated with human intelligence: reasoning, learning, adaptation, self-correction, sensory comprehension, and interaction [1,2].

AI is seen as a strategic lever to improve access, quality, and efficiency of health care and services [3]. For example, by exploiting exhaustive data sets from complex systems, it could contribute to improving clinical decision making (eg, diagnosis, screening, and treatment), service organization (eg, flow optimization, triage, and resource allocation), and patient management and follow-up (eg, drug administration and compliance) [4].

However, research on the application of AI in health focuses primarily on technological performance in experimental contexts or on ethical issues. Although relevant, these studies do not fully address the broader systemic policy questions surrounding their use in a real-world context of care and services. In a recent meta-analysis, Lieu et al [5] concluded that despite a diagnostic performance equivalent to that of health care professionals, the diagnostic applications of AI have not been externally validated in a real-world context of care and services. Poor reporting is also prevalent in studies on AI, which limits the reliable interpretation of results. Thus, before being integrated into clinical routine, AI applications should overcome what is called the *AI chasm*, that is, the gap between reported performance in laboratory conditions and its performance and impacts in a real-world context of care and services [6]. AI raises issues of different types, but they are, in practice, closely interconnected: economic, professional, organizational, clinical, human, cognitive, legal, ethical, and technological. To date, few scholars have examined these issues in a systemic and holistic manner [7].

In this viewpoint paper, relying on the *health technology assessment (HTA) core model* [8], which is a methodological framework used to facilitate production and sharing of HTA information [9], we examine, based on our own experience as HTA academics and practitioners and in light of the emerging literature on the subject, issues raised by the use of AI. More specifically, we contrast the expectations specific to the health sector and the risks that should be mitigated for AI to be deployed responsibly. We limit our analysis to AI-based applications for clinical use (eg, diagnostic), some of which would be classified by the US Food and Drug Administration (FDA) as *software as a medical device*: "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device" [10]. They are subject to formal regulatory approval [6,11].

In this paper, we offer critical observations and reflections that are informed by our various roles in HTA as health technology governance experts, researchers-evaluators, and/or decision makers. The analysis primarily adopts the perspective of payers (ie, health system organizations and agencies) because of their central role in regulating, funding, and reimbursing technologies [12].

On the basis of the *HTA core model*, we summarize key challenges posed by AI in a real-world context of care and services, which include (1) technological, (2) clinical, (3) human and cognitive (patient and clinician), (4) professional and organizational, (5) economic, and (6) legal and ethical dimensions (Textbox 1). We provide examples for each of these dimensions and underline how decision makers could approach them in a more systemic and holistic manner.

XSL•FO
RenderX

Textbox 1. Synthesis of some key challenges posed by artificial intelligence.

Technological

- Laboratory performance versus a real-world context of care and services
- Data quality and representativeness of the general population or other contexts
- Black box: how and why the decision is made?
- Is artificial intelligence (AI) reliable and free of biases or technical failures?
- How AI would react in situations where input data deviate from initial data?
- Cybersecurity: data and model (algorithm)
- Interoperability: fragmented systems and unstructured data

Clinical

- Reproduction of tropism of practice models
- Actual clinical added value in a real-world context of care and services: difficult to distinguish the effect of the AI's decision from the rest of the preventive and/or therapeutic strategy
- The level of accuracy of AI in diagnosis and recommendations (reference standard) in a real-world context of care and services

Human and cognitive (patients)

- Evolution of the nature and quality of the clinician-patient relationship
- Loss of human contact: isolation of some people
- Unrealistic expectations in some patients regarding clinical outcomes
- Black box: could be perceived as a restriction on the patient's right to make a free and informed decision
- AI could be beneficial for one part of the population and not be for others: what is the good target population?

Human and cognitive (clinicians)

- How to integrate AI into the electronic health record (EHR) and clinical routine with minimal effort and disruption for clinicians?
- Nonintuitive technologies: weigh-down workflows and burden for clinicians and cognitive overload
- Disruption of interpersonal communication styles (eg, clinician-clinician and clinician-patient)
- AI as clinical mind: challenge of clinician's decision-making autonomy
- Absolute confidence in AI: technical dependence

Professional and organizational

- How will it fit into the patient care and services trajectory?
- How will it be integrated into the clinical-administrative processes and workflows of organizations and health system?
- What changes will result in terms of service organization (eg, waiting time, primary care and specialized services relationships)?
- How will it impact on professional jurisdictions (eg, reserved activities, responsibility, training, new skills, and expertise)?

Economic

- Investments required: continuous performance tests, software and data quality tests, infrastructure and equipment upgrades, human expertise, and training
- Clinical tropism and reimbursement/billing biases: costs for patients, clinicians, organizations, and health system
- Need of new financing mechanisms, appropriate remuneration and/or reimbursement models, and insurance models

Legal and ethical

- When is AI considered as a decision-making support tool? When is it considered as a decision-making tool?
- What are the limits of technology and their potential legal implications?
- If the AI makes a mistake (eg, black box), who will be held responsible? If the patient is harmed, who will pay for the repairs?
- What would be the consequence if the clinician does not comply with the recommendations of an AI and this leads to an error?
- AI needs access to data from different sources: consent is becoming more complex, as patients will be asked to authorize the use of diversified amounts of data

XSL•FO
RenderX

- Protection and confidentiality: origin of the data, how consent was obtained, and authorization to use and/or reuse the data
- Who owns the data? Who is responsible for it? Who can use (or reuse) it and under what conditions?

Technological Dimension

Generalizability and Reproducibility

Studies that focus on technological issues indicate that AI should provide the same level of performance in a real-world context of care and services as that obtained in laboratory conditions. However, this requirement is difficult to achieve [13-16]. The majority of AI applications reported in the literature are not exploitable in clinical practice [17]. AI is often trained with so-called *clean* (exclusion of poor-quality images) and complete data sets (elimination of imperfect data) [18]. It may not be operational in other contexts where data are incomplete or of poor quality (electronic health record [EHR] with missing data and/or erroneously entered data) [19-21]. This applies to some categories of the patient population (eg, low economic status and psychosocial problems) who receive care and services in a fragmented way in several organizations (institutional wandering) [21-24]. In addition, AI is usually trained on data specific to certain sites (hospital) and patients who are not necessarily representative of the general population. This includes decontextualized data (lack of psychosocial and organizational indicators) and data about disproportionately sick individuals (data enriched by metastases cases), men, and those from a European origin (ethnodiversity) [23,25-28].

Health organizations and systems produce and manage data in different ways. Variations may exist in clinical protocols (eg, diagnosis, procedures, and vital parameters) and devices (eg, different types of scanners, EHRs, and laboratory devices) on which AI applications are trained and those on which they are expected to operate [29,30]. These variations could affect the AI performance in a real-world context of care and services [31]. For example, an AI application trained on data from 2 hospitals in the United States performed poorly in a third hospital [13,32]. In its decision, the AI application had as predictors the image characteristics (magnetic resonance imaging machines specifications), imperceptible to humans, specific to the technological systems of the hospitals where it was trained. The AI solution had adapted to *noise* rather than to the signal of clinical interest [33]. When used in the third hospital, it was *deprived* of these expected predictors (*noise*), which affected its anticipated performance [32]. In the same vein, the use of data from the *Framingham Heart Study* to predict the risk of cardiovascular events produced biased results, which both overestimated and underestimated risk when AI was used in non-white populations [34,35]. The ability of AI to operate without bias or confounding factors on different devices and protocols remains a major challenge [36,37]. Thus, the fact that an algorithm was trained on large data sets does not mean that its results are generalizable.

Interpretability and Transparency

The interpretability and transparency of AI are important issues. The *black box* logic makes some AI applications vulnerable and at risk to false discoveries via spurious associations: how is the decision made and on what basis (justification and process description) [24,38,39]. This issue is central because these technologies will be diffused on a large scale. The error of a defective AI could have a greater impact (several patients) than a clinician's error on a single patient [20,31,40].

Interpretability and transparency are also necessary to identify the origin of errors, biases, or failures that should be prevented and/or avoided in the future [3,21,41]. For example, an AI application could lead to many undesirable impacts related to: (1) poor-quality training data, which could lead to erroneous or biased knowledge (*garbage in, garbage out*), whereas technology may further amplify how poor data produce poor results (noisy data and missing values); (2) the presence of a technical flaw in the algorithm (code), which could lead to erroneous inferences, even if good-quality data are used; (3) decision-making criteria that may not be universally acceptable; and (4) the emergence of new situations for which AI could not adapt, even with good-quality data and code [21,30,42-45]. For example, the emergence of new treatments or practices may require changes in clinical protocols; however, at present, AI applications are not developed to manage temporal data naturally in a real-world context of care and services. However, diseases and treatments evolve in a nonlinear manner [18,45]. The question thus remains regarding how AI would react, with observable indicators, in situations where input data deviate from initial data (EHRs and real-time monitoring devices), in the medium and long term [45,46].

The risk of cyberattacks is also a major concern. The data could be modified and/or fed by other false or biased data in a way that is difficult to detect [1]. For example, a slight intentional modification of laboratory results in a patient's EHR resulted in significant changes in the estimates of a *well-trained* AI of the same patient's risk of mortality [24]. For AI, the issue is two-fold because it is necessary to ensure the security of the data and that of the model (the algorithm). Interoperability is also a significant issue. The integration of AI in fragmented and noninteroperable information technology systems and organizations could create more problems than it will solve; to deliver its full potential, AI needs integrated and interoperable systems with fluent and optimal data circulation and exchange [17].

Finally, addressing interpretably and transparency in AI could be compromised by intellectual property issues, competitive strategy, and financial advantages that make companies reluctant to disclose their source codes [3].

Clinical Dimension

AI can entrench and disseminate practice models specific to particular contexts (organizations or health systems) and not necessarily accepted or used in others (tropism) [38]. For example, clinicians in some countries stopped using *IBM Watson for Oncology* because it reflected US specificity in cancer treatment [1,47].

XSL•FO
RenderX

To use AI in their decision making, clinicians should understand how it makes decisions in the first place [38,45,48]. They need the evidence to support a given conclusion to be able to carry out the necessary verifications or even corrections [14]: Why this decision (what information or image—or part of the image—tipped the final decision of the AI)? Why not another option (or choice)? When may I consider that the decision is correct? When should I accept this decision? How can I correct the error when it occurs?

AI should provide clinically added value for the patient. In a real-world context of care and services, much information, decisions, and diagnoses could intersect (eg, symptom assessment, laboratory tests, and radiology). At present, it is difficult to distinguish the effect of an AI-based decision from the overall preventive and/or therapeutic strategy of patient care [49,50].

Another clinical issue is determining the level of accuracy of AI for diagnosis and recommendations. In practice, decisions physicians make could diverge or even contradict each other in many situations. The *gold standard* is not always easy to define in a process that involves complex judgments [38,51,52]. In this case, should the standard reflect that of the lead clinician (or clinicians) in the organization? Or the one accepted by the majority of clinicians? Or the one reported in similar contexts? Some authors believe that for technologies that aim to provide pragmatic solutions under suboptimal conditions, AI performance should correspond to clinically acceptable practice in a given context and not necessarily to recommended practices [32]. This last point is likely to be problematic, particularly in a context where health systems are trying to overcome the challenge of practice variations to be able to provide equitable and quality services for all citizens.

Human and Cognitive Dimensions

AI could affect the nature and quality of the clinician-patient relationship and their expectations for care and follow-up [53,54]. The loss of human contact could lead to increased isolation of some people (replacement of health care providers) [1]. Some patients may feel able to control and manage their disease, with passive surveillance and/or less contact with the clinician, whereas others may feel overwhelmed by additional responsibilities [55]. AI may also create unrealistic expectations in some patients regarding clinical outcomes, which could have a negative impact on their care and service experience [56]. In addition, some AI-based decisions could be perceived as a restriction on the patient's right to make a free and informed decision [1,53]. Cultural and social aspects could play an important role in how patients will respond to AI and therefore how effective it can prove in practice [57]. Hence, it is important to know on which basis one may define the target population that can benefit from it [58]. In this regard, the question of social acceptability (acceptable risk and public confidence) also needs to be considered, which goes beyond the simple question of the effectiveness and usability of AI [59].

For clinicians, the challenge is to integrate AI into the EHR and clinical routine with minimal effort while respecting their decision-making autonomy [24]. Nonintuitive technologies could encumber workflows and become a burden for clinicians without improving service delivery [30,60]. Otherwise, the ability of AI to combine data from the scientific literature with learning from practice data could generate a repository of clinical practices (*clinical mind*), which could give AI an *unwanted* power or authority [35]. In some situations, AI may reduce the clinician's ability to take into account patient values and preferences. In contrast, some clinicians may develop absolute confidence and become dependent on AI, thus relinquishing their responsibility to verify or double-check its decisions [1].

In short, if clinicians feel overloaded and workflows become more complex, AI may be rejected because of self-perceived inefficacy and performance, alert fatigue, cognitive overload, and disruption of interpersonal communication routines [54,61-63].

Professional and Organizational Dimensions

Global appreciation of the added value of AI should take into account the nature and magnitude of the professional and organizational changes required for its use [6]. For example, the FDA has approved an AI application used for diabetic retinopathy screening, which may be used in primary care clinics [11]. As in some countries, the screening procedure is performed by an ophthalmologist (specialist), some questions arise: How will this technology fit into patient care and services trajectory? How will it be integrated into the clinical-administrative processes of organizations and the health system? If used at the primary care level, will general practitioners, nurses, or optometrists be allowed to supervise the AI? If so, under what conditions? What will be the impact on professional jurisdictions (regulated activities, remuneration, and training)? What changes will result in terms of service organization and clinical-administrative workflows (waiting time at primary care level, primary care, and specialized services relationships)?

Thus, AI could lead to a redistribution of work between different professional scopes of practice and highlight the need for other clinical, administrative, and technical skills and expertise. This will require clarifying new rules and processes (clinical and administrative), negotiating and reframing professional jurisdictions, responsibilities, and privileges associated with them and reassessing the number of positions needed and the new skills required to work (with) and/or perform other tasks that accompany its use. This will have to take into consideration how new roles in terms of skills in informatics and data science and the ability to liaise may be introduced within clinical teams [64].

Finally, today, most AI applications are developed to perform a single task or a set of very specific tasks (eg, diagnosing only diabetic retinopathy and macular edema) [65]. They are unusable for other diagnoses for which they are not trained (eg, nondiabetic retinopathy lesions and eye melanoma) and are unable, at least for the moment, to replace a complete clinical examination [66]. Payers will thus be tasked to determine whether AI provides sufficient added value in relation to the nature and magnitude of the clinical, cognitive, professional, and organizational changes it could generate.

XSL•FO
RenderX

Economic Dimension

To adapt an AI to a local environment, considerable investments and expenditures may be necessary. The evolution of AI in a real-world context of care and services, by integrating large amounts of data of various types and sources, requires additional resources to ensure its proper functioning and stability: continuous performance tests, software and data quality tests, infrastructure and equipment upgrades, human expertise, and training [3,67]. However, many health organizations do not have a secure and scalable technological and data infrastructure as well as adequate human resources to ensure proper collection of the data necessary for the training and adaptation of AI to their local population and clinical environment [17]. The literature on AI's promises as well as innovation policies that support its development downplays the capital-intensive requirements that are required to properly deploy AI, compared with the day-to-day work of managers in organizations.

In health systems where activity-based financing is the basis for funding health organizations, some clinicians tend to enter the *highest paying* codes for each clinical activity (ie, the most complex case of an intervention) to increase performance and maximize revenue. An AI application trained on data from these organizations (EHR with invoicing or reimbursement data) could amplify biases inherent in such practices that do not necessarily reflect the actual clinical condition [23,44,68]. The replication and entrenchment at a large scale of these biases could result in significant costs for patients, clinicians, organizations, and the health systems [35].

Similarly, some AI applications may be *too cautious*, resulting in an increase in requests for unnecessary testing and treatment, leading to overdiagnosis or overprescription [69]. Their recommendations, which are not necessarily associated with improved patient outcomes, could lead to increased costs and expenses for patients and the health system.

Legal and Ethical Dimensions

Many AI technologies are still considered today as *decision-making support tools* for clinicians. It could then be argued that the legal responsibility for the decision still rests with the clinician. However, with the growing performance of AI, clinicians may be increasingly influenced and may more easily accept AI decisions, even when there is clinical ambiguity. Determining the clinician's degree of responsibility becomes more complex [30]. The challenge here is to distinguish between several situations: When is it considered a decision-making support tool? When is it considered a decision-making tool? This distinction is key in defining who is legally responsible in the event of an error or a malfunction (professional misconduct) [30,51,70].

For example, if the clinical decision is based on an erroneous clinical recommendation from the AI (delayed or erroneous treatment), who will be held responsible? Is it the technology developer, technology provider, clinician, organization, or do they all share responsibility (and how)? In some jurisdictions, to confirm professional misconduct, it is necessary to prove that the *standard of care* was not followed. This standard is blurred when AI comes into play [2]. In addition, the likely consequence if the clinician does not comply with the recommendations of an AI and if this leads to an error must be anticipated [2]. It could be argued that the responsibility should rest with the human controller of AI, but such a responsibility becomes difficult to clarify when autonomous technologies are used [57]. In this regard, standards may shift over time: "What happens if medical practice reaches a point where AI becomes part of the standard of care?" Medical insurers and regulators will have to be able to distinguish errors inherent in the tool from those resulting from misuse by the clinician, the organization, or even the patient, an issue exacerbated by the *black box* of AI [51,71].

To generate a complete picture of the patient, AI will need access to data from different organizations (hospitals and insurers) [45]. The risk of disclosing sensitive information about patients or certain populations is real [45]. For example, some AI applications can reidentify an individual from only three different data sources [25,38,72]. In the same vein, the issue of consent is becoming more complex, as patients will be asked to authorize the use of increasingly large and diversified amounts of data about them: medical records, audio, videos, and socioeconomic data [58]. Problems could arise if the patient only consents to sharing parts of his or her data. Usually, confidentiality means that the clinician can withhold certain information—at the patient's request (or not)—and avoid entering it into the EHR. Incomplete data make AI less efficient and does not allow patients to benefit from the best possible services. AI may not be fully operational in a real-world context of care and services if specific restrictions on data access and use are applied [38].

Protection and confidentiality requirements imply the obligation to know several things: the origin of the data, how consent was obtained, and authorization to use and/or reuse the data for training and in a real-world context of care and services. As the data may come from different sources and contexts, different conditions and precautions will need to be considered [73]. Regulators will need to determine who owns the data and, in the context of public-private partnerships, who is responsible for its collection, use, transmission to third parties, and under what conditions [17]. As the answers will vary according to the nature of the data, the jurisdictions, and the purpose of use, the task at hand is sizable [73]. Finally, payers will have to recognize that the ethical implications of AI affect, directly or indirectly, all the other dimensions discussed earlier.

Conclusions

The purpose of this viewpoint paper is to provide a structured roadmap of the issues surrounding the integration of AI into health care organizations and systems. To the best of our knowledge, this is one of the few papers that offers a multidimensional and holistic analysis on the subject [7]. It contributes to current knowledge by providing a necessary basis for reflections, exchanges, and knowledge sharing among the various stakeholders concerned with AI in health care.

In light of the issues we identified, it becomes clear that regulatory and decision-making organizations as well as HTA agencies are facing unprecedented complexity: evaluating and approving so-called disruptive technologies, especially AI, requires taking several issues into consideration altogether.

XSL•FO
RenderX

Many studies have reported significant technical performance of AI technologies, but very few have adopted a holistic standpoint that can situate their impacts and associated changes and transformations in health systems. Technical studies are rarely adapted to the complexity surrounding AI applications, as they overlook the context-dependent changes or adjustments the implementation and use of technology requires (variations, clinical and organizational interactions, and interdependencies) [74]. According to *the frame problem* [62,75], which highlights the difficulty for AI, beyond the specific tasks it masters, to update its set of axioms to capture the context in which it is implemented and used (eg, patient preferences, environment and social support, clinical history, personality/cultural characteristics and values that influence clinical outcomes, and empathy in medicine), the complexity inherent in the use of AI applications in the real-world context of care and services may seem difficult to overcome [62].

For informed decision making, there is a real need for evaluations that address AI as a lever of health system transformation. Given the magnitude of the implications it could have at all levels, the evaluation of AI's value proposition should go beyond its technical performance and cost logic to incorporate its global value based on a holistic analysis in a real-world context of care and services. In this vein, technology brings value when its use in a real-world context of care and services contributes to the aims of the health system and aligns with the values of society. Global value appreciation could be based on the *quintuple aim*: (1) better quality and experience of care and services for patients; (2) a better state of health and well-being for the entire population; (3) reducing costs for responsible and sustainable resource management; (4) a better quality of work and satisfaction of health care providers; and (5) equity and inclusion to avoid exacerbating health disparities in the population [76]. From this perspective, further research on the evaluation of AI should no longer be limited to a technological approach, which only demonstrates quality from an engineering point of view and costs—motivated mainly by a logic of short-term savings—but should broaden its horizons to include the dimensions this paper underscored [77,78].

Real-world evaluations could be a major asset in informing AI decision making. In the context of uncertainty, iterative and reflective evaluation approaches should be developed to encourage dialog and collaboration among all relevant stakeholders (eg, payers, health care providers, technology providers, regulators, citizens/patients, academic researchers, and evaluation agencies) [63,78,79]. In addition, an early dialog between these stakeholders is needed to identify the evidence required to inform decision making [63,78]. This approach would also help AI providers to better understand the expectations of the health system [78]. This change implies that HTA should play an active role as a mediator and facilitator of transparent dialog between different stakeholders who are implicated throughout the technology's life cycle [78,80].

Decision making for innovative technologies is inherently complex, in particular because of visions, perceptions, and objectives that may differ between the stakeholders involved: *risk sharing* is essential to strive to find a balance between uncertainty and added value [81]. In this regard, "major radical innovations never bring new technologies into the world in a fully developed form" but "appear in a crude and embryonic state with only a few specific uses" [81]. It is their use in a real-world context of care and services, through a process of *learning by doing* (improving users' skills) and *learning by using* (improving users' knowledge), which makes it possible to appreciate their global value [81]. With the complexity associated with AI, value appreciation becomes even more complex, challenging the traditional methodological foundations that are the basis for decision making about innovative technologies [82]. This also presents a unique opportunity for HTA to evolve and adapt (evaluative framework and contextualized data), particularly in view of the importance of contexts in the appreciation of the value of innovative technologies [83,84]. It is necessary for HTA scholars and practitioners to explore and exploit other avenues, complementary to traditional methods, to collect data and information that can better inform AI-related decisions [85].

Finally, this new context implies mechanisms for continuous collective learning and sharing of lessons. To do so, there is a need for learning and flexible health organizations and systems that are able to adjust and operate under uncertainty. In this regard, creating the political, regulatory, organizational, clinical, and technological conditions necessary for proper innovation is the first step. This requires building trust to ensure stakeholder engagement to guide AI developments, rapidly generate knowledge in a real-world context of care and services, and draw lessons to translate them into action.

Acknowledgments

HA was supported by the Canadian Institutes of Health Research (CIHR)'s Health System Impact Fellowship. This program is led by CIHR's Institute of Health Services and Policy Research, in partnership with the Fonds de recherche du Québec–Santé and the Institut national d'excellence en santé et services sociaux. The authors would like to thank the reviewers and the editorial team for their insightful comments and suggestions, as these comments and suggestions led to an improvement of the manuscript.

Authors' Contributions

HA and PL produced the first draft of this manuscript and received input from YA, MG, MG, JS, DR, RF, MA, and JF. All authors have read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Scott RE, Mars M. Principles and framework for ehealth strategy development. J Med Internet Res 2013 Jul 30;15(7):e155 [FREE Full text] [doi: 10.2196/jmir.2250] [Medline: 23900066]
2. Massalha S, Clarkin O, Thornhill R, Wells G, Chow BJ. Decision support tools, systems, and artificial intelligence in cardiac imaging. Can J Cardiol 2018 Jul;34(7):827-838. [doi: 10.1016/j.cjca.2018.04.032] [Medline: 29960612]
3. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
4. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplasty 2018 Aug;33(8):2358-2361. [doi: 10.1016/j.arth.2018.02.067] [Medline: 29656964]
5. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297 [FREE Full text] [doi: 10.1016/s2589-7500(19)30123-2]
6. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. Digit Med 2018;1:40.
7. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
8. Lampe K, Mäkelä M, Garrido MV, Anttila H, Autti-Rämö I, Hicks NJ, European Network for Health Technology Assessment (EUnetHTA). The HTA core model: a novel method for producing and reporting health technology assessments. Int J Technol Assess Health Care 2009 Dec;25(Suppl 2):9-20. [doi: 10.1017/S0266462309990638] [Medline: 20030886]
9. Kristensen FB, Lampe K, Wild C, Cerbo M, Goettsch W, Becla L. The HTA core model-10 years of developing an international framework to share multidimensional value assessment. Value Health 2017 Feb;20(2):244-250 [FREE Full text] [doi: 10.1016/j.jval.2016.12.010] [Medline: 28237203]
10. Software as a Medical Device (SaMD): Key Definitions. International Medical Device Regulators Forum (IMDRF). 2013. URL: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf [accessed 2020-04-23]
11. FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems. US Food and Drug Administration. 2018. URL: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [accessed 2020-04-23]
12. Godman B, Bucsics A, Bonanno PV, Oortwijn W, Rothe CC, Ferrario A, et al. Barriers for access to new medicines: searching for the balance between rising costs and limited budgets. Front Public Health 2018;6:328 [FREE Full text] [doi: 10.3389/fpubh.2018.00328] [Medline: 30568938]
13. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 2018 Nov;15(11):e1002683 [FREE Full text] [doi: 10.1371/journal.pmed.1002683] [Medline: 30399157]
14. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology: current applications and future directions. PLoS Med 2018 Nov;15(11):e1002707 [FREE Full text] [doi: 10.1371/journal.pmed.1002707] [Medline: 30500815]
15. Wong J, Horwitz MM, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. Curr Epidemiol Rep 2018 Dec;5(4):331-342 [FREE Full text] [doi: 10.1007/s40471-018-0165-9] [Medline: 30555773]
16. Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, ACTERREA Research Group. Big data and targeted machine learning in action to assist medical decision in the ICU. Anaesth Crit Care Pain Med 2019 Aug;38(4):377-384. [doi: 10.1016/j.accpm.2018.09.008] [Medline: 30339893]
17. Panch T, Mattie H, Celi LA. The 'inconvenient truth' about AI in healthcare. Digit Med 2019;2:77.
18. Sherbet GV, Woo WL, Dlay S. Application of artificial intelligence-based technology in cancer management: a commentary on the deployment of artificial neural networks. Anticancer Res 2018 Dec;38(12):6607-6613. [doi: 10.21873/anticanres.13027] [Medline: 30504368]
19. Redd TK, Campbell JP, Brown JM, Kim SJ, Ostmo S, Chan RV, Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. Br J Ophthalmol 2018 Nov 23:- epub ahead of print. [doi: 10.1136/bjophthalmol-2018-313156] [Medline: 30470715]
20. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018 Sep 7;1(5):e182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]
21. Walsh CG, Chaudhry B, Dua P, Goodman KW, Kaplan B, Kavuluru R, et al. Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. JAMIA Open 2020;3(1):9-15 [FREE Full text] [doi: 10.1093/jamiaopen/ooz054]

XSL•FO
RenderX

22. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 2018 Nov 1;178(11):1544-1547 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3763] [Medline: 30128552]
23. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
24. Wang F, Casalino LP, Khullar D. Deep learning in medicine-promise, progress, and challenges. JAMA Intern Med 2019 Mar 1;179(3):293-294. [doi: 10.1001/jamainternmed.2018.7117] [Medline: 30556825]
25. Crigger E, Khoury C. Making policy on augmented intelligence in health care. AMA J Ethics 2019 Feb 1;21(2):E188-E191 [FREE Full text] [doi: 10.1001/amajethics.2019.188] [Medline: 30794129]
26. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. J Am Med Assoc 2017 Aug 8;318(6):517-518. [doi: 10.1001/jama.2017.7797] [Medline: 28727867]
27. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature 2018 Jul;559(7714):324-326. [doi: 10.1038/d41586-018-05707-8] [Medline: 30018439]
28. Bejnordi BE, Veta M, van Diest PJ, van Ginneken B, Karssemeijer N, Litjens G, The CAMELYON16 Consortium, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. J Am Med Assoc 2017 Dec 12;318(22):2199-2210 [FREE Full text] [doi: 10.1001/jama.2017.14585] [Medline: 29234806]
29. Keel S, van Wijngaarden P. The eye in AI: artificial intelligence in ophthalmology. Clin Exp Ophthalmol 2019 Jan;47(1):5-6. [doi: 10.1111/ceo.13435] [Medline: 30485615]
30. Harwich E, Laycock K. Thinking on its Own: AI in the NHS. London, UK: Reform Research Trust; 2018.
31. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
32. Nevin L, PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. PLoS Med 2018 Nov;15(11):e1002708 [FREE Full text] [doi: 10.1371/journal.pmed.1002708] [Medline: 30500811]
33. Tiffin PA, Paton LW. Rise of the machines? Machine learning approaches and mental health: opportunities and challenges. Br J Psychiatry 2018 Sep;213(3):509-510. [doi: 10.1192/bjp.2018.105] [Medline: 30113285]
34. Gijsberts CM, Groenewegen KA, Hoefer IE, Eijkemans MJ, Asselbergs FW, Anderson TJ, et al. Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. PLoS One 2015;10(7):e0132321 [FREE Full text] [doi: 10.1371/journal.pone.0132321] [Medline: 26134404]
35. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med 2018 Mar 15;378(11):981-983 [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]
36. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. Med Image Anal 2016 Oct;33:170-175 [FREE Full text] [doi: 10.1016/j.media.2016.06.037] [Medline: 27423409]
37. Mateos-Pérez JM, Dadar M, Lacalle-Aurioles M, Iturria-Medina Y, Zeighami Y, Evans AC. Structural neuroimaging as clinical predictor: a review of machine learning applications. Neuroimage Clin 2018;20:506-522 [FREE Full text] [doi: 10.1016/j.nicl.2018.08.019] [Medline: 30167371]
38. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018 Apr;15(141):- [FREE Full text] [doi: 10.1098/rsif.2017.0387] [Medline: 29618526]
39. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online 2014 Jul 5;13:94 [FREE Full text] [doi: 10.1186/1475-925X-13-94] [Medline: 24998888]
40. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 1;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
41. Kuan R. Adopting AI in Health Care Will Be Slow and Difficult. Harvard Business Review. 2019. URL: https://hbr.org/2019/10/adopting-ai-in-health-care-will-be-slow-and-difficult [accessed 2020-04-23]
42. Fox J, Das SK. Safe and Sound: Artificial Intelligence in Hazardous Applications. Menlo Park, CA: AAAI Press; 2000.
43. Bibault J, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. Cancer Lett 2016 Nov 1;382(1):110-117. [doi: 10.1016/j.canlet.2016.05.033] [Medline: 27241666]
44. Yu K, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
45. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
46. Derrington D. Artificial Intelligence for Health and Health Care. Office of the National Coordinator for Health Information. 2017. URL: https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf [accessed 2019-04-23]
47. Ross C, Swetlitz I. IBM Pitched Its Watson Supercomputer as a Revolution in Cancer Care. It's Nowhere Close. Stat. 2017. URL: https://www.statnews.com/2017/09/05/watson-ibm-cancer/ [accessed 2020-04-23]

XSL•FO
RenderX

48. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. J Am Med Assoc 2016 Dec 13;316(22):2366-2367. [doi: 10.1001/jama.2016.17563] [Medline: 27898977]
49. Belfort MA, Clark SL. Computerised cardiotocography-study design hampers findings. Lancet 2017 Apr 29;389(10080):1674-1676 [FREE Full text] [doi: 10.1016/S0140-6736(17)30762-6] [Medline: 28341516]
50. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018 Mar;286(3):800-809. [doi: 10.1148/radiol.2017171920] [Medline: 29309734]
51. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. J Am Med Assoc 2019 Oct 4:-epub ahead of print. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]
52. Kim J, Hong J, Park H. Prospects of deep learning for medical imaging. Precis Future Med 2018 Jun;2(2):37-52 [FREE Full text] [doi: 10.23838/pfm.2018.00030]
53. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019 Mar;69(2):127-157 [FREE Full text] [doi: 10.3322/caac.21552] [Medline: 30720861]
54. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019 Mar 20;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
55. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
56. Luxton DD. Should Watson be consulted for a second opinion? AMA J Ethics 2019 Feb 1;21(2):E131-E137 [FREE Full text] [doi: 10.1001/amajethics.2019.131] [Medline: 30794122]
57. Luxton DD. Artificial intelligence in psychological practice: current and future applications and implications. Prof Psychol Res Pr 2014;45(5):332-339. [doi: 10.1037/a0034559]
58. Lovejoy CA, Buch V, Maruthappu M. Technology and mental health: the role of artificial intelligence. Eur Psychiatry 2019 Jan;55:1-3. [doi: 10.1016/j.eurpsy.2018.08.004] [Medline: 30384105]
59. Poder TG, Bellemare CA, Bédard SK, Lemieux R. Social acceptance and population confidence in telehealth in Quebec. BMC Health Serv Res 2015 Feb 21;15:72 [FREE Full text] [doi: 10.1186/s12913-015-0727-1] [Medline: 25889230]
60. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. J Am Med Assoc 2018 Dec 4;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
61. McCullagh LJ, Sofianou A, Kannry J, Mann DM, McGinn TG. User centered clinical decision support tools: adoption across clinician training level. Appl Clin Inform 2014;5(4):1015-1025 [FREE Full text] [doi: 10.4338/ACI-2014-05-RA-0048] [Medline: 25589914]
62. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
63. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 9;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
64. Gillan C, Milne E, Harnett N, Purdie T, Jaffray D, Hodges B. Professional implications of introducing artificial intelligence in healthcare: an evaluation using radiation medicine as a testing ground. J Radiother Pract 2018 Oct 3;18(1):5-9 [FREE Full text] [doi: 10.1017/s1460396918000468]
65. Syed AB, Zoga AC. Artificial intelligence in radiology: current technology and future directions. Semin Musculoskelet Radiol 2018 Nov;22(5):540-545. [doi: 10.1055/s-0038-1673383] [Medline: 30399618]
66. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J Am Med Assoc 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
67. Sendak M, Gao M, Nichols M, Lin A, Balu S. Machine learning in health care: a critical appraisal of challenges and opportunities. EGEMS (Wash DC) 2019 Jan 24;7(1):1 [FREE Full text] [doi: 10.5334/egems.287] [Medline: 30705919]
68. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
69. Loder J, Nicholas L. Confronting Dr Robot: Creating a People-Powered Future For AI in Health. Nesta: The Innovation Foundation. 2018. URL: https://media.nesta.org.uk/documents/confronting_dr_robot.pdf [accessed 2020-04-23]
70. Toratani M, Konno M, Asai A, Koseki J, Kawamoto K, Tamari K, et al. A convolutional neural network uses microscopic images to differentiate between mouse and human cell lines and their radioresistant clones. Cancer Res 2018 Dec 1;78(23):6703-6707 [FREE Full text] [doi: 10.1158/0008-5472.CAN-18-0653] [Medline: 30254144]
71. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? AMA J Ethics 2019 Feb 1;21(2):E160-E166 [FREE Full text] [doi: 10.1001/amajethics.2019.160] [Medline: 30794126]
72. Shokri R, Stronati M, Song C, Shmatikov V. Membership Inference Attacks Against Machine Learning Models. In: Proceedings of the IEEE Symposium on Security and Privacy. 2017 Presented at: SP'17; May 22-26, 2017; San Jose, CA, USA. [doi: 10.1109/sp.2017.41]

XSL•FO
RenderX

73. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018 Nov;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
74. Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs--same or different? Value Health 2009 Jun;12(4):402-404 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00476_1.x] [Medline: 19138306]
75. McCarthy J, Hayes PJ. Some philosophical problems from the standpoint of artificial intelligence. Read Artif Intel 1981:431-450 [FREE Full text] [doi: 10.1016/b978-0-934613-03-3.50033-7]
76. Matheny M, Israni S, Ahmed M, Whicher D. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. E-Health Research. 2019. URL: https://ehealthresearch.no/files/documents/Rapporter/Andre/2019-12-AI-in-Health-Care.pdf [accessed 2020-04-23]
77. Fuchs S, Olberg B, Panteli D, Perleth M, Busse R. HTA of medical devices: challenges and ideas for the future from a European perspective. Health Policy 2017 Mar;121(3):215-229. [doi: 10.1016/j.healthpol.2016.08.010] [Medline: 27751533]
78. Facey K, Henshall C, Sampietro-Colom L, Thomas S. Improving the effectiveness and efficiency of evidence production for health technology assessment. Int J Technol Assess Health Care 2015 Jan;31(4):201-206 [FREE Full text] [doi: 10.1017/S0266462315000355] [Medline: 26646858]
79. Pomey M, Brouillard P, Ganache I, Lambert L, Boothroyd L, Collette C, et al. Co-construction of health technology assessment recommendations with patients: an example with cardiac defibrillator replacement. Health Expect 2020 Feb;23(1):182-192 [FREE Full text] [doi: 10.1111/hex.12989] [Medline: 31691439]
80. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. Med Decis Making 2012;32(2):376-388. [doi: 10.1177/0272989X11416870] [Medline: 21987539]
81. Angelis A, Kanavos P. Multiple criteria decision analysis (MCDA) for evaluating new medicines in health technology assessment and beyond: the advance value framework. Soc Sci Med 2017 Sep;188:137-156 [FREE Full text] [doi: 10.1016/j.socscimed.2017.06.024] [Medline: 28772164]
82. Wahlster P, Goetghebeur M, Kriza C, Niederländer C, Kolominsky-Rabas P, National Leading-Edge Cluster Medical Technologies 'Medical Valley EMN'. Balancing costs and benefits at different stages of medical innovation: a systematic review of multi-criteria decision analysis (MCDA). BMC Health Serv Res 2015 Jul 9;15:262 [FREE Full text] [doi: 10.1186/s12913-015-0930-0] [Medline: 26152122]
83. van der Wilt GJ, Gerhardus A, Oortwijn W. Toward integration in the context of health technology assessment: the need for evaluative frameworks. Int J Technol Assess Health Care 2017 Jan;33(5):586-590. [doi: 10.1017/S0266462317004354] [Medline: 29256849]
84. Poder TG, Bellemare CA. Importance of contextual data in producing health technology assessment recommendations: a case study. Int J Technol Assess Health Care 2018 Jan;34(1):63-67. [doi: 10.1017/S0266462317004469] [Medline: 29482668]
85. Poder TG. Using the health technology assessment toolbox to facilitate procurement: the case of smart pumps in a Canadian hospital. Int J Technol Assess Health Care 2017 Jan;33(1):54-62. [doi: 10.1017/S0266462317000125] [Medline: 28578750]

Abbreviations

AI: artificial intelligence
CIHR: Canadian Institutes of Health Research
EHR: electronic health record
FDA: Food and Drug Administration
HTA: health technology assessment

Edited by G Eysenbach; submitted 11.01.20; peer-reviewed by HS Gan, T Poder, M Reynolds, B Chaudhry, A Zaidi, C Tziraki, SGS Shah; comments to author 22.03.20; revised version received 25.04.20; accepted 13.05.20; published 07.07.20.

Please cite as:
Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, Roy D, Fleet R, Ag Ahmed MA, Fortin JP
Artificial Intelligence and Health Technology Assessment: Anticipating a New Level of Complexity
J Med Internet Res 2020;22(7):e17707
URL: https://www.jmir.org/2020/7/e17707
doi:10.2196/17707
PMID:32406850

©Hassane Alami, Pascale Lehoux, Yannick Auclair, Michèle de Guise, Marie-Pierre Gagnon, James Shaw, Denis Roy, Richard Fleet, Mohamed Ali Ag Ahmed, Jean-Paul Fortin. Originally published in the Journal of Medical Internet Research

XSL•FO
RenderX

(http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Blockchain in Health Care: Hope or Hype?

Rania El-Gazzar[1], PhD; Karen Stendal[1], PhD

University of South-Eastern Norway, Hønefoss, Norway

Corresponding Author:
Rania El-Gazzar, PhD
University of South-Eastern Norway
Bredalsveien 14
Hønefoss, 3511
Norway
Phone: 47 31009151
Email: Rania.El-gazzar@usn.no

Abstract

There has been an increasing interest in blockchain technology from the health care sector in the last couple of years. The value proposition for using blockchain technology in the health care sector is to share sensitive patient data among health care entities securely and to empower patients. Blockchain technology allows patients to have an active role in developing and updating their own patient data. However, is blockchain technology really the silver bullet it seems to be? With this paper, we aim to understand the benefits and challenges of blockchain technology in the health care sector. We discuss innovation and security implications concerning blockchain technology in health care. Furthermore, we show that there is a need for more use cases to ensure the secure sharing of data within the health care sector. In our opinion, blockchain technology will not solve the issues encountered by the health care sector; in fact, it may raise more issues than it will solve.

(J Med Internet Res 2020;22(7):e17199) doi:10.2196/17199

KEYWORDS

blockchain; health care; innovation; security; implications

Introduction

Beyond its first application as the peer-to-peer payment system Bitcoin, blockchain technology is anticipated to revolutionize industries and sectors [1,2]. The implementation of blockchain technology has been clearly dominant in the financial industry [3], the supply chain industry [4], the payments industry [5,6], and e-commerce [7]. The health care sector can potentially benefit from blockchain technology by making health care information systems patient-centric and facilitating health data–sharing securely and efficiently [8]. A plethora of studies have proposed various potential use cases for using blockchain in health care [8-10]; however, the vast majority of those proposed use cases were not implemented [10]. In their recent literature review, Hasselgren et al [8] analyzed 39 studies on blockchain in health care that proposed solutions that were implemented as proof of concept.

The common use cases that benefit from blockchain-based solutions for a patient-centric health care information system include patient-managed health records, enhanced insurance claim processes, enhanced health care research, and advanced medical records shared among patients and health care providers [11]. Despite the suitability of blockchain solutions for problems and innovation needs in health care information systems, the feasibility of fully implementing those solutions is scarce to moderate [8]. Even if some solutions have proven feasible to implement into practice, they require reductions in data size and operating costs, as well as better protection of personal information to maintain privacy and security [12].

Further, barriers to the feasibility of fully implementing blockchain-enabled patient-centric electronic medical records include interoperability and scalability issues [9]. Interoperability issues manifest in the lack of standards among various blockchain-based solutions [9]. However, proposals have been introduced to address those issues [8]. Due to the high volume of clinical data, scalability issues arise, as blockchain-based solutions have data size limitations [9,13]. Patient engagement seems to be a benefit of blockchain-based solutions in health care; however, it is likely not the case for all types of patients, as not all patients are enthusiastic about managing their own data [9].

Blockchain-based solutions have not been adequately assessed for their compliance to individuals' rights protected by the General Data Protection Regulation (GDPR) [8,14]. The abovementioned opportunities, barriers, and concerns have implications for patients, health care providers, and researchers.

XSL•FO
RenderX

Furthermore, the legal, security, and privacy implications deserve further exploration. This may lead to some workarounds to make the blockchain architecture GDPR compliant, which have not yet resulted in something concrete [14-16].

Through this viewpoint article, we aim to discuss innovation and security implications that blockchain brings to the health care sector. We did not attempt to conduct a thorough literature review on the topic, as there are already comprehensive literature reviews on blockchain in health care [8-10,14]. Rather, we discuss our viewpoint on the implications of blockchain for the health care sector, with support from the current body of literature.

Thus, we aim to answer our research question: What implications may blockchain technology bring to the health care sector?

First, we present a background on blockchain technology. Second, we discuss implications of innovating with blockchain in health care. Third, we discuss the security implications of blockchain in the health care sector. Fourth, we summarize our discussion, and finally, conclude this paper.

Background of Blockchain Technology

Blockchain technology is described as a disruptive innovation that brings opportunities and challenges to various industries and sectors, and it deserves further exploration [17-19]. There is an ongoing debate as to which came first, blockchain or Bitcoin [1,2,20]. Blockchain is the underlying technology with broader capabilities and characteristics. Bitcoin is just an application area for trading that inherits blockchain characteristics [21]. Stuart Haber and W Scott Stornetta invented the notion behind blockchain first, when they proposed a framework for a "timestamping digital document" to calculate hash values that uniquely identify documents and save them in certificates with a timestamp [22]. These documents are linked by a data structure with the hashes of previous records. Nakamoto [2] adopted the framework proposed by Haber and Stornetta, creating the first Bitcoin peer-to-peer payment system based on timestamped blocks of transactions, which are chained using the hash values of previous blocks. Bitcoin then became commonly known as a means for trading with cryptocurrency [23].

Swan [13] defined blockchain as a decentralized transparent ledger with transaction records. Blockchain contains a set of data blocks, each of which contains data on multiple transactions (ie, transactions list, timestamp, nonce, hashes of the transactions and their root hash or block hash, and the hash of the previous block). As more blocks are added to the chain, the distributed ledger becomes a complete transaction history book [24]. Before adding the new transactions to the ledger, the consensus mechanism is applied by multiple participants to validate the transaction and the block. Transactions reside in the block for a specified time until the consensus process is done. Then, the block of transactions is stored in the ledger, where the information cannot be changed [24]. If the hash of a block is modified, the block is no longer valid [25], which makes subsequent blocks invalid as well, and this will require verifying the block after recalculating its hash and the hashes of subsequent blocks [26].

There are two basic deployment forms of blockchain; these are public permissionless and private permissioned blockchains [1,27]. Public permissionless blockchains are open and decentralized, where anyone can join and leave the network as reader and writer at any time (eg, Bitcoin). The network has no central authority to monitor it and no one owns and controls the network. Private permissioned blockchains only authorize a limited set of readers and writers (eg, Hyperledger). The network has a central authority that assigns the right to individuals to read and write operations.

Several definitions of blockchain mainly refer to the characteristics of public permissionless blockchains, such as absolute immutability, anonymity, decentralization in running the consensus mechanism, and openness [1,21,26]. The definitions do not provide a description of private permissioned blockchains, which are managed by a central trusted authority that controls the consensus process, in which the identities of participants are predefined and access permissions are restricted [28].

Blockchain technology is claimed to be an "accelerating force of innovation" that promises a wide range of benefits [17]. However, the claims about blockchains being tamperproof and offering strong security are challenged by a long list of security threats [29]. Blockchains are claimed to be immutable and unable to be hacked, but this has been proven invalid [30,31]. Furthermore, blockchains are energy-consuming, which entails considerable costs (eg, network performance problems) [25,32]. This poses a concern about whether the benefits and promises brought by blockchain can be taken for granted or whether they will become a threat to the ambitions for innovation and better security. Consequently, practice and academia still have questions to address regarding benefits and risks that arise from blockchains, including whether blockchain is a radical or incremental innovation in nature [17].

Innovation and Security Implications of Blockchain in Health Care

Innovation and Security Needs

Whether blockchain is a blessing or a curse, in terms of innovation and security, it is a matter of what it adds to a no-blockchain situation. It can add information technology (IT) or business costs or complexity [33]. An ever-present challenge for any industry or sector is the balance between implementing modern IT solutions and keeping information assets safe from security threats. Blockchain technology emerged with the promise to address this challenge; it enables innovation by implementing a modern decentralized information infrastructure [17,19,34].

The health care sector has a long history of heavy regulation and bureaucratic inefficiency that has decelerated its innovation [35-37], and an increasing number of data breaches have been reported in recent years [38]. It is claimed that innovating with blockchain ensures the privacy and security of highly vulnerable

XSL•FO
RenderX

and sensitive patient data in the cyber world [35]. However, there are more experiments of proposed blockchain solutions than full implementations in the health care sector [36,39,40].

Implications of Innovating With Blockchain in Health Care

Blockchain is a disruptive innovation that can leverage health care information systems' abilities to improve patient care; however, this has considerable regulatory, financial, and operational implications [41,42]. Private permissioned blockchains are a proper option for the health care sector to deal with sensitive patient data [8,27]. This type of blockchain deployment has beneficial implications for the use of blockchain in health care. The use cases suggested by researchers for blockchain in health care include patient-managed medical records, improved insurance claim processes, accelerated medical research with the use of shared anonymous patient data, and an advanced health data ledger maintaining clinical transaction logs, pharmaceutical supply chains, and consent recording [8,11,43,44].

The exploration of blockchain's compliance with GDPR is scarce despite its importance [8]. Private permissioned blockchains have implications for GDPR due to the central authority controlling the network and access to personal data [45]. In the case of using patient data to support health care researchers [11], a pseudonymization technique is required to protect patients' sensitive data [16], which may pose a risk of reidentification (ie, linking the pseudonym code or metadata to the patient's health data), raising a conflict with GDPR [45]. This requires a careful consideration of the use case and the design of the blockchain-based health information system [46].

Blockchains are immutable; therefore, it is not possible to delete a block. Thus, blockchain does not comply with the requirement under GDPR that stipulates that data subjects have the right to request for their data to be erased, including health sensitive data [9,47]. A proposed workaround is to store the patient data off-chain and have the pseudonym codes stored on-chain [16,47]. However, this implies that the pseudonym code and any transaction records on the patient data that are stored on-chain would still be existent even after deleting the patient data that were stored off-chain [47]. To reverse the immutability of blockchain, a proof-of-concept prototype for a "forgetting blockchain" was proposed to delete old data from private permissioned blockchains; however, the prototype still has limitations to address [15].

Beck and Müller-Bloch [19] argued that blockchain is a radical innovation that outdates the conventional distributed systems approach with a different architecture and characteristics [19,48]. Thus, radical innovations are difficult to implement and they bring more complex challenges, which require organizational readiness and the updating of old organizational knowledge and IT infrastructure [49]. This has financial implications for using blockchain technology in the health care sector; despite the fact that it has the potential to improve the quality of medical services, it may create financial uncertainties [41]. The top challenges facing the adoption of blockchain in health care include computational overhead, lack of interoperability and standardization, privacy concerns, and the uncertainty about who is responsible for the cost of technology implementation and who profits from it [14,50]. Barriers to adopting blockchain in the health care sector include immaturity of the technology itself, insufficient skills to understand and implement it, lack of buy-in, and lack of clear return on investment [51]. The lack of buy-in goes back to the unfamiliarity of blockchain, the negative attitudes of medical doctors toward the use of blockchain [52], and the fact that not all the patients are interested in managing their health records [9].

Beck and Müller-Bloch [19] suggested that in order to manage a radical innovation with blockchain, 3 competencies are needed to realize its benefits: discovery, incubation, and acceleration. Discovery refers to recognizing and articulating the blockchain opportunities and building research communities. Incubation involves designing blockchain use cases and experimenting with them (ie, proof of concept). Acceleration involves proposing the blockchain implementation and investing in the implementation of a full-functioning blockchain logic and infrastructure. The proof of concept of blockchain technologies strives to replicate real-world conditions in order to evaluate the feasibility of blockchain in health care and address its challenges [8,12]. Even though the required improvements are tested and have provided successful results, they come at the expense of other important aspects in the health information system. A proof of concept for a blockchain-based patient-centric information exchange between patients and providers has provided promising results; however, the real-world implementation is expected to provide different results [53]. Using blockchain to improve health data exchange and patient engagement can come at the expense of performance due to the dynamic regeneration of smart contracts. Additionally, the experimentation files will never be at the same size as the actual patient data [53]. Data size is considered to be one of the important considerations for the feasibility of blockchain solutions for health care [12].

Approaches to implement blockchain in the health care sector can be evolutionary or revolutionary. The evolutionary approach involves integrating blockchain with legacy electronic health records systems, which can compromise the availability of patient information and cause the relaxing of security countermeasures [14]. The revolutionary approach is a bottom-up approach that aims at building the entire health care information system as a blockchain-enabled system and then migrating to it [14]. Both approaches create uncertainties around the cost of implementing or integrating blockchain-based solutions for health care and provide unclear returns on investment. It has been claimed that implementing blockchain-based solutions negatively affects the financial metrics in the short term but pays off in the long term [41]. Reducing operating costs is an important consideration to test the feasibility of blockchain solutions [12].

With blockchain technology, transactions are processed and verified by an automated programmable logic with predefined rules, which reduces transaction costs (ie, effort and time spent on bureaucracy) [17,54]. The consensus mechanism of blockchain ensures the integrity of the data, but calculating the hashes for a single block in the chain is time-consuming and energy-consuming [13,25,32]. Consequently, complex or

XSL•FO
RenderX

computation-intense systems are not the best use cases for blockchain [18]. For health care information systems, performance, real-time communication, coordination, data sharing, and medical service availability are critical in life-threatening situations [55].

Some of the major challenges with the current health information systems are interoperability, integration complexities, and the inability of current legacy systems to communicate directly and share health records [40,56,57]. However, the use of blockchain in health care is found to have interoperability challenges, and blockchain, as a radical innovation, brings integration and implementation complexities [9,10,19]. Even though there are proof-of-concept suggestions and experiments to improve the interoperability, the challenges still exist [8]. This implies an unclear difference between no-blockchain and blockchain situations.

Security Implications of Blockchain in Health Care

It is claimed that private permissioned blockchain deployment brings the most benefits for health care applications [8,9]; however, it brings security risks at the same time [11,41]. Private permissioned blockchains are limited to trusted and predefined participants, and a central authority manages the rights to read and write operations of the blockchain [27]. This feature provides more control by assuring that only authorized participants can perform read or write operations on the patient data [9]. This has positive implications for the confidentiality and integrity of the data. Additionally, the immutability enables tracking of patient-generated data for medical research purposes, transactions on insurance claim processes to detect fraud, and pharmaceutical supply chains for quality assurance [11,41]. Private permissioned blockchain can also enable the availability of audit trails and progress traceability.

In the case of using patient-generated health data for research purposes, smart contracts enable patients to give consent and permission for researchers to access their health data [58]. However, data integrity can be compromised, as the patient data entry point, which is the patient's device, can be used to impersonate the patient [59]. Sharing patient health data with researchers poses a threat to the privacy of the patient; even if the data are pseudonymized, there is a risk of reidentification [60]. However, the attempts to enhance patient privacy in blockchain environments and design blockchain features for privacy are still in the pilot phase, and there is no guarantee they will preserve privacy [60,61].

Private permissioned blockchains are most prone to a 51% attack [21]. This happens when the central trustworthy node is compromised by the attacker; since the validation of the transactions is centralized, the attacker gains the authority to control the computational power of the network, causing a transaction to happen twice. Hence, the integrity of the transaction data is affected and the resources of the network are exhausted. This has negative implications on the integrity of the data and service availability, which are critical for health care applications [11].

Private permissioned blockchains have limitations in saving patient data with transaction data for the purpose of preventing distributed denial-of-service (DDoS) attacks [12]. This represents an obstacle, as the volume of patient health data is growing over time [12]. Addressing the data size limitation in private blockchain would need to accommodate the increasing volume of patient data, exposing the network to DDoS attacks [12]. Additionally, validating a block of a large data size consumes much power and entails further operational costs [12]. In either case, the service availability, which is critical for health care services, would be compromised.

The security of patient health data with blockchain technology is still in its proof-of-concept phase, and security and privacy are not fully guaranteed so far. The attempts to address security and privacy of blockchain in health care appear to be at the expense of other important features of blockchain technology itself or the needs of the health care sector [62].

Discussion of the Implications

In this viewpoint paper, we have examined the various innovation and security implications of using blockchain technology in the health care sector. Based on that, we revisit our research question: What implications may blockchain technology bring to the health care sector?

Blockchain technology is not new; however, exploring the feasibility of blockchain applications for the health care sector is in its infancy. The current state of innovating with blockchain in health care is in the proof-of-concept phase [8,58]. Blockchain is a technological innovation that brings benefits and challenges [42]. The health care sector is expected to benefit from blockchain in terms of empowering patients and increasing immutability and traceability [11,41,58]. The needs of the health care sector include sharing vast amounts of patient health data across involved entities (ie, interoperability), regulatory compliance (eg, GDPR), data confidentiality, data integrity, privacy, and data and service availability. The feasibility of using blockchain in health care is dependent on the capability of storing and processing vast amounts of patient health data, ensuring privacy, and reducing operating costs [12,46]. Customizing private permissioned blockchain solutions to fit the needs of the health care sector may result in manipulating the characteristics of blockchain technology [15] or manipulating the needs of the health care sector [12,46]. This is a known trade-off approach happening in the proof-of-concept attempts to apply blockchain solutions in health care [62]. This trade-off involves compromising between two desirable but incompatible features. For instance, complying with GDPR requirements involves manipulating the immutability of blockchain. Additionally, the block size limitation in blockchain is intended to reduce performance overhead and prevent DDoS attacks. However, this compromises the scalability needed to accommodate the vast amount of patient health data [12], and manipulating the size of patient health data is difficult. On the other hand, if the blockchain is designed to process large data, it will cause extra operating costs due to the performance overhead, and it will expose the network to DDoS attacks.

We see the need to distinguish between the benefits and challenges that are unique to blockchain and those that are common across other technological innovations. For example,

XSL•FO
RenderX

interoperability is not a challenge specific to blockchain per se; rather, it is a common challenge when adopting any technological innovation.

Based on the topics and views debated in this paper, we summarize the implications of blockchain for health care in terms of both the patients and the health care providers (see Table 1).

Table 1. Blockchain implications for the health care sector.

Group	Benefits	Challenges
Patients	• Patients are empowered with self-sovereignty through self-managing personal patient-generated health data • The identity of the patients is anonymized	• Some patients may not be interested in self-managing their health data
Health care providers	• Providing a decentralized database with identical copies of the same complete health information, which is made accessible to all parties in the health care chain • Facilitating collaboration and data sharing • Claimed immutability of a transaction's history	• Interoperability is a challenge, and complex systems are not the best use cases for blockchain • DDoS[a] attacks are likely to happen and affect the availability of the patient health data • A 51% attack, specific to blockchain, affects the integrity of transactions' data and consumes the network resources • Compliance issues with GDPR[b] • Blockchain can be resource consuming when all entities in the chain have to approve a large-sized data block

[a]DDoS: distributed denial-of-service.

[b]GDPR: General Data Protection Regulation.

We argue that blockchain technology is surrounded by a controversy between marketing hype and realistic criticism. The marketing hype has manifested in the claim that blockchain is immune to common security attacks that threaten data confidentiality, integrity, and availability. Meanwhile, there are realistic criticisms that show that blockchains are hackable in many ways [21]. A comprehensive list of security threats in blockchain and their causes has already been rendered [29]. This serves as an incentive for research on security improvements for blockchain in general as well as in health care [14,21].

In terms of medical research, there is potential for using blockchain technology. It is possible to store participants' informed consent to ensure a more transparent, traceable, and tamperproof research method for medical research [63].

This work has implications for further research. Health care researchers can benefit from anonymized health data, which can be shared and aggregated to generate new insights into improving patient health or health care services while maintaining the privacy of patients. Further research is needed to increase awareness about blockchain and to clear the misconceptions and the hype around it. More context-specific use cases need to be designed to avoid generic arguments about blockchain's applicability across sectors. Future research efforts can aid health care providers in developing the required competencies to innovate with blockchain (ie, discovery, incubation, and acceleration) [19].

Conclusion

Throughout this paper, we have presented and discussed various views on blockchain technology and the positive and negative issues related to it. Blockchain technology is regarded as a promising technology for securely sharing health data. However, it is not clear if blockchain is really the solution to all the issues regarding highly sensitive data.

Throughout this work, we have highlighted the myths and important challenges concerning blockchain technology. Further, we have questioned the applicability of blockchain technology to the health care sector. Governments may want to examine feasible scenarios in which to use blockchain in the health care sector as well as the challenges associated with the traditionalism of such a sector and the immaturity of blockchain. This requires a careful consideration of the trade-offs that may be made when designing and implementing blockchain solutions for health care.

In this paper, we identified blockchain technology's positive and negative implications for patients and health care providers, which opens up unlimited opportunities for future research to delve into.

Conflicts of Interest

None declared.

References

1. Underwood S. Blockchain beyond bitcoin. Commun. ACM 2016 Oct 28;59(11):15-17. [doi: 10.1145/2994581]
2. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin 2008:1-9 [FREE Full text] [doi: 10.2139/ssrn.3440802]

XSL•FO
RenderX

3. Eickhoff M, Muntermann J, Weinrich T. What Do FinTechs Actually Do? A Taxonomy of FinTech Business Models. 2017 Presented at: Proceedings of the 38th International Conference on Information Systems; Dec 10-13, 2017; Seoul, South Korea p. 1-19 URL: https://pdfs.semanticscholar.org/0c8f/b6756bdbbb9fce9465a50ca3ebcea479e7cb.pdf
4. Korpela K, Hallikas J, Dahlberg T. Digital Supply Chain Transformation toward Blockchain Integration. 2017 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Waikoloa, HI p. 4182-4191. [doi: 10.24251/hicss.2017.506]
5. Holotiuk F, Pisani F, Moormann J. The Impact of Blockchain Technology on Business Models in the Payments Industry. 2017 Presented at: Proceedings of the 13th International Conference on Wirtschaftsinformatik; Feb 12-15, 2017; St Gallen, Switzerland p. 912-926.
6. Lindman J, Rossi M, Tuunainen V. Opportunities and Risks of Blockchain Technologies in Payments? A Research Agenda. 2017 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Waikoloa, HI p. 1533-1542. [doi: 10.24251/hicss.2017.185]
7. Frey R, Wörner D, Ilic A. Collaborative Filtering on the Blockchain: A Secure Recommender System for E-commerce. 2016 Presented at: Proceedings of the 22nd Americas Conference on Information Systems; Aug 11-14, 2016; San Diego, CA p. 1-5.
8. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134(2020):1-10 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
9. Agbo C, Mahmoud Q, Eklund J. Blockchain Technology in Healthcare: A Systematic Review. Healthcare (Basel) 2019 Apr 04;7(2):56 [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
10. Hölbl M, Kompara M, Kamišalić A, Nemec Zlatolas L. A Systematic Review of the Use of Blockchain in Healthcare. Symmetry 2018 Oct 10;10(10):470-422. [doi: 10.3390/sym10100470]
11. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
12. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility. J Med Internet Res 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
13. Swan M. In: McGovern T, editor. Blockchain: Blueprint for a New Economy. Sebastopol, CA: Reilly Media; 2015:1-152.
14. Al-megren S, Alsalamah S, Altoaimy L, Alsalamah H, Soltanisehat L, Almutairi E, et al. Blockchain Use Cases in Digital Sectors?: A Review of the Literature. 2018 Presented at: IEEE Conf Internet Things, Green Computing and Communications, Cyber, Physical, and Social Computing Smart Data, Blockchain, Computer and Information Technology Congress Cybermatics; July 30-Aug 3, 2018; Halifax, NS p. 1417-1424. [doi: 10.1109/cybermatics_2018.2018.00242]
15. Farshid S, Reitz A, Roßbach P. Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; Jan 8-11, 2019; Maui, HI p. 7087-7095. [doi: 10.24251/hicss.2019.850]
16. Lima C. Blockchain-GDPR Privacy by Design How Decentralized Blockchain Internet will Comply with GDPR Data Privacy. IEEE Blockchain Group. 2018. URL: https://blockchain.ieee.org/images/files/pdf/blockchain-gdpr-privacy-by-design.pdf [accessed 2002-03-20]
17. Avital M, King J, Beck R, Rossi M, Teigland R. Jumping on the Blockchain Bandwagon: Lessons of the Past and Outlook to the Future. 2016 Presented at: 37th International Conference on Information Systems; Dec 11-14, 2016; Dublin, Ireland.
18. Glaser F. Pervasive Decentralisation of Digital Infrastructures: A Framework for Blockchain enabled System and Use Case Analysis. 2017 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Waikoloa, HI p. 1543-1552. [doi: 10.24251/hicss.2017.186]
19. Beck R, Müller-Bloch C. Blockchain as Radical Innovation: A Framework for Engaging with Distributed Ledgers. 2017 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Waikoloa, HI p. 5390-5399. [doi: 10.24251/hicss.2017.653]
20. Om Bitcoin Internet. Norges Bitcoin- og Blockchainforening. 2019. URL: https://norgesbitcoinforening.no/bitcoin/om-bitcoin/ [accessed 2019-02-04]
21. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where Is Current Research on Blockchain Technology?-A Systematic Review. PLoS One 2016;11(10):e0163477 [FREE Full text] [doi: 10.1371/journal.pone.0163477] [Medline: 27695049]
22. Haber S, Stornetta WS. How to time-stamp a digital document. J Cryptology 1991 Jan;3(2):99-111. [doi: 10.1007/bf00196791]
23. Kazan E, Tan C, Lim E. Value Creation in Cryptocurrency Networks: Towards A Taxonomy of Digital Business Models for Bitcoin Companies. 2015 Presented at: 19th Pacific Asia Conference on Information Systems; July 5-9, 2015; Singapore p. 1-15.
24. Nofer M, Gomber P, Hinz O, Schiereck D. Blockchain. Bus Inf Syst Eng 2017 Mar 20;59(3):183-187. [doi: 10.1007/s12599-017-0467-3]
25. Gipp B, Jagrut K, Breitinger C. Securing Video Integrity Using Decentralized Trusted Timestamping on the Blockchain. 2016 Presented at: Proceedings of the 10th Mediterranean Conference on Information Systems (MCIS); Sep 4-6, 2016; Paphos, Cyprus p. 3.

XSL•FO
RenderX

26. Crosby M, Pattanayak P, Verma S, Kalyanaraman V. BlockChain Technology: Beyond Bitcoin. Applied Innovation Review 2016 Sep;2(6-10):5-19 [FREE Full text]
27. Wüst K, Gervais A. Do you Need a Blockchain? 2018 Presented at: 2018 Crypto Valley Conference on Blockchain Technology (CVCBT); June 20-22, 2018; Zug, Switzerland p. 45-54. [doi: 10.1109/cvcbt.2018.00011]
28. Ølnes S, Ubacht J, Janssen M. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. Government Information Quarterly 2017 Sep;34(3):355-364. [doi: 10.1016/j.giq.2017.09.007]
29. Mosakheil J. Security Threats Classification in Blockchains Internet [thesis]. St. Cloud State University. St Cloud, MN; 2018. URL: https://repository.stcloudstate.edu/msia_etds/48 [accessed 2020-06-23]
30. Orcutt M. Once Hailed as Unhackable, Blockchains are Now Getting Hacked Internet. MIT Technology Review. 2019. URL: https://www.technologyreview.com/2019/02/19/239592/once-hailed-as-unhackable-blockchains-are-now-getting-hacked/ [accessed 2019-05-02]
31. Halaburda H. Blockchain revolution without the blockchain? Commun. ACM 2018 Jun 25;61(7):27-29. [doi: 10.1145/3225619]
32. Rückeshäuser N. Do We Really Want Blockchain-Based Accounting? Decentralized Consensus as Enabler of Management Override of Internal Controls. 2017 Presented at: 13th International Conference on Wirtschaftsinformatik; Feb 12-15, 2017; St Gallen, Switzerland p. 16-30.
33. El-Gazzar R, Stendal K. Cloud Computing, Blockchain, and IOT: a blessing or curse? 2019 Presented at: Proceedings from the Annual NOKOBIT Conference; Nov 26-27, 2019; Narvik, Norway p. 1-16.
34. Schweizer A, Schlatt V, Urbach N, Fridgen G. Unchaining Social Businesses - Blockchain as the Basic Technology of a Crowdlending Platform. 2017 Presented at: Proceedings of the 38th International Conference on Information Systems; Dec 10-13, 2017; Seoul, South Korea p. 1-21.
35. Smith KJ, Dhillon G. Blockchain for Digital Crime Prevention: The Case of Health Informatics. 2017 Presented at: Proceedings of the 23rd Americas Conference on Information Systems; Aug 10-12, 2017; Boston, MA p. 1-10.
36. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: A distributed architecture model to integrate personal health records. J Biomed Inform 2017 Jul;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
37. Sullivan C, Burger E. E-residency and blockchain. Computer Law & Security Review 2017 Aug;33(4):470-481. [doi: 10.1016/j.clsr.2017.03.016]
38. The Biggest Healthcare Data Breaches of 2018 (So Far). HealthcareITNews. 2018. URL: http://www.healthcareitnews.com/projects/biggest-healthcare-data-breaches-2018-so-far [accessed 2018-01-04]
39. e-Health Records. E-estonia. 2018. URL: https://e-estonia.com/solutions/healthcare/e-health-record/ [accessed 2018-03-31]
40. Ekblaw A, Azaria A, Halamka J, Lippman A. A Case Study for Blockchain in Healthcare: "MedRec" Prototype for Electronic Health Records and Medical Research Data. 2016 Presented at: Proceedings of IEEE Open & Big Data Conference; Dec 5-8, 2016; Washington, DC p. 1-13. [doi: 10.1109/obd.2016.11]
41. Zhang R, George A, Kim J, Johnson V, Ramesh B. Benefits of Blockchain Initiatives for Value-Based Care: Proposed Framework. J Med Internet Res 2019 Sep 27;21(9):e13595 [FREE Full text] [doi: 10.2196/13595] [Medline: 31573899]
42. Clohessy T, Acton T. Investigating the influence of organizational factors on blockchain adoption. IMDS 2019 Aug 12;119(7):1457-1491. [doi: 10.1108/imds-08-2018-0365]
43. Mettler M. Blockchain Technology in Healthcare: The Revolution Starts Here. 2016 Presented at: IEEE 18th Int Conf e-Health Networking, Applications, and Services; Sep 14-17, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749510]
44. Clohessy T, Clohessy S. Blockchain Vigilant Information Systems: A Model for the Pharmaceutical Industry. In: Blockchain and Distributed Ledger Technology Use Cases. Cham, Switzerland: Springer; 2020:51-68.
45. Finck M. Blockchain and the General Data Protection Regulation: Can distributed ledgers be squared with European data protection law? European Parliament Think Tank. 2019. URL: http://www.europarl.europa.eu/thinktank [accessed 2019-05-05]
46. Hawig D, Zhou C, Fuhrhop S, Fialho AS, Ramachandran N. Designing a Distributed Ledger Technology System for Interoperable and General Data Protection Regulation-Compliant Health Data Exchange: A Use Case in Blood Glucose Data. J Med Internet Res 2019 Jun 14;21(6):e13665 [FREE Full text] [doi: 10.2196/13665] [Medline: 31199293]
47. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
48. Johansen S. A Comprehensive Literature Review on the Blockchain Technology as an Technological Enabler for Innovation Internet. Department of Information Systems, Mannheim University. 2018. URL: https://www.researchgate.net/publication/312592741 [accessed 2018-06-05]
49. Clohessy T, Acton T, Rogers N. Blockchain Adoption: Technological, Organisational, and Environmental Considerations. In: Business Transformation through Blockchain. Cham, Switzerland: Palgrave Macmillan; 2019:47-76.
50. Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood) 2005;24(5):1103-1117. [doi: 10.1377/hlthaff.24.5.1103] [Medline: 16162551]
51. IBM. Healthcare Rallies for Blockchains. IBM Institute for Business Value. URL: https://www.ibm.com/downloads/cas/BBRQK3WY [accessed 2018-07-02]

XSL•FO
RenderX

52. Hau YS, Lee JM, Park J, Chang MC. Attitudes Toward Blockchain Technology in Managing Medical Information: Survey Study. J Med Internet Res 2019 Dec 09;21(12):e15870 [FREE Full text] [doi: 10.2196/15870] [Medline: 31815676]
53. Hylock RH, Zeng X. A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. J Med Internet Res 2019 Aug 31;21(8):e13592 [FREE Full text] [doi: 10.2196/13592] [Medline: 31471959]
54. Jarvenpaa S, Teigland R. Trust in Digital Environments: From the Sharing Economy to Decentralized Autonomous Organizations. 2017 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Waikoloa, HI p. 5812-5816.
55. Esmaeilzadeh P, Mirzaei T. The Potential of Blockchain Technology for Health Information Exchange: Experimental Study From Patients' Perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
56. Engelhardt MA. Hitching Healthcare to the Chain: An Introduction to Blockchain Technology in the Healthcare Sector. TIM Review 2017 Oct 27;7(10):22-34. [doi: 10.22215/timreview/1111]
57. Zhang P, White J, Schmidt DC, Lenz G. Applying Software Patterns to Address Interoperability in Blockchain-based Healthcare Apps Internet. arXiv. 2017. URL: http://arxiv.org/abs/1706.03700 [accessed 2018-03-03]
58. Maslove DM, Klein J, Brohman K, Martin P. Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
59. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and Scalable mHealth Data Management Using Blockchain Combined With Client Hashchain: System Design and Validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]
60. Glicksberg BS, Burns S, Currie R, Griffin A, Wang ZJ, Haussler D, et al. Blockchain-Authenticated Sharing of Genomic and Clinical Outcomes Data of Patients With Cancer: A Prospective Cohort Study. J Med Internet Res 2020 Mar 20;22(3):e16810 [FREE Full text] [doi: 10.2196/16810] [Medline: 32196460]
61. Jones M, Johnson M, Shervey M, Dudley JT, Zimmerman N. Privacy-Preserving Methods for Feature Engineering Using Blockchain: Review, Evaluation, and Proof of Concept. J Med Internet Res 2019 Aug 14;21(8):e13600 [FREE Full text] [doi: 10.2196/13600] [Medline: 31414666]
62. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design Choices and Trade-Offs in Health Care Blockchain Implementations: Systematic Review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
63. Olesch A. Is There a Chance for a Symbiosis of Blockchain and Conservative Healthcare? HIMSS Europe. 2018. URL: https://www.himss.eu/himss-blog/there-chance-symbiosis-blockchain-and-conservative-healthcare [accessed 2019-08-09]

Abbreviations

DDoS: distributed denial-of-service
GDPR: General Data Protection Regulation
IT: information technology

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by T Clohessy, A Hasselgren; comments to author 16.12.19; revised version received 07.02.20; accepted 13.05.20; published 10.07.20.

Please cite as:
El-Gazzar R, Stendal K
Blockchain in Health Care: Hope or Hype?
J Med Internet Res 2020;22(7):e17199
URL: http://www.jmir.org/2020/7/e17199/
doi:10.2196/17199
PMID:32673219

©Rania El-Gazzar, Karen Stendal. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Tango of Problem Formulation: A Patient's/Researcher's Reflection on an Action Design Research Journey

Michael B Twomey[1]; David Sammon[1]; Tadhg Nagle[1]

Business Information Systems, Cork University Business School, University College Cork, Cork, Ireland

Corresponding Author:
Michael B Twomey
Business Information Systems
Cork University Business School
University College Cork
College Road
Cork
Ireland
Phone: 353 021 4902136
Email: michael.twomey@ucc.ie

Abstract

This paper reports on the reflection of the lead researcher, a 48-year-old patient with cystic fibrosis (CF), and aims to portray his real-life experience of a 10-month action design research (ADR) project. Playing a dual role, as both a patient and researcher, the lead researcher reflects deeply on his ADR experience with particular emphasis on the problem formulation stage of creating a simple yet impactful checklist to aid memory recall of CF patients or caregivers during a medical appointment. Using Driscoll's model of reflection, a real-life unsanitized ADR experience is carefully imparted via a series of 4 vignettes, including 4 key learnings, which highlight the connection between a meticulous considered approach to problem formulation and truly effective outcomes. By providing this rich account of problem formulation within ADR, it is hoped that this reflection will help researchers to better understand the complexity of problem formulation in design-oriented research; to avoid making assumptions and becoming fixated on solutions; and to move instead to an end point where several possible ways of examining a problem have been considered, explored, and understood—an end point where successful end results are reached through grit and determination. This paper advocates for the inclusion and portrayal of the actual realities or ups and downs of this dynamic and evolving stage of ADR, capturing the often-tacit knowledge of problem formulation and begetting a sense of realism and humanity to ADR serving as knowledge contributions in their own right. The lead researcher is the patient and researcher in this ADR project. This is my story!

(J Med Internet Res 2020;22(7):e16916) doi:10.2196/16916

KEYWORDS

action design research; patient; reflection; problem formulation; checklist; cystic fibrosis

Introduction

It is a windy Saturday afternoon in February 2020, and I am busy packing up a box of checklist booklets that I designed to aid the memory recall or information retrieval of patients with cystic fibrosis (CF) and their caregivers during their medical appointments. The box is on its way to the Royal London Children's Hospital, Whitechapel, London, for distribution to caregivers of patients with CF. Caregivers know the reality of having a sick child in a medical appointment, the stress of trying to remember medical history, and the difficulty of trying to recall afterward what took place within a medical appointment. I tape up the box carefully, making sure it is secure for its journey ahead.

A journey that would not have come to pass had I not been invited to present my action design research (ADR) study at the International CF Clinical Conference held in Killarney, Ireland, on January 30, 2020. Out of that, came the dispatching of the checklist booklet to Spain, Sweden, Israel, and Australia. There is no doubt that the checklist booklet is beginning to travel far and wide. Earlier, in January 2019, the booklet was distributed by Cystic Fibrosis Ireland to every patient or caregiver within the Republic of Ireland (1300 patients with CF). Afterward, in April 2019, it was shipped to the Czech Cystic Fibrosis Association and Austria for review.

Moreover, in October 2019, I was invited to present my research on the booklet to over 100 clinicians at a *Hot Topics in CF* conference held in Birmingham, England. Nevertheless, the real

XSL•FO
RenderX

impact of the checklist booklet is best depicted by the mother of a 7-year-old child with CF:

> *We just wanted to say we received our medical appointment check list today, and we just wanted to say THANK YOU so much, we love it and it's going to be incredibly handy for us, although it's just a book to our little boy now, in a few years he'll know how great and simple it is as well.*

As I contemplate the same, a warm feeling envelops me, I feel I am beginning to make a difference, a difference to people like me.

The box is ready to go, a thought enters my mind, *So was my ADR journey easy*, the answer is definitely no. *Did it take determination and patience?* Yes, for sure! *And did I sometimes think of giving up? Did it test me to my limits?* I would have to say yes on both accounts. So how do I account for the success of the checklist booklet? Earlier I mentioned determination; I was unyielding in my quest to understand the problem, the problem that I and others like me experience every time we are in a medical appointment. To help you understand, however, I need to take you on a voyage, a voyage of reflection. It is only by coming on this journey that you will come to comprehend the end point, where through grit and determination, success can be achieved.

"Humans have always reflected on experiences and feelings" [1]. According to Boud et al [2] reflection is "an important human activity in which people recapture their experience, think about it, mull it over and evaluate it." As a researcher, I felt an obligation to share my unsanitized lived experience of *problem formulation* in ADR, not only as a practitioner but also through the eyes of a patient researcher living with a chronic illness.

To this end, I found that a reflection would be the most appropriate instrument to aid me with the *mental process of trying to structure or restructure* [3] my real-life experience of *problem formulation*, to put together or capture the *existing knowledge or insights* [3] from the project that I lived through over a 10-month period. I hope that the insights imparted herein may serve not only as insightful to ADR practitioners in their *problem formulation* endeavors but also to highlight the importance of this stage of ADR in achieving successful outcomes. Moreover, I hope that any patients reading my reflection may be inspired to enter the stimulating world of research, making real-world impacts within their own patient communities, as I have strived to do in mine.

The paper is structured as follows. First, I present a very brief background on ADR, the methodology that I used in my research exploration, going somewhat deeper on the *problem formulation* stage of the methodology (the focus of my reflection). Next, I endeavor to set the scene, giving the reader a deep candid sense of the patient researcher behind the reflection, followed by a very brief section on why a reflection was the correct tool for my deliberations, and the rationale behind the model of reflection I selected. I subsequently organize my reflection through a series of 4 vignettes, which are used to explain the lessons that I learned from my experience of *problem formulation* within ADR and how crucial this was to the effects my research is having now. Finally, I bring my musings to a close in the concluding remarks section of the paper.

Background

Action Design Research

Design science research (DSR) accentuates a *construction-oriented* interpretation of information systems (IS) research, which at its core lies the design and build of innovative information technology (IT) artifacts and which is deemed appropriate when research aims to produce artifacts that address so called *wicked problems* or ill-structured problems [4]. Although this approach provides IS researchers with the ability to go beyond mere elucidation, toward research that spawns design knowledge relevant to practitioners [4], it still fails to "fully recognize the role of organizational context in shaping the design as well as shaping the deployed artefact" [5].

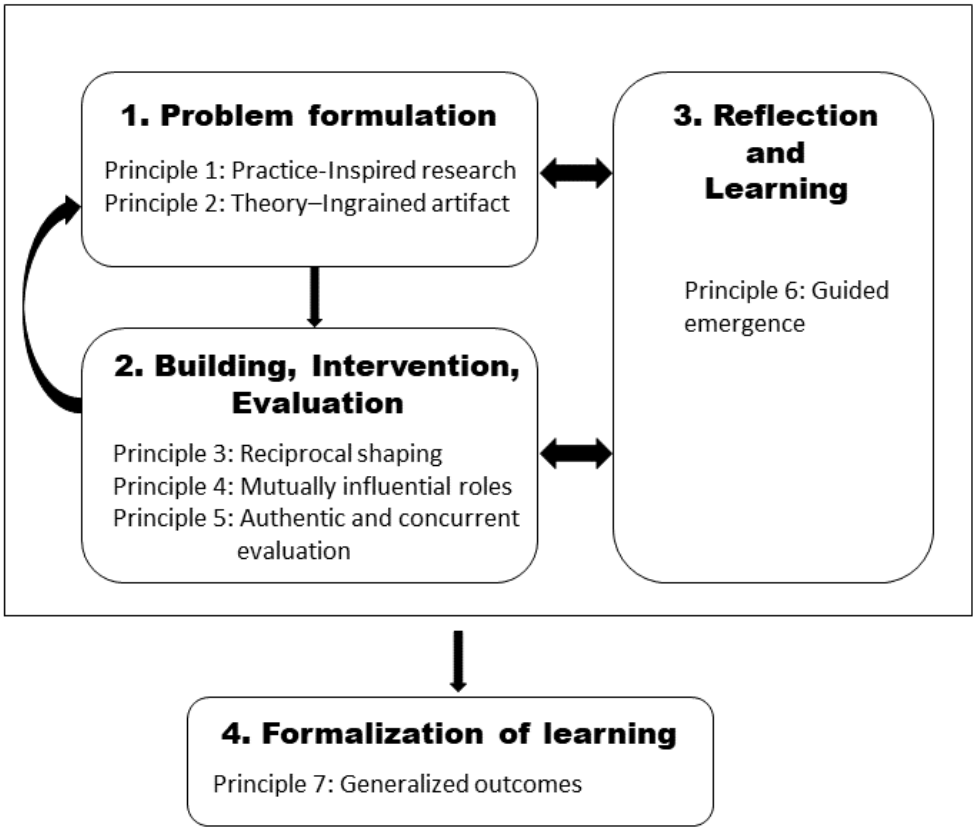
In their seminal paper, Sein et al [5] proposed a variant of DSR (Figure 1), which is called action design research (ADR), and clearly acknowledged IT artifacts as "shaped by the interests, values, and assumptions of a wide variety" [6] of stakeholders at the same time retaining the essence of design research. ADR targets the creation of innovative artifacts in an *organizational context* but at the same time produces knowledge contributions from the intervention while tackling a problematic situation [4,7]. Sein and Rossi [8] argue that the "embedding of the context in the design through intervention in an organisation, a single-entry point (problem-centered), and, inductive epistemology, is the characteristics of ADR that validate knowledge claims of emergent knowledge co-produced with practice."

ADR differs from other design approaches in so far as it draws on DSR, which centers on the utility of an artifact, and action research, which primarily focuses on learning from an environment, believing that "[t]he only way to produce conditions of practice is to move to practice" [9]. Moreover, at the core of ADR is an inquiry with rigorous evaluation, which is highly iterative in nature, consisting of nested loops [8], where each iteration concludes with a consideration of the artifact. This evaluation acts as the impetus for thorough *reflection and learning*, which then feeds back into *problem formulation*, thereby challenging "organizational participants' existing ideas and assumptions about the artefact's specific use context in order to create and improve the design" [5]. It is these very characteristics that make ADR so successful as a methodology.

It comes as no surprise, therefore, that ADR has been used very effectively in a wide array of research projects, and "because of its ever-expanding applications, the ADR concepts and process model continue to grow and evolve to meet the demands of new and challenging environments" [10], including those within the health domain. For example, the successful solution by Bretschneider et al [11] helps to leverage the innovative idea potential of patients better than traditional communication forums.

XSL•FO
RenderX

Figure 1. Action design research method: stages and principles.



As a pragmatist who is "more interested in utility and usefulness than in an abstract notion of truth" [12], I deemed ADR as a suitable methodology for my research endeavors as it is grounded in and grounds research from practice, academia, and empirical data. Its primary focus is on learning from designing an artifact or intervention within an environment. This is important as it was imperative that the solution that I created worked within the complex environment of the medical appointment. My reflections later on in this paper reveal my experience of ADR as a patient and researcher and serve to contribute to the recent open and engaging discussions regarding problem *formulation* in ADR, consistent with the original authors' contention, who contend that ADR "is still an open endeavor" [13]. First, let us delve into the *problem formulation* stage of ADR in greater depth.

Problem Formulation

In 1998, Berthon et al [14] stated that *problem formulation* was the least researched of problem-solving activities. In 2018, Mullarkey and Hevner [10] reported on the challenges they had regarding the *problem formulation* stage of ADR. More specifically, they discovered that they "needed to better understand the problem space" [10] and maintained that the ADR model by Sein et al [5] required an explicit *diagnosis* stage (with a clear separation from design) model "to analyse the importance of the problem domain and the relevance of the IT solution class to research and practice" [10].

What Is Problem Formulation?

Simply put, *problem formulation* is the sum of problem ID (perception), problem definition (conceptualization), and problem structuring (instrumental reasoning). "The first requirement with any complex problem is to try and understand it as a totality. How has it arisen, and why? Where is it going and what route is it taking? Is it changing its nature or structure as it develops?" [15]. The *problem formulation* stage in ADR (Textbox 1 presents the tasks in the stage) identifies and conceptualizes (using both divergent and convergent metacognitive processes) a research opportunity based on existing theories and technologies [4], where the research activity is said to be problem inspired [16,17].

How important is the *problem formulation* stage of ADR? The value of a suitable definition has been established empirically [18]: "The more of the context of a problem that a scientist can comprehend, the greater are his chances of finding a truly adequate solution" [19]. Mintzberg et al [20] argue that *problem formulation* is "probably the single most important routine, since it determines in large part, however implicitly, the subsequent course of action." However, Mitroff et al [21] maintained that problem forming and defining are as critical, if not more so than, problem solving. This is probably not surprising as our understanding of a problem greatly influences our selection of solutions [22] and helps avoid type III errors, solving the wrong problem [23].

XSL•FO
RenderX

Textbox 1. Tasks in the problem formulation stage of action design research [5].

- Identify and conceptualize the research opportunity
- Formulate initial research questions
- Cast the problem as an instance of a class of problems
- Identify contributing theoretical bases and prior technology advances
- Secure long-term organizational commitment
- Set up roles and responsibilities

Why Is Problem Formulation So Challenging?

According to Mitroff and Featheringham [24], one of the most important challenges of the problem-solving activity is solving the *wrong* problem by adapting a formulation that is either too narrow or inappropriate. So, one may well ask why *problem formulation* is so challenging? In Textbox 2, I have tried to encapsulate some of the key challenges to *problem formulation* that have been reported in literature.

How Might Problem Formulation Be Done Better?

In their ADR process model, Mullarkey and Hevner [10] argued that every iteration should go through *problem formulation* and that reflection and learning should also be executed in every cycle—mainly as it informs the *problem formulation* for the next ADR iteration. They also included a fourth ADR stage *evolution*, as a means to address the very temporal and evolving nature of both the artifact and the problem environment.

Sein and Rossi [8] agreed that these modifications were a *valuable* addition to their initial model (which did not go into details) in terms of making it more *transparent* and *accessible to researchers;* they do however point out that this was intentional "because in launching a new method of doing DSR, we wanted to keep it at a broad enough level of abstraction to allow the actual application of the method to emerge in use.".

Textbox 2. The challenges of problem formulation.

- Problems, by their very nature, are complex and ill-structured [25-27].
- "In real life there is not a single, static, well-defined problem, but a constantly changing problem." [28]
- "The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required." [29]
- Stakeholders may have conflicting interpretations of a problem resulting from different life experiences, competencies, goals, and values [23].
- Human biases: "fixated on these unwarranted assumptions, and this fixation interferes with the insight needed to solve the problem" [30].
- People often are too quick to move on to evaluative stages of problem solving rather than gaining a complete understanding of the problem [31].
- "We may be too ready to re-use features of known existing designs, rather than to explore the problem and generate new design features." [32]
- People often only identify the most obvious symptoms, or those to which they are most sensitive, resulting in the problem being described inappropriately [24,27].
- Most companies are not adequately thorough in actually defining the problems they are trying to solve [33].
- "Problem formulation" has been shown to be highly dependent on the mode of problem presentation [34-36].
- Changing problem presentation modes has considerable effects on mental model formation [34], where Simon and Hayes found that "innocent changes in language had major effects on problem formulation." [37]

When several stakeholders are affected by a problem, all viewpoints must be taken into consideration for a solution to be deemed successful [23]. After declaring the problematic gap, it is advisable to specify evidence supporting the presence of that gap. Indeed, Brody [38] raises the question, should problem statements include an "as evidenced by" clause? Moreover, Mitroff et al [21] advocated the use of assumptional analysis to question any assumptions, projections, and explanations lying beneath the problematic statement, whereas Lyles and Mitroff [25] also proposed that alternative views of the problem be sought to improve *problem formulation*.

The use of conceptual processing and mental models is also encouraged in DSR literature [39,40] to assist the *problem formulation* stage. Interestingly, Lesgold [41] discovered that experts expended additional time deciding an appropriate mental model of a problem than did novices. This may be explained because "what we understand and how we understand a situation depends on the information we bring to a given situation, and the longer we think about the situation the more its cognitive representation changes. It may be assumed that cognitive elaboration activates more schemata" [42] and hence enriched *problem formulation.*

Like any good story, it is important to first set the scene, facilitating an appreciation for key contextual elements of the story that I feel are important to comprehend, and so, we now move to the next section, that is, the research setting.

XSL•FO
RenderX

Research Setting

It is 11:27 AM, it is a pleasant day outside, the sun is shining, summer shows visible signs of its arrival in the garden outside, with many perfumes and aromas creating an exciting olfactory feast. It seems appropriate that I start this paper on this date, May 25, 2019, as it is my sister Jane's 50th birthday, and she was one of the key motivators and sources of inspiration for my research journey. Unfortunately, I cannot celebrate this special occasion with her as she passed away with CF on July 29, 1997. I feel an intense sense of sadness mixed with emotions of happiness as the memories of her sail across the horizon of my memory.

I have witnessed the effect that her passing has had on my family. I have seen the agony and the physical effects that the death of my sister has had on my dear parents. I will never forget that day. I will never forget the life-support machine flat-lining, the tears, the pain in that intensive care unit; it will live with me forever. The experience of her end has left an indelible mark on me not only because of the pain of her passing but also as I have CF myself. I have also lived through the hardship that one endures with CF. I have faced the dark shadows that come into a room when gasping for breath, where one's mortality becomes all too real.

After my sister's death, I had a choice; to let this disease define me and become a bitter, negative person who craved self-pity or to embrace the positive aspects that the disease had carved out or sculpted into my heart and mind. You might ask what do I mean by positive? I mean the appreciation for life, for family, for fun, and for being able to breathe. I mean the ability to empathize with others and to be compassionate to another person's suffering. I wanted to make a difference, to give something back, as others have given to me, which has resulted in my own good health. I wanted to help others with CF who are going on their own difficult journey, a passage that has many dark and difficult days.

Although I admit that living with CF is not always easy, I have always been a fighter, I had to be! I would not give into myself or my condition, I love life too much. Life has so much to offer, alas many take their lives for granted. Moreover, I think when you are faced with the very serious question of your mortality at a young age, you learn to be truly grateful for the gift of life. That is the real reward of a chronic condition, and it became the match that ignited the fight and passions within me.

My appetite for learning and wanting to help other patients with CF and their caregivers was really kindled in October 2014 when I returned to postgraduate executive education, and it continues to burn brightly. For those who know me best, it probably comes as no surprise that I have become a researcher. Even at an early age, I was quite inquisitive and sought to explore and understand the world around me. However, it may surprise you to know that my first attempt at research failed miserably. I was just 7 years old, and I was trying to make sense of how one could get a liquid (in this case, petrol) to rise in a tube. I thought it fascinating until I imbibed or inhaled a gulp of it and also flooded the neighbor's driveway. I was the talk of the neighborhood for about a week. Thankfully, it did not end in complete disaster, nor did it quell my inquiring mind, but I am sure I scared my poor parents half to death. This event became known as the *petrol episode* (Figure 2, in yellow).

It is a day that we at home do not speak of too often, except to remind me that I am capable of some awful blunders and that research does not always go according to plan. However, I like to think that that day in May 1979 gives you the reader a little glimpse of the character that I am, spirited, curious, and not afraid to explore!

Many would contend that, as a patient with CF with over 47 years of experience living with CF, I was the perfect candidate to research or explore and understand the associated problems of memory recall within the medical appointment. And yes, there is no doubt that I had the ability to impart rich CF-related information to the project; however, quite quickly, I became very cognizant of the fact that my illness journey (although sharing similarities with others) was my own unique voyage, with all the biases of any individual. Recognizing this and taking advantage once again of the robust ties that I had within the CF community, I enrolled 2 caregivers of patients with CF into a design team and a clinician specializing in CF to reduce the bias that I brought to the ADR project and to enrich the knowledge of the collective.

My patient-inspired investigation was driven by what eventually became the following motivation: how might I augment the memory recall or information retrieval of patients with CF and their caregivers within the elicitation and elucidation phases of the medical appointment?

Having attained ethical approval from the University College Cork Social Research and Ethics Committee, my primary focus eventually (discussed later) became the design of a checklist artifact over a 10-month period (September 2016 to April 2017). A checklist that would not only ameliorate the challenges of memory recall within the CF medical appointment but also augment our actual comprehension of these same challenges.

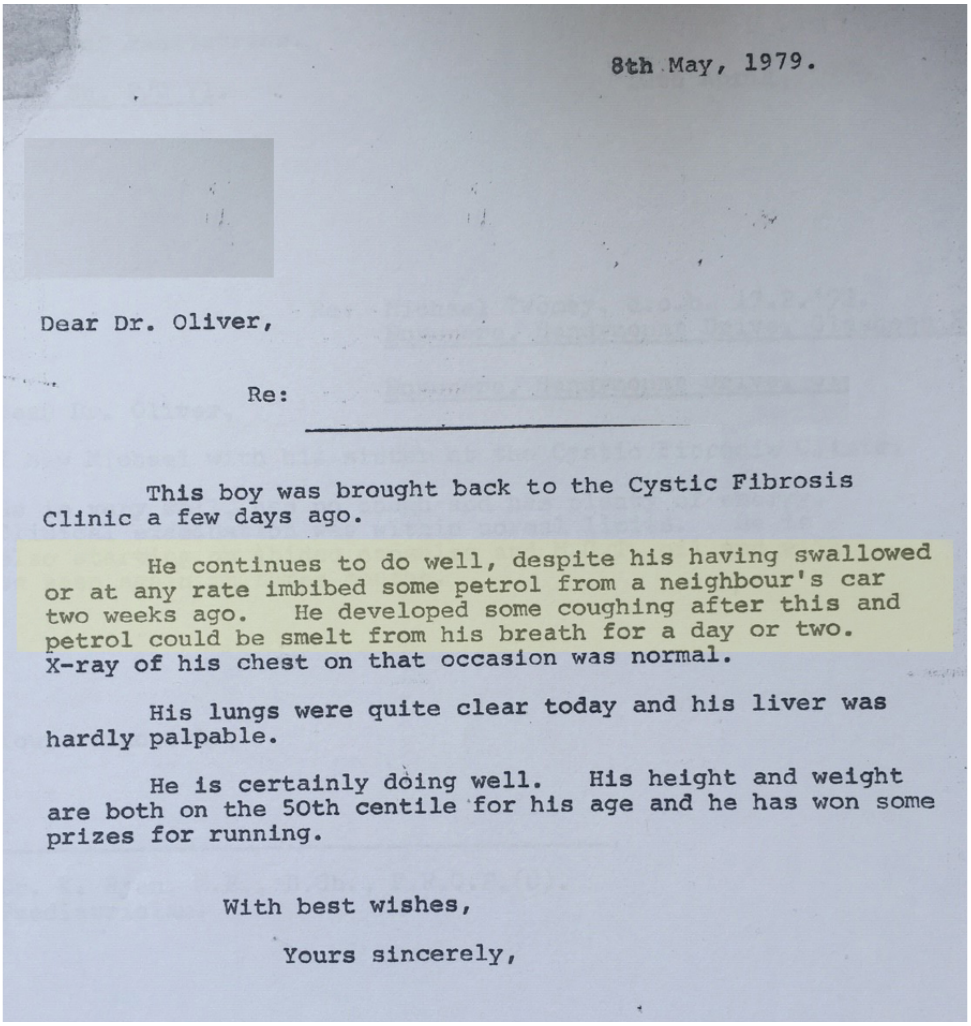
My research activity occurred in 3 iterative ADR cycles. Each iteration comprised rigorous naturalistic evaluation, involving subjective ex post interviews between myself and the evaluation team (7 adult patients with CF and 11 caregivers of children with CF), regarding their use of the artifact within their real-life medical appointments. In these evaluation interviews, I used qualitative metrics evaluating completeness, usability, robustness, and impact (Multimedia Appendix 1) [43], which helped shape our sensemaking process.

In tandem with this, I also sought expert opinions from clinicians on their appraisal of the checklist design and its subsequent use by patients with CF and their caregivers. These activities were followed up with learnings, reflections, and frequently additional consultation with the literature, followed by conformity by the design team on the design enhancements to be executed in any subsequent iteration, incorporating the intervention strategy of the same.

I now visit the next section of the paper, starting with the model of reflection I opted for, and the reflection (through 4 vignettes) on *problem formulation* within my ADR journey and its significance to efficacious outcomes.

XSL•FO
RenderX

Figure 2. The petrol episode.



8th May, 1979.

Dear Dr. Oliver,

Re:

This boy was brought back to the Cystic Fibrosis Clinic a few days ago.

He continues to do well, despite his having swallowed or at any rate imbibed some petrol from a neighbour's car two weeks ago. He developed some coughing after this and petrol could be smelt from his breath for a day or two. X-ray of his chest on that occasion was normal.

His lungs were quite clear today and his liver was hardly palpable.

He is certainly doing well. His height and weight are both on the 50th centile for his age and he has won some prizes for running.

With best wishes,

Yours sincerely,

Reflection on My Action Design Research Journey

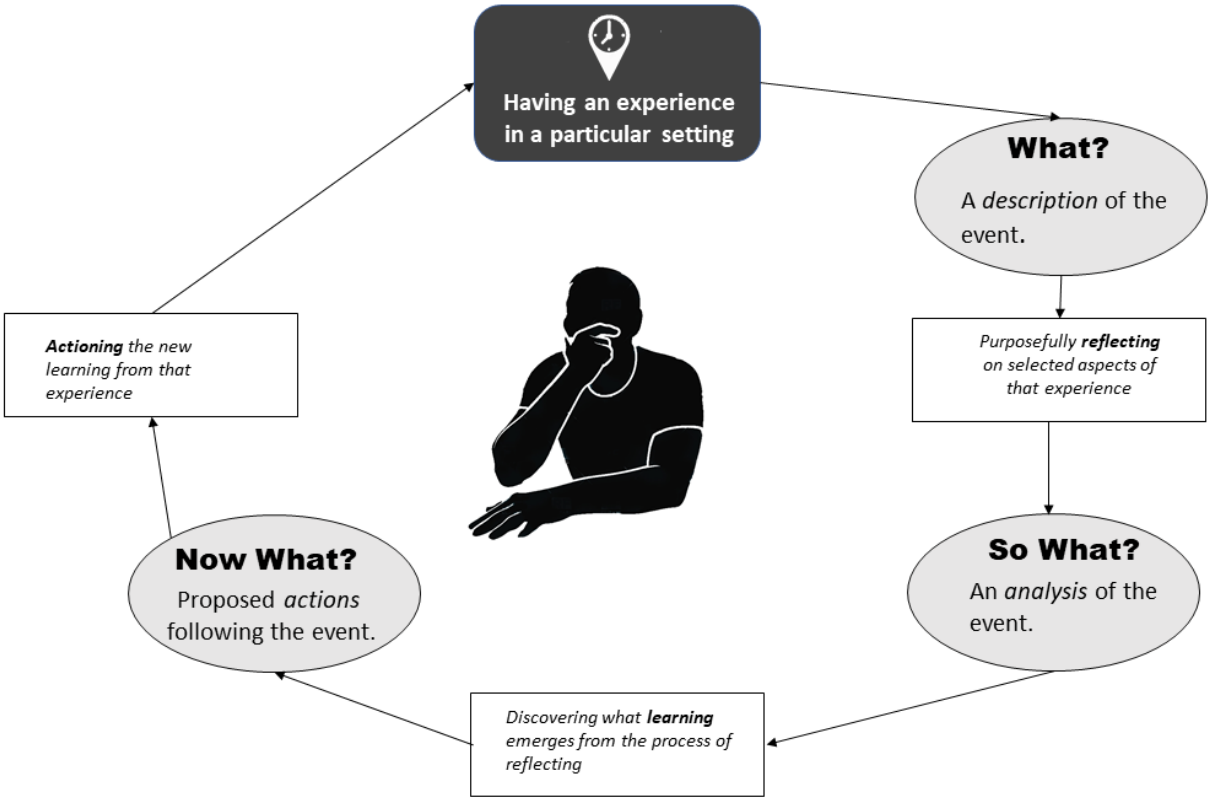
Model of Reflection

In 1988, Gibbs [44] argued that it was "not sufficient simply to have an experience in order to learn; without reflecting upon this experience it may quickly be forgotten, or its learning potential lost. It is from the feelings and thoughts emerging from this reflection that generalisations or concepts can be generated. And it is generalisations that allow new situations to be tackled effectively." After all, "we learn from reflection on experience. Reliving of an experience leads to making connections between information and feelings produced by the experience" [45].

Many of the seminal works on reflections or reflecting served as the initial stepping stone of my reflection in this paper. Although many models exist as possible viewpoints from which one might reflect, I opted to use Driscoll's What? Model of reflection [46] (Figure 3), as I felt it resonated with me the most as an instrument to steer my reflection through the often-murky waters of my inner self.

XSL•FO
RenderX

Figure 3. The What? Model of reflection - adapted from Driscoll.



Furthermore, as a reflection can be very personal and tacit in nature, it can be quite a challenging exercise or experience in and of itself. For example, it can take considerable time and may be painful and may even create a crisis of confidence [47]. That being said, it "offers distinctively grounded opportunities to pursue the connections between biography and social structure" [48]. Moreover, "reflection allows us to draw conclusions about our past experiences and develop new insights that we can apply to our future activities" [49].

Therefore, I felt it important to select the correct lens for me, one that would facilitate the recapture of my real-world experience rather than curtail or hinder the narration of my ADR journey as a patient and researcher. I reflect (using a series of 4 vignettes and Driscoll's model) on an aspect of my ADR journey that I feel quite passionate about, that of the *problem formulation* stage of ADR, and of course its value to the fruitful outcomes of my study. I have also added a lesson learned in each vignette; this is merely to reinforce the key message I wish you, the reader, to take away from my meanderings.

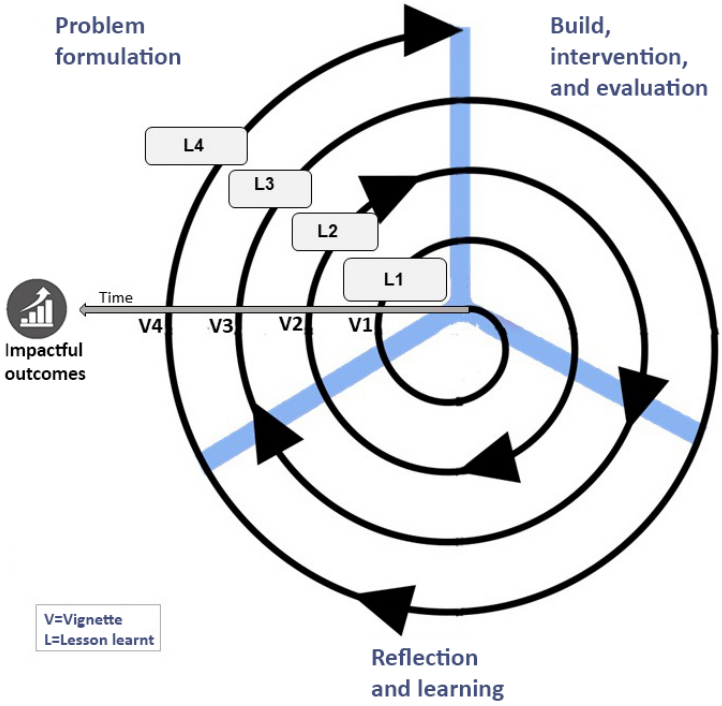
Although my ADR journey is still ongoing, the paper-based checklist aspect of the project, the majority of which took place over a 10-month period (from September 2016 to April 2017), culminating in the creation of a checklist booklet in January 2019, is for now complete. It is this period that I wish to reflect upon, a *reflection on action* or a reflection through review as described by Schon [50], the process of review to inform learning [50].

However, it is also important at this juncture to accentuate how and where these vignettes and learnings arose in the context of the overall ADR project. To this end, I include a simple diagram (Figure 4) that depicts the vignettes and learning in the context of the emergent, cyclical nature of the ADR project and its eventual outcomes. Although this reflection focuses on the *problem formulation* stage of ADR, it is critical to never lose sight of the fact that this stage moved in tandem with the other key stages of ADR.

These stages of ADR are akin to dancers performing the Tango, moving together in a closed embrace, individually and yet as one, influencing each other, each receptive to the other's movements, shaping and being shaped by each other, all to a combination of an on-and-off-beat rhythm. It is in this spirit that I wish the reader to embark with me on my ADR journey. Although I try to minimize the use of any references, where I have used them, it is merely to enforce or enrich the musing of my ADR journey. And so, we move to my first vignette.

XSL•FO
RenderX

Figure 4. Visualisation of reflection within the context of the ADR project, where V=vignette and L=lesson learnt.



Vignette 1: Tragic Thursday—September 2016

What Happened?

It was September 8, 2016. I was giving a presentation to my class on my efforts to date. I stood there like a proud peacock, chest out, boldly claiming the problem statement to be as follows:

> *There is currently no Patient Electronic Medical Records (PEMR) System that caters for the needs of CF patients or caregivers.*

Therefore, patients with CF and their caregivers needed an app to manage their medical data. What is more, my design team and I were going to deliver it. We had created a number of wire-frames for each screen of the app, and here I was presenting them, "at long last CF patients and caregivers would no longer struggle in their medical appointments when asked by physicians about their medical histories" I said. Moreover, "when they leave the medical appointment, they will not struggle to remember the information imparted to them by the physician, it will sit on their phones and tablets and eventually in the cloud."

And then it came, "I'm not convinced, I don't think they will use it, I don't think you have a handle on this yet" my innovation lecturer said. It was like he had pulled the rug from under me; I stood there shocked! The conversation that ensued between me and the lecturer (I am embarrassed to say) got somewhat heated, and what made the whole situation worse was it all happened in front of the entire class!

I returned home; the day had not gone as planned; I was very frustrated and quite upset; I had been publicly challenged regarding my solution and indeed the problem I thought I was trying to solve. I was now sitting in the kitchen with my head in my hands; I was not in a good place. Why was I having such difficulties with the *problem formulation* stage of ADR? My wife and child came into the kitchen and knew by my demeanor that all was not well. "What happened, did the presentation not go well?" she asked. "No, it was awful" I replied. "Let's talk later" she said with a quick glance at our little boy.

Later on, that evening, I proceeded to give her all the gory details of what had happened. However, I did not realize my little boy (of 7.5 years) was listening as he had left the kitchen and had gone upstairs to his bedroom before the regurgitation of my day. However, as you know, children occasionally have a happy knack of overhearing or eavesdropping on discussions that they should not be privy to. This, despite our best efforts to protect them from the trivialities of adult conversations, and so, after ascertaining the gist of what had happened to me, he enters the kitchen with the swagger of a man who was going to expound some wonderful insight, and says "Daddy wouldn't you think that after all the trips to the doctor that you have had, that you would understand what happens at a medical appointment?" I was left dumbstruck. I now refer back to that particular day as *tragic Thursday*, and yet on reflection, there was nothing tragic about it, in fact, the complete opposite is true; from that day onward, my eyes were going to be opened! I was going to realize the importance of this day as a turning point in my research.

So What?

I refer to the earlier event merely to illustrate my toil within the *problem formulation* stage of ADR. Although Sein et al [5] outline this as the first stage in ADR, they, unfortunately, do not refer to *how* one might or should go about *problem formulation*. I, probably like many others before me, ran into the solution space, convinced that I understood *the problem* that needed to be solved. I really thought after the presentation *on*

XSL•FO
RenderX

tragic Thursday that I was the issue and that I may not be the patient or researcher for the job in hand. So, was it just me?

Well, yes and no. Let me clarify; as the earlier background section on *problem formulation* exemplifies, there is a lot more to *problem formulation* than one might initially think. Unsurprisingly, as we saw, we humans are not always logical in how we approach problems. We hold many biases and repeatedly make suppositions becoming fixated on unjustifiable assumptions [30]. These faults within humans often run counter to the successes that we strive to achieve and to the challenges that we strive to overcome.

Furthermore, solution fixation often results, leading to adverse consequences, precluding or hindering in-depth questioning or interrogation of problems, and prematurely freezing a problem space before it can fully form. I think the most unfortunate negative effect of poor *problem formulation* is on the divergent exploration of the creative possibility in design. I was an exemplar of this behavior, as a patient living with CF for 48 years, with countless visits to physicians, I thought I knew the problem *inside out*. As far as I was concerned, I had conducted an initial survey of 305 patients with respiratory illnesses and their parents back in 2015, and 77.9% (280/359) said they had difficulty in remembering their medical history. Furthermore, over 95% of them said they would use a secure app if one existed. I had taken offense to being challenged. I remember thinking, *what the hell does he know?* My ego had thwarted my ability to hear and appreciate the very sound advice that was actually being imparted to me.

Thankfully, I am not a stubborn fellow, and sense prevailed. I have since become a lot humbler and more open to criticism. I suppose, as a patient who has faced the question of my mortality at a very young age, I have become somewhat resilient, accustomed to picking myself up, dusting myself down (this often involves giving myself a good telling off, including the words—"*stop feeling sorry for yourself, remember why and who you are doing this for*") and getting on with it. After all, someone had to resolve the issue, and if not me, then who?

Moreover, had *tragic Thursday* not happened, I would be sitting here on this Saturday afternoon in February 2020 (like many others before me) with a failed app. I most definitely would not be boxing up a solution for departure to the Royal London Children's Hospital, London. A solution that, in the coming weeks ahead, will grace the laps of caregivers all over southeast England, within their real-world medical appointments.

Now What?

As we established earlier, I was not alone in my thoughts and tribulations regarding *problem formulation*. Mullarkey and Hevner [10] and others had also reported complications with this stage of ADR. My challenge then was to be mindful of the real need for an in-depth implicit problem analysis and to understand and define the information problem that patients with CF and their caregivers face during their medical appointment in tandem with a solution/s to ameliorate the same. My next design workshop was calling.

My first lesson learnt (L1) was as follows: If we wish to achieve successful outcomes, *problem formulation* requires a conscientious focus on problem comprehension, avoiding *solution fixation* and other assumptions.

Vignette 2: So, What Is the Problem Again? —September 2016

What Happened?

I was sitting in the car in mid-September 2016. I was on my way to a workshop, looking forward to working with my design team, I was excited but also extremely nervous, it was only a week ago that my ship (I call the patient innovator) had taken a flurry of shots across the bow and nearly sunk, joining all the other vessels who had failed to survive on the wicked high seas in the world of innovation.

Many thoughts had been incubating in my head since *tragic Thursday*. My metacognitive processes were working furiously; my thoughts were a mix of emotions and ideas; to anyone brave enough to look inside my head, it would look *really messy*, like my room when I was a kid. I replayed the many medical appointments that I had attended in my mind, in tandem with all the literature that I had read in the area since autumn 2014. Although I had identified a problem or anomaly in the medical appointment, the issue that I now endeavored to solve was to understand and define the problem that patients with CF and their caregivers face during their medical appointment and delineate it in a way that made sense to me and others. Suddenly, a memory popped into my mind, vis-à-vis the time I endeavored to explain the issue to my darling wife for ≥10 min, after which she turned around and said, *so*, "what's the problem again?" Moreover, I needed to be able to classify and represent the problem in a way/s that would assist my design team and I to see how we might best deal with it.

I was struggling; my experiences as a patient with CF alone were not sufficient to articulate and solve the problem. I felt as if my mind was constantly being polluted by irrelevant details and assumptions. I could feel my heart beating faster; questions flooded my mind; how was I going to structure or represent the problem? How broad was the problem? What were the constraints? What knowledge was needed to understand and solve it, and what gaps existed in my or our current knowledge base? What external and social factors would come into play? What strategy would we adapt? What did success look like? It was going to be a thought-provoking, challenging workshop!

So What?

Let us be honest; we all face problems of one kind or another every day of our lives. A 1-year-old may face the problem of how to stand unaided, or how to escape from their cot, the Alcatraz of their world. On the other hand, teenagers face the challenge of living in the evolving world of social media, acceptance, bullying, and so on. Problems come in all forms, some are simple, some are quite complex in nature, and others have an undeniably *wicked* composition.

Although there is a myriad of difficulties within the medical appointment, I needed clarity regarding the specific enigma (or part thereof) that I was going to focus on. I needed to avoid or at least be aware of the symptoms that were polluting and confusing the situation I was trying to remedy. Those related

XSL•FO
RenderX

to other problems and yet overlapped with mine; otherwise, my ability to make sense of the issue with my design team was going to be a long, arduous process, one which would most likely end in failure. My design team and I needed to find an appropriate representation of the problem we had identified within the medical appointment, one that would give us insight/s to an appropriate solution pathway. Although my experience was beneficial and useful, it was only one CF patients' voyage and nothing more; there were many more patients with CF and their caregivers that also had their story to tell, stories that would enrich our comprehension of the problem space. However, how should I go about gathering such insights?

The problem I found with the ADR methodology is twofold; first, it appears to hold a rather technocentric view of innovation; by this, I mean it does not seem to take into account the often-idiosyncratic nature of humans. This is evident in the lack of guidance on *how* we might or should come to truly understand the people behind a problem. For example, Who has the problem? Why is it a problem? What do they think? What really matters to them? What do they feel as they toil within a problem space? and so on. The ability to garner such fundamental human insights is crucial to disentangling, understanding, and defining a problem. Second, not enough emphasis seems to be placed on the *problem formulation* stage of ADR. It is almost portrayed as if *problem formulation* is a straightforward process, when in reality, the opposite is often true; it is frequently wicked and ill-defined.

Now What?

Arlin [51] argues that for a problem to be real, there needs to be a *felt need* to eradicate any impediments to an objective. Considering these words further, they conjure or evoke thoughts of sentiment, of emotions, the very essence of what makes us human. Therefore, to pursue ADR within the context of the social environment of the medical appointment, I also needed a deeply human-centered mindset, an approach that was profoundly human in and of itself. Unfortunately, I found ADR wanting in this regard.

In contrast, design thinking focuses on a user's experiences and the emotions that are encapsulated in such events. Design thinking is a human-centered design methodology that "relies on our ability to be intuitive, to recognize patterns, to construct ideas that have emotional meaning as well as functionality, to express ourselves in media other than words or symbols" [52]. As one might imagine, some of the core principles of design thinking are empathy with users and discipline of prototyping.

Emotions are an integral part of what makes us human. Therefore, my design team and I began a series of design thinking workshops, beginning with a number of design thinking tools, from personas (Multimedia Appendix 2) to empathy maps (Multimedia Appendix 3) and journey maps (Multimedia Appendices 4 and 5). This meant spending a great deal of time with fellow patients with CF and their caregivers discussing the many facets of living with CF, their experiences (building on mine) within the medical appointment, and capturing their reality of being a patient with CF or having a child with CF, which are unique perspectives that were both profound and deeply insightful and often times quite moving. Empathy became the key to helping define the problem, but why empathy?

Many different interpretations of empathy exist from sharing the feelings of others and comprehending the emotions of others [53-56] to feeling what another experiences and the ability to appreciate the views of others [57]. Although we shared a medical condition, our life journeys are our own, unique, molding us, and shaping us into individuals. I came to understand the experiences of other patients with CF and their caregivers, assisting me to feel what they felt, I came to comprehend the stress they experienced living with CF, something that they (and I) have learned to manage daily. I heard of their experiences of holding a conversation with physicians while being short of breath, described by an adult patient with CF:

> *I was gasping for air, and trying to remember stuff for the doctor.*

A young mother of a child with CF explained to me what it was like trying to list the medications that her child was on at a particular medical appointment:

> *My 5-year-old child was really sick with a chest infection, she was crying due to the pain in her lungs, it was impossible to concentrate, it felt like I had 500 things going through my head, I was so stressed. I remember thinking, what if I make a mistake? What if I leave something out? I felt so guilty and helpless at that moment.*

Although insightful, I confess that, at times, I found this aspect of the project very difficult from a personal perspective. However, I learned to manage my own feeling quite quickly, forcing myself to compartmentalize my thoughts and feelings when and as required. I knew that this was critical if I wished to succeed and avoid floundering on the rocks in the sea of my own internal thoughts and emotions.

Nevertheless, it is important to note that it was through this appreciation of the problem through the senses and experiences reported by other patients and caregivers that we (my design team and I) would come to realize that a technocentric approach to the problem was not an appropriate starting point. It was only by really listening to what the patients and caregivers were telling us that we came to grasp what we were really dealing with. They were not, in fact, talking about apps or technology at all; they were complaining about their needs regarding their information and their frustrations regarding access to their medical history when and as required. A patient explained:

> *If I'm away traveling, I need to have my medical information at hand in case I get sick. I can't walk into a doctor who knows nothing about me or my condition. This happened to me before and I was put on the wrong treatment. It was very upsetting; I could have died.*

My design team and I eventually settled on the following new problem statement: "CF patients and caregivers are not having their information needs adequately addressed within their medical appointments."

XSL•FO
RenderX

Interestingly, renowned cardiologist Eric Topol [58] argued that "patients exist in a world of insufficient data, insufficient time, insufficient context.". In line with the principles of design thinking, we decided to engage in pretotyping (paper-based prototyping) in the form of a checklist to aid in our comprehension of initial interest and actual usage by users of our solution. The pretotype was designed for patients with CF and their caregivers to fill out before and during the physician's appointment. Pretotyping (conceived by Alberto Savoia [59]) also made sense as it enabled the smallest investment of time and money possible, while still facilitating the capture of distinctive insights from users of the checklist within the context of the medical appointment. The checklist pretotype was also prudent as a precursor to any digital solution that we may eventually embark on. It also steered me clear of *falling in love* with any early solutions. I wished to avoid another tragic Thursday. Checklist iteration 1 was designed and released at the end of September 2016 (Multimedia Appendix 6) [43] with evaluations conducted at the end of October and through November 2016 (Multimedia Appendix 1) [43]. Interestingly, even at this early stage, the checklist demonstrated positive outcomes, with 81% of participants reporting better memory recall as a result of using the checklist.

My second lesson learnt (L2) was as follows: *Problem formulation* requires in-depth human-centric exploration—scrutinizing a problem thoroughly through the senses of those who experience it, understanding how it affects them, and culminating in the articulation of an accurate problem definition and more positive solution outcomes.

Vignette 3: Teaching an Old Dog New Tricks—May 2017

What Happened?

It was a beautiful morning in May 2017 and I was out walking my dog, *Suzy*, down in the local forest not far from home. There was a very light breeze, the trees gently swayed to the dawn chorus of bird acoustics, a melody of song and calls. I had recently completed iteration 3 [43] (Figures 5 and 6) of the checklist, and thankfully, the evaluations were very positive (Multimedia Appendix 1) [43], with the artifact having major impacts within the CF community. Here, patients and caregivers reported reduced stress and increased empowerment as a result of having their medical information with them during their medical appointments and, of course, afterward. The checklist really appeared to work, but why was it working so well?

What had we done that seemed to be tipping the results in our favor? Our enhanced understanding of the problem definitely seemed to be manifesting itself in the delivery of an improved solution? However, I had this innate feeling that I was missing something, that we still had not achieved an absolute sense of why the checklist was so impactful? I was bothered.

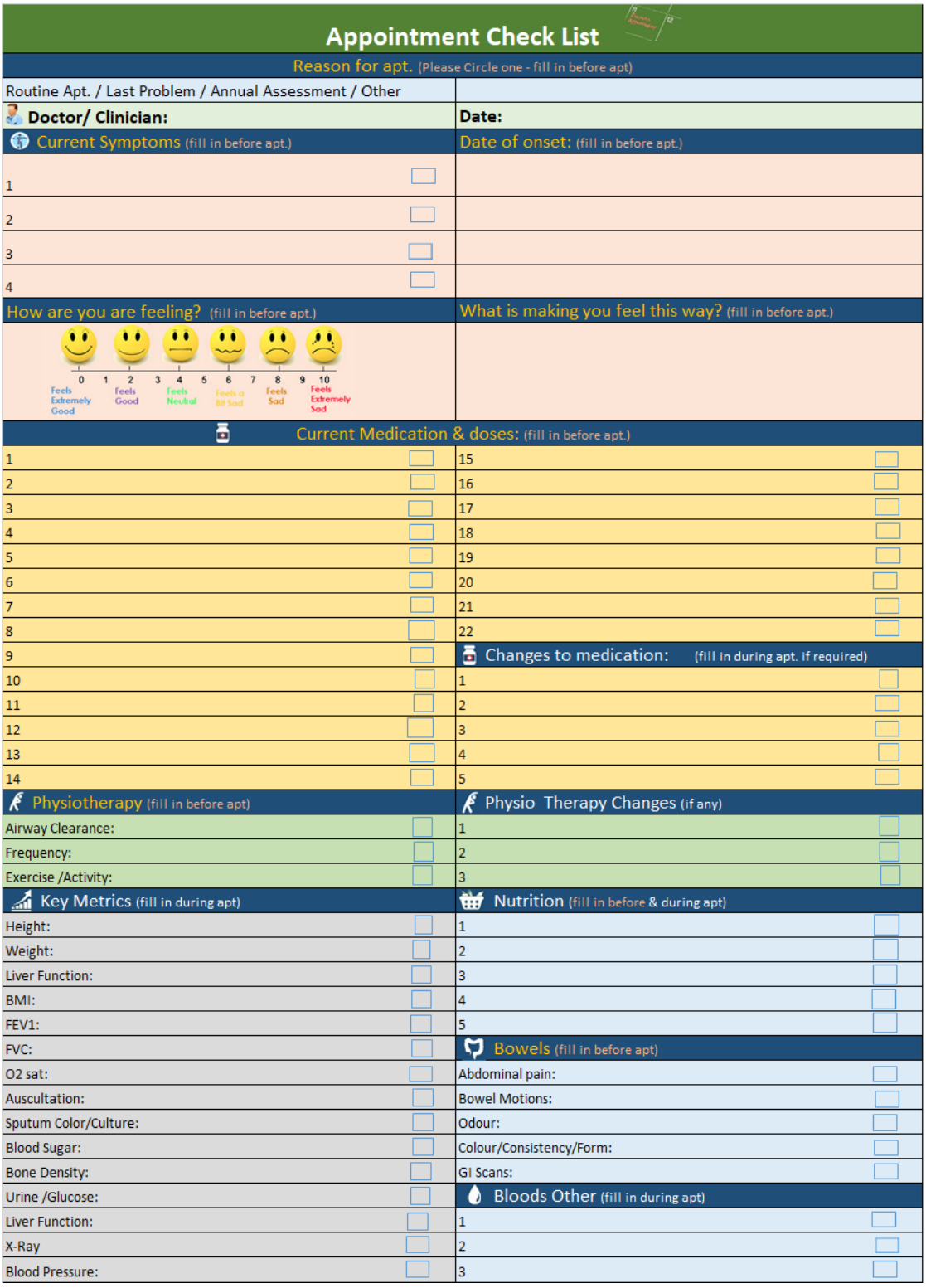
I proceeded down the forest path toward this small brook in which my dog was playing. She was play acting with a stick, when much to her dismay she dropped it, and it disappeared into some sort of small drain. I watched her with as much a sense of curiosity as amusement, as she endeavored to retrieve her prized possession. She approached the drain from one angle, then another, and then back again. This all went on for a number of minutes, it was clear from her expression that she was engaging the very limits of her cognitive abilities, as she tried to make sense of the predicament. Eventually, she managed to work it out, after several failed attempts doing it one way, she suddenly approached the problem differently, in a simpler fashion, she had her stick again!

My mind wandered back to my own thoughts, was there other ways that I should be tackling or viewing the information challenges that patients with CF and their caregivers were experiencing as well? Surely there were additional ways from which I could view the problem and solution, possibilities yet to be considered, ways that may come to enlighten me as to why the checklist was begetting such light into the often dark setting of the medical appointment with regard to memory recall. I was convinced that there was more to learn and more to understand, and yet, every time I tried to think about it, I found myself back where I started. It was like a neural impasse, as if the mental pathways in my mind were predetermined, immovable objects, defeating my abilities to explore new possibilities and new ways of thinking. I grew frustrated; my mind was tiring. I kicked a stone into the brook, "if only we had a more advanced memory, akin to some sort of futuristic form of Artificial Intelligence then there wouldn't be this stupid problem" I thought. And then just like my dog a few moments earlier, a new thought entered my mind, "If we had perfect memories, we would not need a solution."

The problem I thought was not really *information needs* as much as the limitations of human memory, arguably defective, often resulting in an inability to remember, a failure to recall memories on demand, and a malfunction of our biological information retrieval system. Inadequate *memory recall* was the real problem, it had been there under my nose all the time, and yet I failed to see it, until now, months later.

XSL•FO
RenderX

Figure 5. Version 3, part 1, of the checklist.



Appointment Check List

Reason for apt. (Please Circle one - fill in before apt)

Routine Apt. / Last Problem / Annual Assessment / Other

Doctor/ Clinician: | **Date:**

Current Symptoms (fill in before apt.) | **Date of onset:** (fill in before apt.)

1 | 2 | 3 | 4

How are you are feeling? (fill in before apt.) | **What is making you feel this way?** (fill in before apt.)

0 Feels Extremely Good; 1 2 Feels Good; 3 4 Feels Neutral; 5 6 Feels a Bit Sad; 7 8 Feels Sad; 9 10 Feels Extremely Sad

Current Medication & doses: (fill in before apt.)

1 2 3 4 5 6 7 8 9 10 11 12 13 14

15 16 17 18 19 20 21 22

Changes to medication: (fill in during apt. if required)

1 2 3 4 5

Physiotherapy (fill in before apt)

Airway Clearance:
Frequency:
Exercise /Activity:

Physio Therapy Changes (if any)

1 2 3

Key Metrics (fill in during apt)

Height:
Weight:
Liver Function:
BMI:
FEV1:
FVC:
O2 sat:
Auscultation:
Sputum Color/Culture:
Blood Sugar:
Bone Density:
Urine /Glucose:
Liver Function:
X-Ray
Blood Pressure:

Nutrition (fill in before & during apt)

1 2 3 4 5

Bowels (fill in before apt)

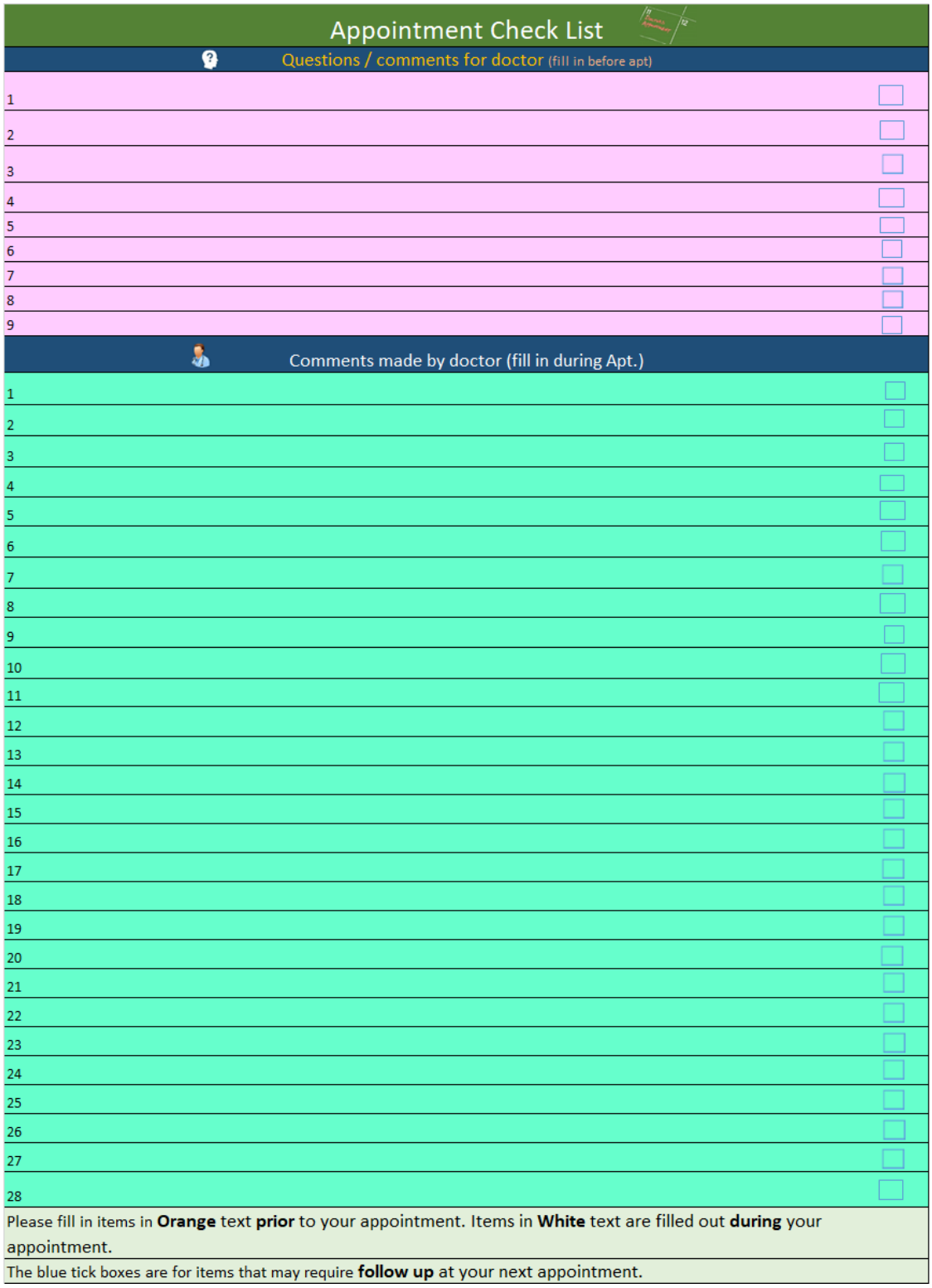
Abdominal pain:
Bowel Motions:
Odour:
Colour/Consistency/Form:
GI Scans:

Bloods Other (fill in during apt)

1 2 3

XSL•FO
RenderX

Figure 6. Version 3, part 2, of the checklist.



Appointment Check List

Questions / comments for doctor (fill in before apt)

1
2
3
4
5
6
7
8
9

Comments made by doctor (fill in during Apt.)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Please fill in items in **Orange** text **prior** to your appointment. Items in **White** text are filled out **during** your appointment.

The blue tick boxes are for items that may require **follow up** at your next appointment.

So What?

The abovementioned vignette reminds me of the American inventor and engineer Charles Kettering when he declared that "a problem well stated is a problem half-solved" [60]. So, what exactly did Kettering mean here? Do we take from his statement that once a problem is stated, we are halfway to a solution? Or does he infer something more arcane, that even with a well-stated problem, there is abundant knowledge yet to be unearthed, comprehended, and considered, regarding how and why solutions perform as they do within their intended environments? Having traveled through my ADR experience, I think he may well have intended for us to appreciate both in combination and individually.

In May 2017, I really realized and appreciated the richness of looking at a problem through multiple lenses. I was also amazed at how static my cognitive frameworks or schemas were. Moreover, I was amazed at how long I had stayed in these

XSL•FO
RenderX

cognitive states, unable to move, paralyzed if you will, this despite numerous evaluation interviews with patients and caregivers and workshops with my design team. Why had it taken so long for the older mental model to be replaced and augmented with a new one that would enhance my explanatory power? Was it the assumptions I was making while trying to make sense of the complex environment of the medical appointment? Had I become locked in on a particular mental representation of what I perceived was the *sweet spot* of the issue?

I mentioned earlier that "what we understand and how we understand a situation depends on the information we bring to a given situation, and the longer we think about the situation the more its cognitive representation changes. It may be assumed that cognitive elaboration activates more schemata" [42]. I was fascinated by this and yet cautious, what other points of view had I not considered? In spite of these contemplations, I decided that I would refrain from tormenting myself and spoiling this moment in the process.

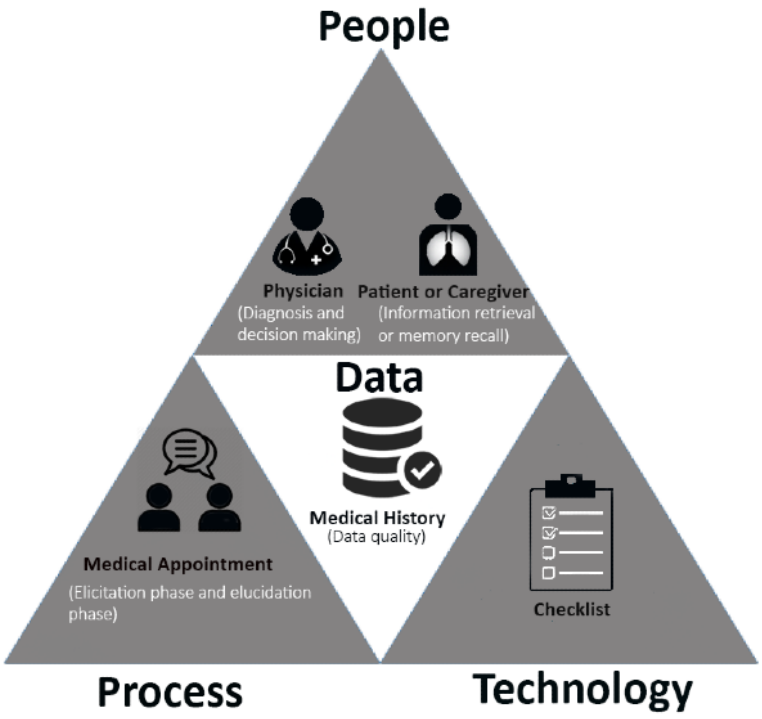
I had no doubt that as I continued on my voyage, I would discover new ways of thinking of framing my understanding. However, a sense of balance and perspective is required; one must avoid entering a state of *analysis paralysis*. As the esteemed English writer Samuel Johnson once said, "nothing will ever be attempted if all possible objections must first be overcome." I would take each enlightenment as it came, relishing the cognitive challenge that each schema would bring. More than one account of a complex system is achievable, where altered portrayals will break the system down in diverse forms, and changed descriptions may also have altered levels of intricacy. I was both relieved and excited that I had discovered a new viewpoint from which to perch my telescope of inquiry. Moreover, I felt an augmented confidence ignite within me, the cause of which was twofold; first, I now felt more assured regarding my grasp of the actual problem, and second, I sensed more confidence in my appreciation as to why the checklist was functioning so well for patients with CF and their caregivers.

Now What?

I relayed my thoughts to my design team, and we came up with a new problem statement: *The challenges of memory recall or information retrieval that CF patients and caregivers have or experience within the medical appointment are not well understood or solved.*

This made a lot more sense to all of us, and so taking a people, process, technology, and data view, we created a model (Figure 7) to depict our new representation or understanding of the problem, including the roles of memory recall or information retrieval within 2 key information stages of the medical appointment. I have also expanded on each of these people, process, technology, and data concepts as they pertain to the medical appointment in Table 1. The usefulness of Figure 7 and Table 1 is twofold: first, to depict my interpretation of the problem, and second, to highlight the advancement of my comprehension of the various interacting or moving components within the problem space.

Figure 7. People, process, technology, and data.



Moreover, I came to realize why the checklist was so effective, it was in effect acting as a tool aiding the long-term declarative memory of the patient and caregiver during their medical appointments. From another perspective, one could say that it achieved this by actually relying less on the patients or caregivers' own memory and more on the checklist within the appointment. Patients and caregivers now came prepared with the information required within the elicitation phase of the medical appointment already written down in front of them.

XSL•FO
RenderX

Moreover, they had a placeholder to capture key information within the appointment as well.

Table 1. People, process, technology, and data model concepts.

Concept	Reference
People	
Physician	
The purpose of the medical appointment is to "make the diagnosis."	[61]
The consequences of poor memory recall or information retrieval are: On the quality of information imparted to a physician, the ability to make an effective diagnosis and treatment decisions, and impacts on patient outcomes.	[62]
Clinician satisfaction	[63]
Patient or caregiver	
Research shows "that memory for medical history, like other forms of autobiographical memory, is likely to be flawed, incomplete and erroneous."	[62]
Process	
The elicitation phase	
Physician and patient or caregiver participate in a bidirectional conversation regarding the patient's medical history, current well-being, current medication, and so on.	[64]
Furnishes the physician with 60%-80% of the data required to make a diagnosis.	[65-67]
The elucidation phase	
Physicians communicate diagnoses, clinical options, and self-care plans, in tandem with overall advice regarding the management of a medical condition/s.	[68]
This phase directly impacts patient adherence and other self-managing activities, such as regime change.	[69]
Technology	
Checklist	
Defined as "a formal list used to identify, schedule, compare, or verify a group of elements or...used as a visual or oral aid that enables the user to overcome the limitations of human memory."	[70]
Data	
Medical history data includes the following: medical appointments, symptoms, illness episodes, encounters with other clinicians, medical therapies and medications.	[62]

Hence, they would not need to rely on memory when they left the medical encounter either. They had it all in the checklist and could refer back to it as required, even when traveling, if they happened to get ill. It probably comes as no surprise then why patients or caregivers were reporting such impacts on stress and empowerment. Additionally, I really came to fathom the potency of precise *problem formulation* vis-à-vis its impacts on actual outcomes. However, as one might expect, as I delved deeper into the area of memory recall during the medical appointment, I realized there was a lot more to this puzzle than I first envisaged or ever imagined on that beautiful summer morning in May 2017.

My third lesson learnt (L3) was as follows: In the *problem formulation* stage of ADR, we must challenge ourselves to look at a problem from different perspectives and from alternative disciplines. If we have not found or considered alternative viewpoints, we may fail to understand a problem well enough, affecting the most appropriate articulation of the problem definition, and the successful design of a solution or comprehension of why a solution functions as it does.

Vignette 4: Breaking It Down—May 2018

What Happened?

It was a warm, humid day on May 25, 2018; the sky was cloud-flecked; and the various bird songs bestowed a pleasing accompaniment to the day. I was on my way to a symposium in my university to give a presentation on my research. I was nervous; my stomach was making noises, clamoring's that I hoped were inaudible to the various scholars gathered in the room. I imagined that none would be too fond of hearing such clamoring ascend from my abdomen.

I hoped they would, however, be very interested in hearing how the checklist we had designed, built, and evaluated, functioned so well within the environment of the CF medical appointment. Moreover, following a 9-month rigorous systematic literature review of memory recall within medical appointments, I sought to impart where my comprehension of the problem had advanced to and why carrying out such an activity was fundamental to unlocking the additional knowledge I required in the *problem formulation* stage of our ADR project.

XSL•FO
RenderX

So What?

I had decided to conduct a meticulous literature review for 2 reasons. First, I wanted to understand why the checklist was functioning so well in the medical appointment. Second, I wanted to see if I was overlooking anything, for example, was defining the problem as memory recall or information retrieval of patients with CF and their caregivers within the medical appointment comprehensive or deep enough? I found that the answers to both questions were, in fact, deeply intertwined.

In the first instance, I came to understand that human declarative long-term memory was analogous to many complex systems consisting of components, in this case different memory types: episodic memory, autobiographical memory, and prospective memory (Multimedia Appendix 7 [71-77]). The components themselves are often simple (or can at least in this instance can be viewed as such) and interact with each other through various routes possible among components, mediated in distinct circumstances.

So why was the checklist functioning so well in medical appointments? Henry Ford is noted for saying, “Nothing is particularly hard if you divide it into small jobs.” Breaking down memory recall or information retrieval into its components, studying the physician and patient narratives (supplied by consenting patients and caregivers from real medical appointments), and assigning declarative memory components to each sentence or group of sentences allowed me to unearth a more profound comprehension of the complexities of dialogues within the medical appointment and the variety of long-term declarative memory components used therein. I also came to truly appreciate the pressures that memory recall places on the patient and caregiver, such as recalling a particular episode (episodic memory), time period/s (autobiographical memory), or remembering to report symptoms at an appointment (prospective memory) or a combination of declarative memory types. In addition, I found that the checklist design actually maps to *aid* the declarative long-term memory component drawn upon by the patient or caregiver during the medical appointment. This deeper, more comprehensive level of understanding of memory recall or information retrieval, breaking it down into its components, afforded me far deeper knowledge from which to view, examine, and indeed make sense of the problem I endeavored to solve, and of course as already put forward, why the checklist functioned so well for patients with CF and their caregivers within the complexities of the medical appointment.

Second, although a single checklist simplifies the capture of information at a particular medical appointment and, thus, aids the episodic memory of the patient or caregiver regarding that specific event, I came to understand that from an autobiographical memory perspective, the checklist was still somewhat inadequate. I wanted to support the autobiographical memory of patients and their caregivers in the best way possible, and, although theoretically, one could file away a single printed checklist in a folder after each medical appointment, I decided that this would not suffice. Patients and caregivers have enough going on in their lives without trying to find another workaround. Being a patient myself, I was sick to death of always having to settle for second rate solutions, solutions that I would later try and adapt to my own needs. On several occasions, I remember thinking, why is it that no one can get this stuff right? Are we (patients) that hard to understand, are our needs that difficult? Or is it that no one really gives genuine thought when designing products for us? I could not be a hypocrite; I had to improve the checklist, no workarounds!

Now What?

First, my design team and I came up with a new problem statement: *The challenges of memory recall information retrieval (and its components) that CF patients and caregivers have or experience within the medical appointment are not well understood or solved.*

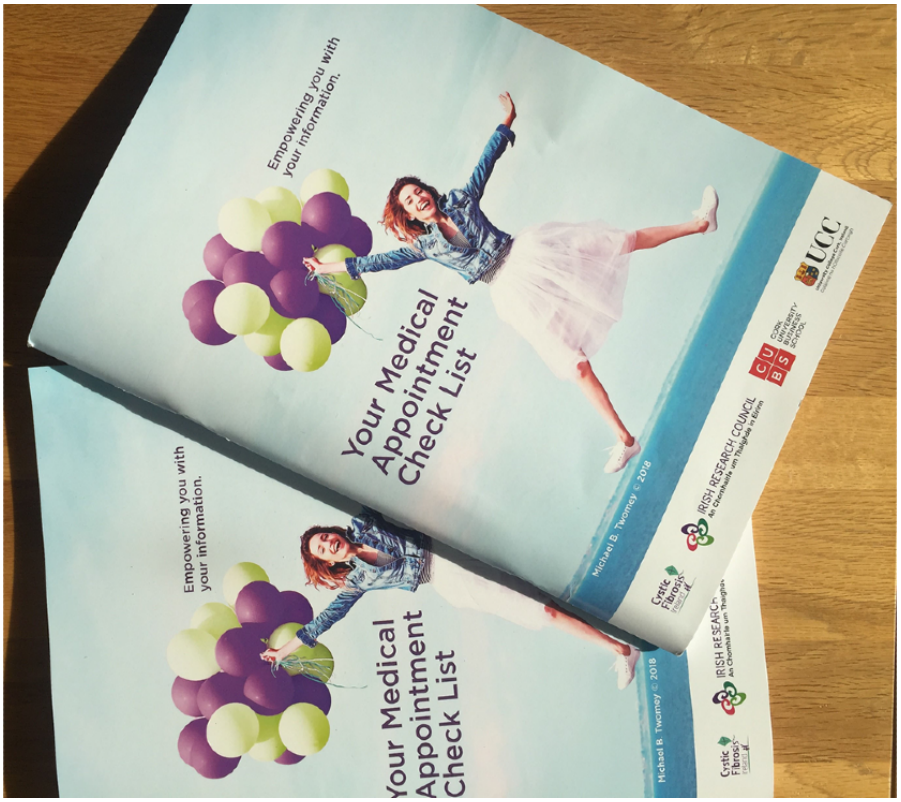
Second, after iteration 3, and my comprehension of long-term declarative memory components, we decided to create a professional, physically robust booklet (Figure 8), with a little help from a professional graphic designer (a friend of mine; Multimedia Appendix 8 shows the final checklist). A booklet containing 28 checklist items, unshackles the patient or caregiver from having to do any workarounds, including any printing. At once, a repository of medical discourse is created, where 28 medical appointments checklists are held together, not only facilitating the episodic and prospective memory of patients and their caregivers but also acting as an autobiographical memory of a specific time span. Searching for a previous appointment/s was now simple and straightforward.

Many may argue that this further exploration was a step too far, unnecessary, and indeed prohibitive for many researchers and organizations in terms of cost, and so on. Although I appreciate these sentiments, I would not agree. If you really care about the user’s experience (and you should) and you want to deliver quality impactful solutions, you must be willing to *go the whole nine yards*. In fact, I believe this is the only way to accomplish truly successful outcomes. To do anything else is to cheat yourself, your organization, and, most importantly, the user from what might have been.

Had I not gone deeper, the checklist as a booklet would never have come to pass. Granted, I would have a checklist (as a single page); however, I cannot say I would be confident that a patient or caregiver would continue to go to the trouble of printing a checklist before every appointment, and then file it away afterward. One must appreciate that patients with chronic diseases and their caregivers are busy trying to lead as normal a life as possible, and they often have very complex and time-consuming treatment schedules. Hence, asking what may seem a simple task can, unfortunately, often be the *straw that breaks the camel’s back* for a patient or caregiver and hence lead to unused or underused solutions. By having a deeper grasp of the problem, I was able to put this knowledge to good use, advancing to a more robust solution. This, I believe, is why so many are now requesting the checklist booklet.

XSL•FO
RenderX

Figure 8. The checklist booklet.



Additionally, when I explain the rationale behind the workings of the checklist booklet to those with the CF community, including clinicians, I can see that they are really intrigued, and if I may be so bold as to say, excited by the solution. This has of late manifested itself in a large amount of dialogue within the community, much of which I am only now becoming aware of, as it has translated into invitations to various clinical conferences, and international requests to see the checklist booklet, and of course boxes of checklist booklets being taped up on a Saturday afternoon. Checklist booklets that will soon find their way into the hands of patients with CF and their caregivers far and wide.

My final and fourth lesson learnt (L4) was as follows: *G*oing deeper in the *problem formulation* stage of ADR will help to explain observed phenomena, highlight shortcomings in a solution, and enrich problem definitions, resulting in a truly comprehensive understanding of a problem domain and the delivery of truly successful impactful solutions.

Concluding Remarks

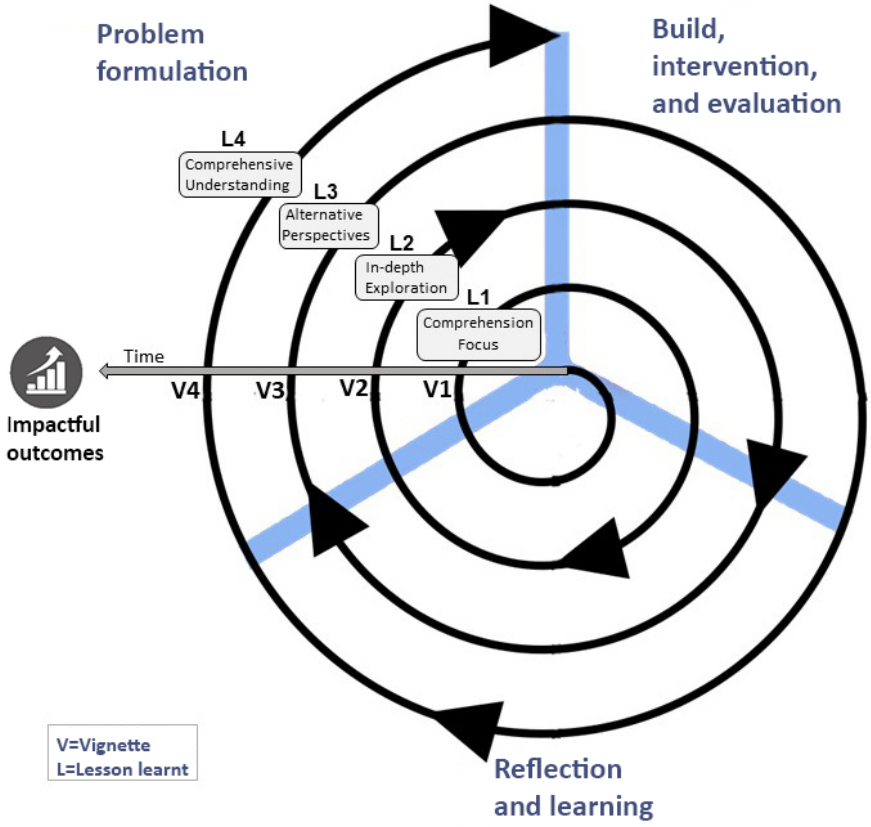
Unfortunately, that a small amount is appreciated vis-à-vis how problems are formulated in ADR seems as true today as it was six decades ago. In many of the ADR papers that I have read, there appears to be a very quick shift in focus to the subsequent stages of the methodology, with little mention or focus on the *problem formulation* stage. Moreover, *problem formulation* is seldom mentioned again in manuscripts. This is despite the iterative nature of ADR (Figure 1), where after *reflecting and learning*, the researcher/s refer back to the *problem formation* stage to ascertain whether or not a problem definition has changed or evolved. This is, of course, not to say that the stages that follow the *problem formulation* stage are not important, quite the contrary, they are also fundamental to an ADR project. Hence, I have included the same again in Figure 9, which, unlike Figure 4 earlier, now includes the lessons learned (as concepts) from each vignette extracted from my ADR journey. You will note in Figure 9 that our impactful outcomes are only attained at the culmination of our journey through iterations of ADR and *problem formulation* exploration and determination.

The question that I still contemplate is, why further regard is not given to this crucial stage of ADR? In my researcher story overlaid on an ADR story, I have bought to bear (through my series of vignettes) how difficult this stage of ADR actually is. I have also tried to portray what can go wrong without a conscientious focus on problem comprehension. Moreover, I have highlighted how beneficial time spent in this stage of ADR is, in terms of research impacts and results. Surely, I am not alone in my struggles as a researcher in *problem formulation*?

As I have already stated, but wish to emphasize once more, it is fundamental to empathize or understand the people behind a problem, what they experience, what they are feeling, and what and how they think, if you wish to deliver truly impactful solutions and sought-after outcomes. Indeed, Southard [78] was the first to articulate the significance of empathy in the physician and patient therapeutic relationship and its role in assisting diagnostic (*problem formulation* within the medical appointment) outcomes. To do otherwise is to deny our humanity, blocking the very comprehension we require to address the often-difficult problems we encounter, as we go about our lives on a daily basis, navigating the many complex systems within which we live.

XSL•FO
RenderX

Figure 9. Visualization of reflection within the context of the action design research project.



In vignette 3, I contended that we challenge ourselves to look at a problem from different perspectives, to be more open to other disciplines, and to be prepared to "repeatedly change our point of view, our way of looking at the problem. We have to shift our position again and again" [79]. This augmented comprehension not only facilitates a more appropriate articulation of a problem but should also liberate insights into why a solution functions the way it does within a particular environment.

However, as illustrated in L4, we must also be willing to probe and dissect these new viewpoints further, atomizing them down into components and concepts, facilitating deeper insights into phenomena, deepening our problem definitions, and thereby enabling a truly inclusive augmented comprehension of a problem. Additionally, such curiosity draws our attention to inadequacies in our solution/s, such as the checklist before it being in booklet form, where it really failed to address autobiographical memory. Through this knowledge, we further enhanced the checklist. Without delving deeper, this would not have happened.

The checklist produced evolved and was shaped, not only by the environment into which it is placed but also by my mind, which also underwent a type of metamorphosis, as empirical findings and knowledge waltzed together to the beat of my heart and the passions and conviction that expounded from within. Time and time again on this ADR project, I felt like I was on a journey of self-discovery, with many twists and turns, good days and bad days, days that taught me some valuable lessons, experiences that sometimes had a real sting to them, but will not be forgotten.

None of it was, of course, in vain; quite the contrary, I know I have made a real difference and continue to change the lives of many patients with CF and their caregivers during their medical appointments. I could not ask for a better outcome in my research endeavors. As pointed out earlier, I put this down to both grit and determination, especially as I have shown, within the *problem formulation* stage of my research. Therefore, fundamental to ADR success is the continuous revisiting of *problem formulation* after each iteration of an artefact, it is only by doing same that we can hope to gain a truly augmented understanding of a problem, and become more confident in designing and in our solution designs. I hope what I have discovered and aimed to impart here proves useful and insightful to those who brave the high seas of *problem formulation* in ADR, helping them to avoid some of the fatuous mistakes that I have made while on this chapter of my ADR voyage. Aiding them reach their intended research destination in one piece, confident that they too, have delivered impactful solutions through an augmentation of problem comprehension.

In tandem with the above viewpoints, I would advocate for the inclusion and portrayal of the actual realities of this stage (as I have endeavored to accomplish in this reflection) to be included and explored by researchers and practitioners. I feel that the insights garnered regarding same would not only bring a sense of realism and humanity to research (a component that I feel is often missing), they would also generate contributions to knowledge in and of themselves, the *how to* or *how I or we* navigated challenges encountered in research.

I miss my dear sister, but I choose to honor her memory by doing something that I know would bring a warm smile to her

XSL•FO
RenderX

face. I will be honest at this moment I have no idea where my researcher voyage will eventually take me, but it feels so right. Like a faint whispering in my ear that gets louder each day, like a fog lifting, giving a clearer aspect to the road ahead. My heart quickens as my quest becomes clearer.

Acknowledgments

This research was supported by the Irish Research Council and Cystic Fibrosis Ireland (project ID: EPSPG/2017/354). The University College Cork Social Research and Ethics Committee has reviewed and approved the submission for this study (application log 2017-090).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist evaluation summary.
[PNG File , 696 KB - jmir_v22i7e16916_app1.png]

Multimedia Appendix 2
Personas.
[PNG File , 414 KB - jmir_v22i7e16916_app2.png]

Multimedia Appendix 3
Empathy map.
[PNG File , 720 KB - jmir_v22i7e16916_app3.png]

Multimedia Appendix 4
Journey Map - Activity Cycle.
[PNG File , 188 KB - jmir_v22i7e16916_app4.png]

Multimedia Appendix 5
Journey maps—Day in the life.
[PNG File , 276 KB - jmir_v22i7e16916_app5.png]

Multimedia Appendix 6
Version 1 of the checklist.
[PNG File , 60 KB - jmir_v22i7e16916_app6.png]

Multimedia Appendix 7
Declarative memory types.
[PNG File , 95 KB - jmir_v22i7e16916_app7.png]

Multimedia Appendix 8
Booklet version of the checklist.
[PNG File , 2457 KB - jmir_v22i7e16916_app8.png]

References

1. Sternberg RJ. The Psychology of Problem Solving. 5th Ed. Cambridge, UK: Cambridge University Press; 2003.
2. Boud D, Keogh R, Walker D. Reflection: Turning Experience into Learning. London, UK: Kogan Page; 1985.
3. Korthagen F. A reflection on reflection. In: Korthagen FA, Kessels J, Koster B, Lagerwerf B, Wubbels T, editors. Linking Practice and Theory: The Pedagogy of Realistic Teacher Education. New Jersey, USA: Lawrence Erlbaum Associates; 2001.
4. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75. [doi: 10.2307/25148625]
5. Sein M, Henfridsson O, Purao S, Rossi M, Lindgren R. Action design research. MIS Q 2011;35(1):37. [doi: 10.2307/23043488]

XSL•FO
RenderX

6. Orlikowski WJ, Iacono CS. Research commentary: desperately seeking the 'IT' in IT research—a call to theorizing the it artifact. Inf Syst Res 2001 Jun;12(2):121-134. [doi: 10.1287/isre.12.2.121.9700]
7. Baskerville R, Wood-Harper AT. Diversity in information systems action research methods. Eur J Inf Syst 1998 May 29;7(2):90-107. [doi: 10.1038/sj.ejis.3000298]
8. Sein MK, Rossi M. Elaborating ADR while drifting away from its essence: a commentary on Mullarkey and Hevner. Eur J Inform Syst 2018 Oct 18;28(1):21-25. [doi: 10.1080/0960085x.2018.1527189]
9. Wieringa R. Relevance and Problem Choice in Design Science. In: Proceedings of the International Conference on Design Science Research in Information Systems. 2010 Presented at: DESRIST'10; June 4-10, 2010; St Gallen, Switzerland. [doi: 10.1007/978-3-642-13335-0_5]
10. Mullarkey MT, Hevner AR. An elaborated action design research process model. Eur J Inform Syst 2018 Apr 23;28(1):6-20. [doi: 10.1080/0960085x.2018.1451811]
11. Bretschneider U, Hartmann M, Leimeister JM. Fostering Ideation Among Patients: Exaptation of Web-based Ideation Platforms as Health 2.0 Tools for Virtual Patient Communities. In: International Conference on Information Systems. 2015 Presented at: ICIS'15; December 13-16, 2015; Texas, USA. [doi: 10.2139/ssrn.3159169]
12. Ågerfalk PJ. Getting pragmatic. Eur J Inform Syst 2017 Dec 19;19(3):251-256. [doi: 10.1057/ejis.2010.22]
13. Purao S, Henfridsson O, Rossi M, Sein MK. Ensemble artifacts: from viewing to designing in action design research, systems, signs & actions, an international journal on information technology. Int J Law Info Tech 2013;7:73-71 [FREE Full text]
14. Berthon PR, Pitt LF, Morris MH. The impact of individual and organizational factors on problem perception: theory and empirical evidence from the marketing-technical dyad. J Bus Res 1998 May;42(1):25-38. [doi: 10.1016/s0148-2963(97)00058-1]
15. Mumford E. Problems, knowledge, solutions: solving complex problems. J Strategic Inf Syst 1998 Dec;7(4):255-269. [doi: 10.1016/s0963-8687(99)00003-7]
16. Mark ML, Majchrzak A, Gasser L. A design theory for systems that support emergent knowledge processes. MIS Q 2002;26(3):179-212. [doi: 10.2307/4132330]
17. Hevner AR, March ST, Park J, Ram S. Design Science in Information Systems Research. In: Design Science in Information Systems Research 2004(28):75-105.
18. Volkema RJ. Problem formulation in planning and design. Manag Sci 1983 Jun;29(6):639-652. [doi: 10.1287/mnsc.29.6.639]
19. Ackoff RL. The development of operations research as a science. Oper Res 1956 Jun;4(3):265-295. [doi: 10.1287/opre.4.3.265]
20. Mintzberg H, Raisinghani D, Theoret A. The structure of 'unstructured' decision processes. Adm Sci Q 1976 Jun;21(2):246-275. [doi: 10.2307/2392045]
21. Mitroff II, Emshoff JR, Kilmann RH. Assumptional analysis: a methodology for strategic problem solving. Manag Sci 1979 Jun;25(6):583-593. [doi: 10.1287/mnsc.25.6.583]
22. Kaufman J, Sternberg R. The Cambridge Handbook of Creativity. Cambridge, UK: Cambridge University Press; 2019.
23. Smith GF. Defining managerial problems: a framework for prescriptive theorizing. Manag Sci 1989 Aug;35(8):963-981. [doi: 10.1287/mnsc.35.8.963]
24. Mitroff II, Featheringham TR. On systemic problem solving and the error of the third kind. Syst Res 1974 Nov;19(6):383-393. [doi: 10.1002/bs.3830190605]
25. Lyles MA, Mitroff II. Organizational problem formulation: an empirical study. Adm Sci Q 1980 Mar;25(1):102. [doi: 10.2307/2392229]
26. Kilmann RH, Mitroff II. Problem defining and the consulting/intervention process. Calif Manage Rev 1979;21(3):26-33. [Medline: 10244106]
27. Watson CE. The problems of problem solving. Bus Horiz 1976 Aug;19(4):88-94. [doi: 10.1016/0007-6813(76)90070-7]
28. Simon HA. Models of Discovery and Other Topics in the Methods of Science. New York, USA: Springer; 1977.
29. Simon HA. Models of Man: Social and Rational. Oxford, UK: Wiley; 1957.
30. Weisberg RW, Alba JW. An examination of the alleged role of 'fixation' in the solution of several 'insight' problems. J Exp Psychol Gen 1981;110(2):169-192. [doi: 10.1037/0096-3445.110.2.169]
31. Basadur M. Leading others to think innovatively together: creative leadership. Leadersh Q 2004 Feb;15(1):103-121. [doi: 10.1016/j.leaqua.2003.12.007]
32. Cross N. Designerly Ways of Knowing. New York, USA: Birkhauser: Basel; 2007.
33. Spradlin W. Are you solving the right problem? Harvard Bus. Rev. 2012;90(9):85-93.
34. Kuechler W, Vaishnavi V. So, talk to me: the effect of explicit goals on the comprehension of business process narratives. MIS Q 2006;30(4):961. [doi: 10.2307/25148761]
35. Roy M, Lerch J. Overcoming ineffective mental representations in base-rate problems. Inf Syst Res 1996 Jun;7(2):233-247. [doi: 10.1287/isre.7.2.233]
36. Dilla WN, Stone DN. Representations as decision aids: the asymmetric effects of words and numbers on auditors' inherent risk judgments. Decis Sci 1997 Jul;28(3):709-743. [doi: 10.1111/j.1540-5915.1997.tb01328.x]

XSL•FO
RenderX

37. Simon HA, Hayes JR. The understanding process: problem isomorphs. Cogn Psychol 1976 Apr;8(2):165-190. [doi: 10.1016/0010-0285(76)90022-0]
38. Brody R. Problem solving. In: Animal Mind — Human Mind: Report of the Dahlem Workshop on Animal Mind — Human Mind. New York, USA: Human Science Press; 1982.
39. Venable J, Pries-Heje J, Baskerville R. A Comprehensive Framework for Evaluation in Design Science Research. In: 7th International Conference on Design Science Research in Information Systems and Technology. 2012 Presented at: DESRIST 2012; May 6-18, 2012; Berlin, Germany. [doi: 10.1145/1555619.1555631]
40. Peffers K, Rothenberger M, Tuunanen T, Vaezi R. Design Science Research Evaluation. In: Proceedings of the International Conference on Design Science Research in Information Systems. 2012 Presented at: DESRIST'12; May 14-15, 2012; Las Vegas, NV, USA. [doi: 10.1007/978-3-642-29863-9_29]
41. Lesgold A. Problem solving. In: Sternberg RJ, Smith EE, editors. The Psychology of Human Thought: An Introduction. Cambridge, UK: Cambridge University Press; 1988.
42. Aaltonen M. The Third Lens: Metaphor and the Creation of Modern Cell Biology. London, UK: Routledge: Taylor & Francis; 2016.
43. Twomey M, Sammon D, Nagle T, O'Raghallaigh P. A Check List Designed to Improve Memory Recall Amongst CF Patients. In: Proceedings of the Twenty-Sixth European Conference on Information Systems. 2018 Presented at: ECIS'18; June 23-28, 2018; Portsmouth, UK p. 23-26 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1154&context=ecis2018_rp
44. Gibbs G. Learning by Doing: A Guide to Teaching and Learning Methods. Oxford, UK: Further Educational Unit, Oxford Polytechnic; 1988.
45. Dewey J. How we think: a restatement of the relation of reflective thinking to the educative process. Am J Psychol 1934 Jul;46(3):528. [doi: 10.2307/1415632]
46. Driscoll J, Teh B. The potential of reflective practice to develop individual orthopaedic nurse practitioners and their practice. J Orthop Nurs 2001 May;5(2):95-103. [doi: 10.1054/joon.2001.0150]
47. Maclean S. National Organisation for Practice Teaching. 2012. Critically Reflective Practice in Social Work URL: https://tinyurl.com/ycntb8kp [accessed 2020-05-19]
48. Anderson L. Analytic autoethnography. J Contemp Ethnogr 2016 Jul 26;35(4):373-395. [doi: 10.1177/0891241605280449]
49. Wade RC, Yarbrough DB. Portfolios: a tool for reflective thinking in teacher education? Teach Teach Educ 1996 Jan;12(1):63-79. [doi: 10.1016/0742-051x(95)00022-c]
50. Schon D. The Reflective Practitioner: How Professionals Think In Action. London, UK: Temple Smith; 1983.
51. Arlin PK. Wisdom: the art of problem finding. In: Sternberg R, editor. Wisdom: Its Nature, Origins, and Development. Cambridge, MA: Cambridge University Press; 1990.
52. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. London, UK: Harper Business; 2019.
53. Abram SV, Wisner KM, Fox JM, Barch DM, Wang L, Csernansky JG, et al. Fronto-temporal connectivity predicts cognitive empathy deficits and experiential negative symptoms in schizophrenia. Hum Brain Mapp 2017 Mar;38(3):1111-1124. [doi: 10.1002/hbm.23439] [Medline: 27774734]
54. du Vaure CB, Lemogne C, Bunge L, Catu-Pinault A, Hoertel N, Ghasarossian C, et al. Promoting empathy among medical students: a two-site randomized controlled study. J Psychosom Res 2017 Dec;103:102-107. [doi: 10.1016/j.jpsychores.2017.10.008] [Medline: 29167035]
55. Fan Y, Wonneberger C, Enzi B, de Greck M, Ulrich C, Tempelmann C, et al. The narcissistic self and its psychological and neural correlates: an exploratory fMRI study. Psychol Med 2011 Aug;41(8):1641-1650. [doi: 10.1017/S003329171000228X] [Medline: 21144117]
56. Groen Y, Wijers A, Tucha O, Althaus M. Are there sex differences in ERPs related to processing empathy-evoking pictures? Neuropsychologia 2013 Jan;51(1):142-155. [doi: 10.1016/j.neuropsychologia.2012.11.012] [Medline: 23174404]
57. Crijns H, Claeys A, Cauberghe V, Hudders L. Who says what during crises? A study about the interplay between gender similarity with the spokesperson and crisis response strategy. J Bus Res 2017 Oct;79:143-151. [doi: 10.1016/j.jbusres.2017.06.010]
58. Topal E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, USA: Ingram Publisher Services; 2019.
59. Savoia A. Pretotyping. 2011. Pretotype It: Make Sure You Are Building the Right It Before You Build It Right URL: https://www.pretotyping.org/uploads/1/4/0/9/14099067/pretotype_it_2nd_pretotype_edition-2.pdf [accessed 2019-02-17]
60. Maclean P. A problem well stated is a problem half solved. In: Cierpka M, editor. Family Assessment: Integrating Multiple Perspectives. Boston, UK: Hogrefe & Huber; 2005.
61. Lazare A, Putnam S, Lipkin M. Three functions of the medical interview. In: Lipkin M, Putnam SM, Lazare A, Carroll JG, Frankel RM, editors. The Medical Interview: Clinical Care, Education, and Research. New York, USA: Springer; 1995.
62. Cohen G, Java R. Memory for medical history: accuracy of recall. Appl Cognit Psychol 1995 Aug;9(4):273-288. [doi: 10.1002/acp.2350090402]

XSL•FO
RenderX

63. Schraa J, Dirks J. Improving patient recall and comprehension of the treatment regimen. J Asthma 1982;19(3):159-162. [doi: 10.3109/02770908209104752] [Medline: 7118821]
64. Sarkar U, Schillinger D, Bibbins-Domingo K, Nápoles A, Karliner L, Pérez-Stable EJ. Patient-physicians' information exchange in outpatient cardiac care: time for a heart to heart? Patient Educ Couns 2011 Nov;85(2):173-179 [FREE Full text] [doi: 10.1016/j.pec.2010.09.017] [Medline: 21035298]
65. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [doi: 10.1136/bmj.2.5969.486] [Medline: 1148666]
66. Sandler G. The importance of the history in the medical clinic and the cost of unnecessary tests. Am Heart J 1980 Dec;100(6 Pt 1):928-931. [doi: 10.1016/0002-8703(80)90076-9] [Medline: 7446394]
67. Kassirer JP. Teaching clinical medicine by iterative hypothesis testing. Let's preach what we practice. N Engl J Med 1983 Oct 13;309(15):921-923. [doi: 10.1056/NEJM198310133091511] [Medline: 6888486]
68. Martin L, Dimatteo M. The Oxford Handbook of Health Communication, Behavior Change, and Treatment Adherence. Oxford, UK: Oxford University Press; 2014.
69. McPherson ML, Smith SW, Powers A, Zuckerman IH. Association between diabetes patients' knowledge about medications and their blood glucose control. Res Social Adm Pharm 2008 Mar;4(1):37-45. [doi: 10.1016/j.sapharm.2007.01.002] [Medline: 18342821]
70. Federal Aviation Administration. 2016. Manuals, Procedures, and Checklists for 14 CFR Parts URL: http://fsims.faa.gov/PICDetail.aspx?docId=8900.1,Vol.3,Ch32,Sec12 [accessed 2020-05-11]
71. Radvansky G. Human Memory. Third Edition. United Kingdom: Taylor and Francis; 2017.
72. Habermas T, Bluck S. Getting a life: the emergence of the life story in adolescence. Psychol Bull 2000 Sep;126(5):748-769. [doi: 10.1037/0033-2909.126.5.748] [Medline: 10989622]
73. McAdams DP. The psychology of life stories. Rev Gen Psychol 2001 Jun;5(2):100-122. [doi: 10.1037/1089-2680.5.2.100]
74. Einstein GO, McDaniel MA. Normal aging and prospective memory. J Exp Psychol Learn Mem Cogn 1990;16(4):717-726. [doi: 10.1037/0278-7393.16.4.717]
75. Harris JE. Remembering to do things: A forgotten topic. In: Everyday Memory, Actions and Absent Mindedness. Cambridge: Academic Pr; 1984.
76. Harris JE, Wilkins AJ. Remembering to do things: a theoretical framework and an illustrative experiment. Hum Learn 1982;1:123-136 [FREE Full text]
77. Harris RJ. Memory observed: Remembering in natural contexts. Ulric Neisser (Ed.). San Francisco: W. H. Freeman, 1982. Pp. xiii + 433. Appl Psycholinguist 1983;4(2):184-185 [FREE Full text] [doi: 10.1017/S0142716400004471]
78. Southard EE. The empathic index in the diagnosis of mental diseases. J Abnorm Psychol 1918;13(4):199-214. [doi: 10.1037/h0072727]
79. Polya G, Conway JH. How to Solve It: A New Aspect of Mathematical Method. Princeton, NJ: Princeton University Press; 1957.

Abbreviations

ADR: action design research
CF: cystic fibrosis
DSR: design science research
IS: information systems
IT: information technology

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by N Pantelli, C Bond; comments to author 03.02.20; revised version received 29.03.20; accepted 09.04.20; published 14.07.20.

Please cite as:
Twomey MB, Sammon D, Nagle T
The Tango of Problem Formulation: A Patient's/Researcher's Reflection on an Action Design Research Journey
J Med Internet Res 2020;22(7):e16916
URL: https://www.jmir.org/2020/7/e16916
doi:10.2196/16916
PMID:32285802

©Michael B Twomey, David Sammon, Tadhg Nagle. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution

XSL•FO
RenderX

License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Codeveloping a Virtual Patient Simulation to Foster Nurses' Relational Skills Consistent With Motivational Interviewing: A Situation of Antiretroviral Therapy Nonadherence

Geneviève Rouleau[1,2], RN, Msc; Jérôme Pelletier[2,3], RN, MSc; José Côté[1,4,5], RN, PhD; Marie-Pierre Gagnon[2,6,7], PhD; Valérie Martel-Laferrière[4,8], MD, MSc; Rock Lévesque[8], RN, MSc; SimforHealth[9]; Guillaume Fontaine[5,10], RN, MSc

[1]Research Chair in Innovative Nursing Practices, Montréal, QC, Canada
[2]Faculty of Nursing, Université Laval, Quebec, QC, Canada
[3]Université du Québec à Rimouski, Rimouski, QC, Canada
[4]University of Montreal Hospital Research Centre, Montreal, QC, Canada
[5]Faculty of Nursing, Université de Montréal, Montreal, QC, Canada
[6]University Hospital Centre of Quebec, Laval University Research Centre, Quebec, QC, Canada
[7]Institute of Health and Social Services in Primary Care, Research Centre on Healthcare and Services in Primary Care, Quebec, QC, Canada
[8]University of Montreal Hospital Centre, Montreal, QC, Canada
[9]Please see acknowledgements for a list of collaborators
[10]Research Centre, Montreal Heart Institute, Montreal, QC, Canada

Corresponding Author:
Geneviève Rouleau, RN, Msc
Research Chair in Innovative Nursing Practices
850 rue St-Denis, Pavillon S
Door S01-110
Montréal, QC, H2X 0A9
Canada
Phone: 1 514 890 8000
Email: genevieve.rouleau.chum@ssss.gouv.qc.ca

Abstract

Background: Although helping people living with HIV manage their antiretroviral therapy is a core competency of HIV nursing care, no educational intervention has sought to strengthen this competency. Thus, we codeveloped a simulation of a virtual patient (VP) having difficulty adhering to treatment to foster the relational skills that nurses require in such situations.

Objective: This viewpoint paper aims to describe the codevelopment process and the content of VP simulation, as well as the challenges encountered and the strategies used to overcome them.

Methods: We use a collaborative and iterative approach to develop the simulation based on qualitative evidence, theoretical approaches (strengths-based nursing, motivational interviewing [MI], and adult learning theories), and expert recommendations. We carried out 2 main phases: (1) planning the simulation development and (2) designing the simulation content, sequence, and format. We created the script as if we were writing a choose-your-own-adventure book. All relational skills (behavior change counseling techniques derived from MI) were integrated into a nurse-patient dialogue. The logic of the simulation is as follows: if the nurse uses techniques consistent with MI (eg, open-ended questions, summarizing), a dialogue is opened up with the VP. If the nurse uses relational skills inconsistent with MI (eg, providing advice without asking for permission), the VP will react accordingly (eg, defensively). Learners have opportunities to assess and reflect on their interventions with the help of quizzes and feedback loops.

Results: Two main challenges are discussed. The most salient challenge was related to the second phase of the VP simulation development. The first was to start the project with divergent conceptions of how to approach the VP simulation—the simulation company's perspective of a procedural-type approach versus the clinical team's vision of a narrative approach. As a broad strategy, we came to a mutual understanding to develop a narrative-type VP simulation. It meshed with our conception of a nurse-patient relationship, the values of strengths-based nursing (a collaborative nurse-patient relationship), and the MI's counseling style. The

XSL•FO
RenderX

second challenge was the complexity in designing realistic relational skills in preprogrammed and simulated nurse-patient dialogue while preserving an immersive learning experience. As a broad strategy, we created a collaborative and work-in-progress writing template as a shared working tool.

Conclusions: Our experience may be helpful to anyone looking for practical cues and guidance in developing narrative VP simulations. As relational skills are used by all nurses—from novices to experts—and other health care practitioners, focusing on this clinical behavior is a good way to ensure the simulation's adaptability, sustainability, and efficiency.

(J Med Internet Res 2020;22(7):e18225) doi:10.2196/18225

KEYWORDS

motivational interviewing; HIV; nurses; education, continuing; virtual patient; simulation; nurse-patient relations; communication

Introduction

The Role of Simulation in Continuing Professional Development

Professional expectations and social accountability require nurses to pursue continuing professional development (CPD) to reinforce and maintain their competencies, so they can provide evidence-based care and ensure patient safety [1,2]. CPD is typically offered through interactive and/or didactic educational meetings such as conferences, workshops, seminars, lectures, and courses [3]. However, many of these approaches require time, money, and human resources, which are often limited in health care environments [4,5]. There is a need to develop accessible and innovative ways to strengthen nurses' learning and expertise while considering resource availability and workplace realities. The use of virtual patient (VP) simulation, a digital learning modality, is one way to address these challenges and is increasingly used in contemporary nursing education [6].

Virtual Patient Simulation and Education for Health Professionals

VP simulation is an interactive computer simulation that depicts real-world scenarios with the goals of training, education, and assessment [7] for health care professionals. A VP enables learning in a nonjudgmental, ethical, and safe environment [4,8-10] as learners acquire knowledge and develop skills by learning from their mistakes, without harming patients [11]. The use of VP simulation in health professional education is growing exponentially, considering the number of reviews published in the field [6,12-15]. Although the populations included in these reviews are mainly undergraduate and postgraduate students and health care providers in medicine, nursing, and other disciplines, we found only 1 integrative review focusing exclusively on web-based simulation in nursing programs [6]. Thus, it is relevant to explore the use of VP simulation as a means of supporting CPD for nurses.

On the basis of the needs identified by nurses providing HIV care in a qualitative study [16], we developed a VP simulation aimed at improving the nurses' relational skills consistent with motivational interviewing (MI) [17] in situations of antiretroviral therapy nonadherence. MI is a person-centered, collaborative communication style that seeks to elicit people's motivation and commitment to change. A systematic review conducted by Shingleton and Pafai [18], which included 41 studies, aimed to characterize the use of technology-delivered MI interventions and their efficacy at changing various health-related behaviors among several populations. One of the authors' recommendations was to provide details, for instance, in methodological papers, on how the relational (eg, empathy) and technical (eg, confidence rulers) components of MI are translated into a technology-delivered intervention.

Insufficient Guidance for Developing a Specific Type of Educational Intervention

Several approaches exist to develop interventions aimed at improving health [19] and supporting behavioral change by health care professionals [20]. Relative to a taxonomy of eight approaches to intervention development [19], we locate ours as a specific type of intervention and behavior [21] because of the following elements: its modality (VP simulation), the action to be performed (adoption of relational skills consistent with MI), the actors performing the action (nurses and, more broadly, all health care professionals), the target behavior (medication adherence), and the recipients of this action (patients who are living with HIV or not). Developing such a specific intervention was challenging, given that we faced a scarcity of operational and practical guidelines for doing so, particularly with regard to creating and translating the relational skills into preprogrammed schemes of communication between the virtual nurse and VP. Although we strongly believe in the value of development frameworks [19,20] to broadly inform intervention development, these frameworks are not intended to offer the extensive and practical description needed to guide the creative process of translating relational skills into an intervention-specific VP simulation. Furthermore, while the evaluation of interventions is well known to be a core research activity, the interventions themselves and their methods remain poorly described in the literature. Cheng et al [10,22] supported the importance of providing a detailed description of the development process of simulation.

Relevance of This Viewpoint Paper

This viewpoint paper attempts to fill the gaps cited earlier. In addition, this study may be helpful to VP simulation developers, educators responsible for CPD, and health care professionals wanting to undertake such an endeavor. At the same time, as our paper ensures transparency in reporting the development methods, the components of the VP simulation (including its content), and the lessons learned, it provides insights into an approach to codeveloping a theory-informed VP simulation aimed at improving nurses' relational skills with patients. This viewpoint paper provides practical cues on how to deliver and

XSL•FO
RenderX

translate MI through technology into preprogrammed nurse-patient dialogues.

Scope, Aim, and Objectives

This viewpoint paper is an opportunity to make explicit the tacit knowledge that comes from our personal experience in codeveloping a VP simulation, supported by theoretical approaches and the literature. The specific objectives are to describe (1) the entire codevelopment process of a VP simulation, including its guiding principles, its content, and features and (2) the challenges encountered and the strategies used to overcome them.

Methods

Codeveloping a Virtual Patient Simulation and Its Content

In this section, we describe the methods and processes involved in 2 main phases: (1) planning the VP simulation development and (2) designing the content, sequence, and format of the VP simulation. We also identify the deliverables (or outputs) of each phase and subphase, such as the qualitative evidence of the needs assessment, the composition of an interprofessional team, the clinical content, and the graphical presentation of the VP simulation. We used a collaborative and creative approach to codeveloping the VP simulation. The term *codevelop* is preferred because 2 nurses worked together to produce the clinical content from scratch, along with the project manager of the VP simulation team, and with the support of a larger interprofessional team. Some models [23,24] and standards of best practice in simulation [25] inspired and broadly informed the general codevelopment processes. Although being very useful, especially in the planning phase, these models did not provide operational guidance for the design phase, particularly with regard to creating the full conversational script of the nurse-patient consultation rooted in relational skills. Therefore, describing this process of codevelopment was a retrospective exercise: it was not preplanned but was rather emergent.

The processes are presented in a linear manner to facilitate understanding. However, in reality, they were conducted concurrently, and some in a preordered sequence (the planning phase was done before the design phase). We used field notes, meeting summaries, and debriefing sessions with team members to keep track of important decisions throughout the project. In this paper, sentences or groups of words in *italics* are used to emphasize important guiding principles or key elements surrounding the codevelopment of the VP simulation.

Phase 1: Planning the Virtual Patient Simulation Development

There are 4 subphases related to the planning phase: (1.1) assessing training needs; (1.2) selecting theoretical approaches to inform the VP simulation development; (1.3) negotiating a detailed partnership contract between the research institution, the researcher, and the VP simulation company; and (1.4) assembling an interprofessional team.

1.1. Assessing Training Needs by Understanding HIV Nursing Practice and Its Challenges

To gain insight into HIV nursing practice and the challenges nurses face when providing antiretroviral therapy–related care, a qualitative study [16] was conducted, identifying 3 challenges. The first was performing nursing roles at the interface of social and biomedical boundaries: nurses sometimes felt unequipped to perform their social role. The second challenge was the lack of alignment in the expectations of nurses and patients regarding antiretroviral therapy adherence. The third focused on dealing with the sociopolitical determinants affecting access to health care resources and services.

From these results, *one challenge was chosen and prioritized because it targeted a clinical behavior amenable to change within the nurse-patient relationship and offered room for practice improvement and change*: the misalignment between the expectations of nurses and patients regarding medication adherence. Some nurses had an optimistic attitude toward antiretroviral therapy; they encouraged patients to take their medication while expecting an outcome of adherence that was not achieved. Nurses felt powerless in such situations of nonadherence, as though they had failed in their role. One of the strategies used by nurses to overcome this challenge was to use the MI [17] approach. This challenge was then transformed into a learning opportunity and thus appeared feasible for translation into a simulated clinical situation.

1.2. Selecting Theoretical Approaches to Inform Virtual Patient Simulation Development

We selected 3 theoretical approaches to lay the groundwork for the VP simulation, each serving a different purpose. First, strengths-based nursing (SBN) was chosen as a philosophy and value-driven approach [26] to clarify the goals and mission of nursing overall. *It offers a lens through which to view the roles of nurses and patients, the focus of nursing care, and the nature of the nurse-patient relationship. The assumptions and values underlying SBN guided the conceptualization of the core of the simulation to be a collaborative partnership within the nurse-patient relationship.* Second, while SBN allows for a broad conception of the role of nurses, *MI is the approach used to design the virtual nurse's key actions in the simulation. MI was used to inform the creation of the VP simulation content (ie, the nurse-patient dialogue, quizzes, and feedback)*. The use of MI aimed to operationalize content creation. Third, *considering the learner-centered vision of the VP simulation, the principles of adult learning theories were identified to support the learning processes and activities*. SBN and adult learning theories were used for the basic structure of the VP simulation.

Strengths-Based Nursing Approach

SBN is grounded in the principles of person and family-centered care, empowerment, relational care, and innate health and healing. *Collaborative partnership*, one value underlying SBN, meshes with our conception of the nurse-patient relationship. The term *collaborative partnership* refers to power, to the way it is distributed, and how it is shared with the person/family to give them a voice in achieving their goals [26]. In this approach,

XSL•FO
RenderX

goals are set by the nurse and patient together, each bringing their own experience, knowledge, and competencies to the relationship [27]. Under a facilitator role, the nurse provides guidance to help patients find their own solutions. The recognition of the patient's strength is a key element in this collaborative relationship.

In Textbox 1, we present the criteria underlying the collaborative partnership adapted from SBN [28], showing how this form of the nurse-patient relationship can guide nursing practice based on different assumptions, principles, and values. The purpose here is to raise awareness and reflect on how nursing activities and actions can potentially be influenced or shaped by broad conceptions of the nurse-patient relationship, whether these conceptions are known or unknown, conscious or unconscious, implicit or explicit.

Textbox 1. Collaborative partnership within the nurse-patient relationship; strengths-based care.

Focus

The person's capacity to be well, have a high quality of life, and experience it in a meaningful manner

Role of the nurse

The nurse is a facilitator. Encourages people to share their perspectives and expertise, participate in shared decision-making processes, develop their autonomy, and use their strengths and resources

Role of the person (includes patient/client and the family)

Acts as an active partner who plays an important role in setting goals and in looking for solutions that best match the person

Nature of the relationship

It is reciprocal and mutual, more symmetrical, with continuous negotiation of goals, roles, and responsibilities. Both partners give and receive and gain and grow

Goal setting

Goals are jointly set

Evaluation

Nurse and person share in the joint assessment of progress in reaching mutually determined goals

Expected outcome

The problem may or may not be solved, but the person's skills for managing current or future problems are reinforced. Joint responsibility is accepted for the outcomes

Motivational Interviewing

The SBN approach and MI share common roots and a core value: collaborative partnership. The person-centered communication style characterizing MI [17] can be established only in a relationship where these 4 values are present: partnership, acceptance, evocation, and compassion. In MI, collaborative partnership is about honoring and respecting the person's autonomy and seeking to understand the person's internal frame of reference.

In addition to these values, the VP simulation is based on a number of elements related to MI. First, the VP simulation is divided into 4 processes: (1) engaging, (2) focusing, (3) evoking, and (4) planning (see definitions given in Textbox 2).

XSL•FO
RenderX

Textbox 2. Definitions of the 4 motivational interviewing processes that are the building blocks of the virtual patient simulation.

1. *Engaging:* "Engaging is the process of establishing a mutually trusting and respectful helping relationship" [17]. Engaging in the relationship can be very brief at times, as is the case in the consultation between Mr Wilson and the nurse who have known each other for a long time. At other times, engaging can be very long. However, in all cases, it goes beyond stock greetings or courtesy. Engaging must be manifest throughout the therapeutic relationship. It is about the development of a *working alliance* [17]
2. *Focusing:* Focusing serves to pinpoint the goal to be achieved to set and maintain a direction. To this end, nurses clarify the following with patients:
   - Their values and goals
   - The nature of the desired change
   - The importance patients attribute to the change
3. *Evoking:* "Evoking involves eliciting the client's own motivation for change" [17]. At this stage, nurses guide patients in exploring their motivation, while simultaneously evoking hope and confidence. To this end, nurses:
   - Explore with patients their perceptions (*ambivalence*) about change
   - Mobilize relational skills to elicit *change talk* (a discourse in favor of changing the patient's health-related behavior such as medication adherence)
   - Mobilize relational skills to respond to *sustain talk* (a patient's discourse in favor of the status quo)
4. *Planning*: A clear plan drawn up by the patient and guided by the nurse is conducive to engagement toward change and success in effecting change. To this end, nurses:
   - Recognize signs or cues indicating that patients are ready to take action (ie, increased *change talk or* decreased *sustain talk*)
   - Avoid slipping (back) into a *directing counseling style* by drawing up the plan for the patients
   - Make a transition from *evoking* to *planning* (eg, "Where does this get you?" "What do you intend to do about it?" and "Let's imagine for a moment that you decide to change: How might you go about it?")
   - Mobilize their relational skills and their practical knowledge to clarify the plan
   - Guide patients in anticipating obstacles and solutions to these

Second, the VP simulation is based on relational skills that are essential to the proficient practice of MI (Multimedia Appendix 1) by creating optimal conditions for relational engagement with patients: asking open-ended questions, using reflective listening, summarizing, affirming the patients' strengths, providing information and advice, evoking a hypothetical change, eliciting and scaling change talk, setting patient-determined goals, and arriving at a plan. These relational skills are called *behavior change counseling techniques* because they are active ingredients that allow providers to initiate or maintain health behavior changes [29-31]. Third, traps or roadblocks (eg, expert and blame traps, a directing counseling style) are also part of the VP simulation (Multimedia Appendix 2). These are interventions, advice, or relational skills inconsistent with MI that are delivered by nurses with the best of intentions to help their patients but are likely to cause relational disengagement and to shut down dialogue with them.

Several reasons justify the use of MI to inform the VP simulation. First, some nurses from the previous work used it in their current practice and found it effective, whereas others clearly expressed the need for training in it [16]. Second, MI training is promising for improving the relational skills of health care professionals working in HIV care [32] and has proven effective in enhancing medication adherence [33-35]. Third, MI is commonly used as a behavioral change counseling approach in nursing, as highlighted in a systematic review conducted by Fontaine et al [30]. This means that some nurses may be familiar with it and associate this approach with prior knowledge and experience.

Adult Learning Theories

Simulation is rooted in certain principles of adult learning theories [36] that describe how adults learn and gain an understanding of clinical expertise. In our case, [37,38] is created by the interaction of nurses with the VP simulation environment. The learners, or nurses (these terms will be used interchangeably), will be actively involved in the simulated experience by practicing and testing relational skills with the VP. They will have opportunities to assess their interventions and reflect on them with the help of quizzes and feedback loops (these will be explained later on). We sought to create a *transformative learning experience* [39]. *Transformative learning* capitalizes on the learners' prior experiences and their interpretation of a situation to build a new way of thinking and then act differently by using critical reflection [36,40]. Feedback is a powerful and effective mechanism to support adult learning and was integrated throughout the VP simulation [41].

In sum, SBN and adult learning theories contributed structuring elements with regard to the goal of the nurse-patient relationship and learning processes. MI informed the integration of concrete communication techniques in the fine content of the VP simulation.

XSL•FO
RenderX

1.3. Negotiating a Detailed Partnership Contract Between the Research Institution, the Researcher, and the Virtual Patient Simulation Company

The coming together of the research team and a representative of the VP simulation company (SimforHealth) was an important aspect of this project. The research team was composed of experts in clinical content, and the company had expertise in digital training for health care providers [42]. A person with extensive experience at the research institution was in charge of negotiating the terms of the contract between the researcher (JC) and the company. These terms were financial considerations, accountability and commitments, mandate duration and timeline, confidentiality duties, intellectual property, and platform use licenses.

1.4. Assembling an Interprofessional Team

An interprofessional team was formed, comprising experts from a VP simulation team and clinical, research, and community-based settings. The VP simulation team was composed of a project manager, a pedagogical engineer, two-dimensional (2D) design professional, 3D graphic designers, and a software engineer. The clinical, research, and community team members included nurses (clinical nurse specialists, a head nurse, researchers, and student researchers, including one with experience using MI with people living with HIV), an infectious disease specialist, the director of an HIV community-based organization, a woman living with HIV, and researchers with experience in developing and evaluating web-based interventions to improve antiretroviral therapy adherence and in health technology assessment.

Furthermore, 2 members of the team (GR and JP) cocreated the clinical content of the VP simulation and worked in close collaboration with the project manager of the VP simulation team. These 3 people made up the working committee. The other team members played a consultative role, providing input on the clinical content and graphical components (eg, the VP, the nurse's office).

Phase 2: Designing the Content, Sequence, and Format of the Virtual Patient Simulation

This second phase includes the following subphases: (2.1) setting the learning objectives and cocreating the clinical content; (2.2) recording the nurse and patient voice-overs; (2.3) designing and validating the 2D learning environment; and (2.4) integrating 3 modes of fidelity to ensure learner engagement and immersion in the VP simulation.

To provide an exhaustive description of the content, sequence, and format of the VP simulation and of how we can translate MI through technology, we have provided detailed information on the following in Multimedia Appendices 1 to 5: the relational skills that are or are not consistent with MI, the key elements of the VP simulation, an excerpt of the writing template, and the table of contents of the glossary. In Multimedia Appendix 3, we offer a comprehensive description of all the key elements (eg, simulation designs such as quizzes and feedback points, learner orientation, exposure, participant groups) constituting the VP simulation, which was adapted from Cheng et al [10,22] and Peddle et al [43].

2.1. Setting the Learning Objective and Cocreating Clinical Content

The learning objectives evolved throughout the clinical codevelopment of the VP simulation and are listed in Multimedia Appendix 3. Cocreating the clinical content involved more than just the wording of the script for the nurse-patient dialogue. It was a creative process that took into account the assumptions, nursing philosophy, and values embedded in the simulated situation (ie, SBN)—the active ingredients of MI and the learners' roles (adult learning theories).

The clinical content as a whole was made up of the following: a prebriefing video featuring the nursing student-researcher and its corresponding text; the simulated clinical situation, also referred to as the "virtual nurse-patient consultation," which includes the patient's electronic record (also named patient's file); a glossary; quizzes and feedback loops; and labels (green and red visual cues).

Key simulation design elements and features, including prebriefing, repetitive practice, MI techniques, quizzes, feedback, labels, and fidelity, were included to optimize interactivity, foster nurses' engagement with and immersion in the training, and to promote their active and transformative learning.

Prebriefing

The simulation-based experience with a prebriefing is considered the best practice in simulation [25]. In our case, the goal of the prebriefing, provided via video and text, was to offer basic information about MI so that all learners could begin the simulated nurse-patient consultation with the same standardized information. It also aimed to set boundaries about the scope of the simulation to help nurses manage their expectations. For instance, some nurses might otherwise have thought that the simulation's goals were to acquire or deepen knowledge of antiretroviral therapy, or that prior experience with people living with HIV was needed (neither of which were the case). The prebriefing was also an opportunity to establish a fiction contract, which refers to a kind of commitment in which nurses are invited to act as though the simulation is real, although acknowledging its limitations [44,45]. Multimedia Appendices 6 and 7 show excerpts of the prebriefing video and a demonstration of the VP simulation.

Patient Electronic Record

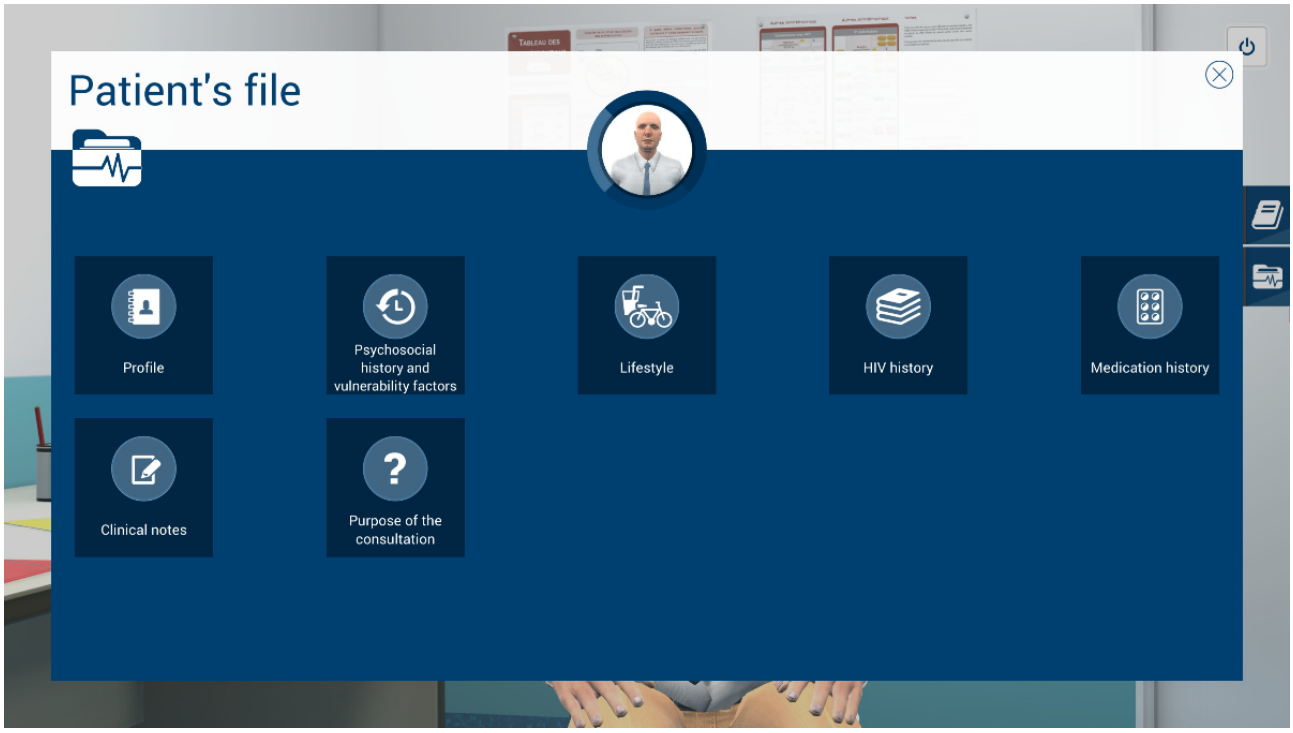
The content of the patient's electronic record was cocreated with the interprofessional team members for 2 reasons: (1) to ensure the credibility and validity of the patient's medical history and (2) to ensure the fidelity with real patient record rubrics in HIV outpatient clinics. The purpose of the patient's consultation is presented in the patient's record and is summarized in Textbox 3. The different rubrics of the patient's electronic record (eg, clinical notes, HIV, and medication history) are illustrated in Figure 1.

XSL•FO
RenderX

Textbox 3. Purpose of consultation.

The story is about Mr Wilson, a 50-year-old man living with HIV since 2011. He generally takes his medication regularly. His viral load was undetectable for 6 years (which is a sign of antiretroviral therapy adherence). Mr Wilson changed his treatment 1 year ago. He was very busy at work, and his routine changed, so it was difficult to take the medication as prescribed. Now, his viral load is over 1000 copies/ml, indicating that the antiretroviral therapy is not as effective (the target is to achieve an undetectable viral load, that is, below 40 copies/ml). What is going on?

Figure 1. Patient's electronic file rubrics.



Full Script of Nurse-Patient Consultation

We created the script as if we were writing a *choose-your-own-adventure* book. We started by writing the *green pathway*, that is, the whole storyline in which the script is based on relational skills that are consistent with MI. Afterward, we identified strategic places in the script to incorporate traps or roadblocks—the *red pathways*. These crossroads between the green and red pathways are introduced by quizzes in which learners must select an intervention. When a red pathway is chosen, a short interaction inconsistent with MI shows learners how some types of communicational traps used by nurses generate negative reactions in the patient's speech. At the end of the interaction, the red pathway stops, written feedback is given to the learners, and they can go back to the crossroads to make a different choice. When the green pathway is chosen, the script (ie, nurse-patient dialogue) becomes automated until the next crossroads. We had to carefully select the strategic places within the script to insert the crossroads. Thus, the arborescence that supports the whole script is made up of the green and red pathways, including quizzes and feedback loops.

All these relational skills, whether they are consistent with MI (eg, open-ended questions) or not (eg, providing advice without asking for permission), were integrated into a nurse-patient dialogue in which learners had to choose between the interventions that generate openness in the patient's speech and those that can shut down communication. This scenario provides learners with safe, constructive, positive, and nonjudgmental spaces that allow for transformative learning and self-reflective practice [36].

The main stages in writing, validating, and producing (ie, filming, voice-over recording) the overall content and testing it on the platform are summarized in Textbox 4. Steps 1 to 4 are more general and account for how we created a script for nurse-patient consultation.

In our various attempts to write the script, we developed multiple versions of a writing template. This was a work in progress throughout. The final template (Multimedia Appendix 4) includes the nurse-patient dialogue (script), quizzes, feedback, labels, guidelines for the VP simulation team regarding the branching, and notes for the actors who would record the voice-overs. This template was a tool that allowed us to write the script in the manner of a *choose-your-own-adventure* book.

A total of 14 quizzes, including multiple-choice and open-ended questions, are part of the VP simulation. Most were formulated to allow learners to select the most appropriate intervention, depending on the script's progress and the patient's speech. Each quiz is supported by synchronous written feedback focusing on the rationale for the good or bad answer selected based on their consistency with or without MI (Multimedia Appendices 1 and 2).

Visual cues, called *green* and *red labels*, were introduced during the nurse-patient consultation to qualify their speech (eg,

XSL•FO
RenderX

open-ended questions, defensive attitude) and in the written feedback. These labels correspond to theoretical MI techniques and provide feedback. The introduction of visual cues into the automated dialogue allows learners to grasp the rationale for the virtual nurse's communication skills and to observe the patient's reaction while limiting the number of pathways (green and red) to be scripted.

Textbox 4. Main stages in writing, producing, and validating the overall content and making it accessible as a virtual patient simulation platform.

1. Writing the green pathways, meaning the overall storyline between the nurse and patient that is consistent with motivational interviewing, by using behavior change counseling techniques (Multimedia Appendix 1), and structured according to its 4 processes: (1) engaging, (2) focusing, (3) evoking, and (4) planning (Textbox 2). *Note:* The script of the nurse-patient dialogue was first written in a linear way to open up onto the patient's experience. At this stage, no crossroads (red pathways) were part of the script
2. Validating the green pathway with the interprofessional team
3. Identifying strategic places along the green pathway where traps or roadblocks could be integrated and then writing the *red pathways* (Multimedia Appendix 2). This step included the creation of quizzes and feedback
4. Creating visual cues that we called *green* and *red labels* to engage the learner observing the preprogrammed dialogue. These labels were integrated into strategic places in the script (eg, when the nurse uses a directing counseling style)
5. Producing the content of the patient's electronic record, the glossary, and the prebriefing (video and text)
6. Validating the whole clinical content with members of the interprofessional team
7. Culturally adapting the script (nurse-patient dialogue). Originally, it was written in French from Quebec (Canada). The text was adapted to *international* French
8. Recording of the nurse's and the patient's speech by professional actors and filming the prebriefing video
9. Integrating all the content in the MedicActiV platform [42]
10. Performing functionality tests and validating the content within the platform
11. Validating the virtual patient simulation with a small group of nurses
12. Launching the French version
13. Translating the virtual patient simulation to English and revising the content in close collaboration with an anglophone nurse who was an expert in motivational interviewing and a member of the Motivational Interviewing Network of Trainers.

Note: Steps 9 to 13 were repeated for the English version.

Glossary

A web-based glossary is available within the VP simulation as supplementary educational material to complement the content of the nurse-patient consultation. It covers theoretical concepts, definitions, and applications. The main topics covered in the glossary are presented in Multimedia Appendix 5.

2.2. Recording the Nurse and Patient Voice-Overs With French- and English-Speaking Actors

The project manager of the VP simulation team was responsible for preselecting French- and English-speaking actors and assisting them during the recording of the nurse and patient voice-overs.

It is important to add stage directions in the writing template, alongside the speech of the nurse and patient, so that the actors can respect the tone of voice and the ambience/vibe in the virtual nurse-patient relationship, consistent with MI.

The nurse's mode of communication had to be respectful, calm, warm, and welcoming, without becoming caricatured, especially in the red pathways. That way, the learners could not guess or deduce the right answer merely from the character's tone of voice or an intervention that is obviously inadequate. As for the patient, if he was feeling worried, the actor had to convey this in his performance.

2.3. Designing and Validating the 2D Learning Environment

At the same time, as the clinical content was being written, the VP simulation team was working on designing the graphical elements of the 2D learning environment. Figure 2 shows a mock-up of the virtual office, including both nurse and patient, and Figure 3 shows the patient only.

These mock-ups allowed us to obtain the team members' opinions of the graphical designs before selecting the final 2D learning environment. It was important to display a chart of antiretroviral therapy on the wall to represent a real-life artifact. Decisions about the virtual nurse and patient had to then be made. For example, would we see the nurse on-screen (third-person view) or not (first-person)? What would the HIV-positive man look like? How would the desk be positioned in the virtual office (ie, between the nurse and patient or on the side)? Ultimately, the first-person view was preferred, where the nurse is not visible on-screen, but we hear her voice and see her speech. In this way, only the VP is visible. The desk was placed on the side to avoid creating distance between the nurse and patient. Figure 4 shows the final design of the virtual nurse's office and the patient (Mr Wilson).

Some features were created to fit with this project, such as the green and red labels, and the interfaces to support synchronous feedback. Furthermore, SimforHealth developed a specific content and dialogue management module (questions/answers)

XSL•FO
RenderX

to meet the clinical team's needs. This module made it possible to speed up the content creation (full script of the nurse-patient consultation) and facilitated content integration into the VP simulation.

All the regular 3D content was created using 3Ds Studio Max (Autodesk), the leading 3D digital content creation solution (Figure 5). Figure 6 shows examples of the wireframes used in the creation of the 3D content. Wireframes are illustrations of the proposed VP simulation components and "assist in visual communication and design of the structure, functionality, learner interface, and positioning of an application" [24].

Figure 2. Mock-up of the virtual office including nurse and patient.



Figure 3. Mock-up of the virtual patient.

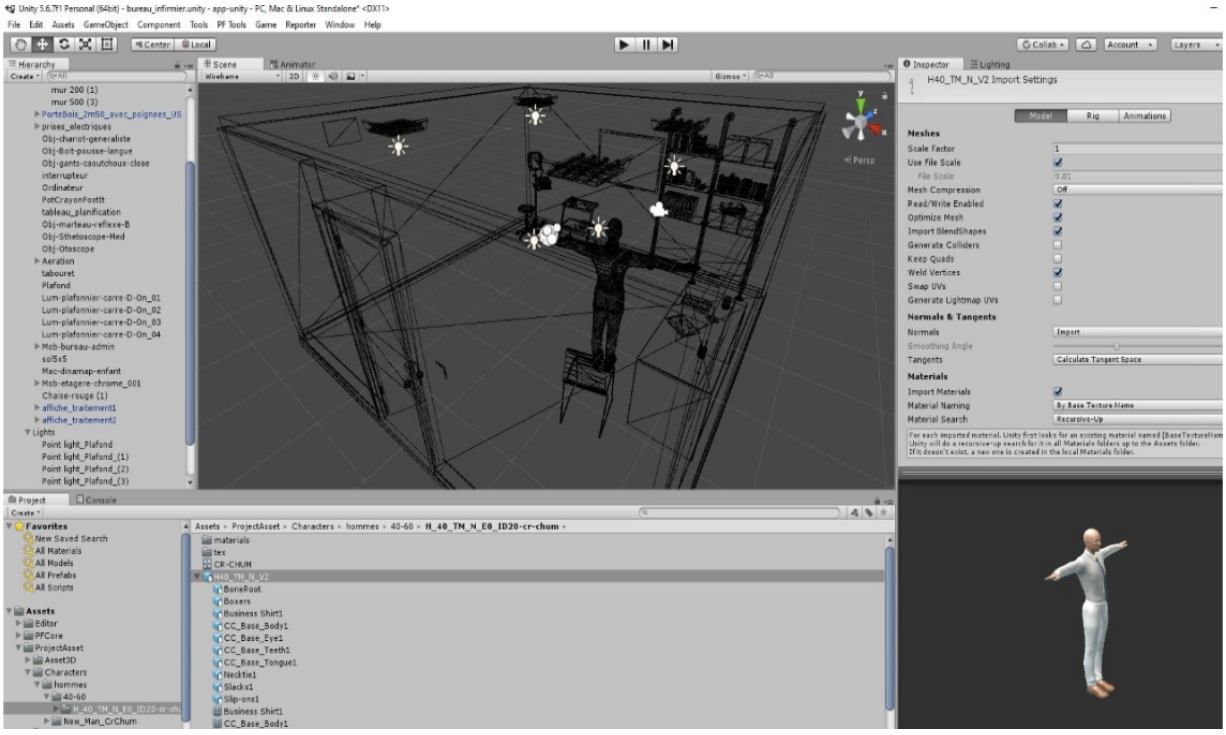


XSL•FO
RenderX

Figure 4. Final design of the virtual patient simulation.



Figure 5. 3D digital content creation solution.



MedicActiV Platform

The MedicActiV platform [42] was designed to support digital training for health care providers and was used to host the simulated nursing scenario. The simulated scenario is a self-directed learning approach that is accessible via a computer or tablet. It is a 2D visual interface (Figure 4) with 3D content and is considered nonimmersive virtual reality [46]. The 2D elements include a graphical user interface, such as colors, icons, and text-based interfaces. The 3D elements are the dynamic *objects* created to have volume, such as the VP and the nurse's office (Figure 6). MedicActiV is a software as a service solution that provides a library of virtual clinical cases. Thanks to the authoring tool, developers (eg, researchers, educators, health care providers) can easily create their own clinical case to share it with other users on the platform. The trainer can set up the training sessions' parameters (eg, date, time, duration) and, immediately after any session, is able to retrieve several key performance indicators (eg, time, errors). Most of the clinical cases are available on multiple devices and operating systems

XSL•FO
RenderX

(eg, Windows, macOS, iOS, Android). However, our clinical case with the HIV-positive VP was not available to everyone because it was embedded in a research process, and thus, was available only to a small sample of research participants, namely, the nurses.

Figure 6. 3D Virtual patient.



2.4. Ensuring Fidelity to Ensure Learner Engagement and Immersion Throughout the VP Simulation

One important feature to include in any simulation, be it virtual or not, is the concept of fidelity, also known as realism or authenticity [47]. Fidelity is about the learner's perceptions of how well a simulation represents or reproduces reality [44] and offers them a realistic learning opportunity [15]. In this context, 3 modes of fidelity represent the ways humans think about reality: (1) physical; (2) conceptual (or semantic); and (3) emotional and experiential (or phenomenal) [44,45,48]. These modes can influence the learner's engagement with and immersion into the simulation as well as their learning process [15,44,48].

The *physical mode* can be described by the *physical* properties of the VP (eg, movements, appearance) and of the nurse's office, including objects and artifacts in that environment. The voice-over acting for the nurse and patient (eg, voice tone, vocabulary) are other elements of fidelity that were taken into account. The *conceptual* or *semantic mode* relates to concepts and their relationships, for example, theories, meaning, or information that are presented via various means (eg, text, voice-overs). It also involves the *if/then* relationship. In our simulated situation, it is globally represented by the following logic: if the nurse uses behavior change counseling techniques (communication skills) consistent with MI, this will open up a dialogue with the VP. If the nurse uses relational skills inconsistent with MI, then the VP will react accordingly (eg, defensively). Finally, the *emotional* and *experiential mode* refers to the learner's emotions, feelings, and beliefs relating to their holistic experience of participating in the simulation. In Table 1, we present strategies used to ensure that the VP simulation had the characteristic of fidelity.

XSL•FO
RenderX

Table 1. Strategies used to ensure 3 modes of fidelity throughout the virtual patient simulation.

Modes of fidelity	Strategies used
Physical fidelity	Giving cues to the VP[a] simulation team for their design of the VP: • Taking a picture of a real nurse's office for the representation of the virtual nurse's office • Adding objects that represent reality (eg, chart of antiretroviral therapy on the wall)
	Getting charts from real patient records to use the same vocabulary in the VP's electronic record
	Designing the VP with human features (eg, facial expression and body movements) based on past experiences of the VP simulation team
	Using real voice-overs for both patient and nurse: • Preselecting some French- and English-speaking actors, listening to their audio tracks and choosing the ones that best fit the spirit of the simulation • Providing descriptive cues in the script besides the nurse's and patient's speech, so that the actors can respect the tone of voice and ambiance and vibe in the nurse-patient *virtual* relationship, consistent with SBN [b] and MI [c]
	Getting written approval for each step of the graphical design before undertaking the subsequent one
Conceptual fidelity	Co-designing the clinical content with a nurse having expertise in HIV care and MI to ensure that the simulation reflects the nursing practice and the validity of the theory injected in preprogrammed interactions
	Meeting a clinical nurse specialist who is an expert in HIV care to discuss real-life situations of people living with HIV having difficulty taking their medication to identify nursing actions, in partnership with physicians and pharmacists, in nonadherence situations
	Validating the content with health care professionals
	Working with pedagogical engineering to make sure that good educational practices are met
Emotional and experiential fidelity	Promoting a positive learning experience and relatedness by creating messages that value and respect nurses' competencies and current practice. In doing so, clinical content inventors must themselves be consistent with the MI values in their way of translating the educational content to the simulation
	Creating opportunities for reflection on action (virtual practice) by incorporating quizzes and feedback that represent what nurses do in their current practice
	Establishing a fiction contract with nurses [44,48]. The following message in the prebriefing video is intended to create such a contract: "The relational aspect of the practice of caring cannot be simulated to perfection. Needless to say, human beings are not preprogrammed to respond to a nurse this way or that. I invite you, then, to immerse yourself in this virtual simulation as if it were real and to pay attention to the interactions between Mr. Wilson and the nurse".

[a]VP: virtual patient.

[b]SBN: strengths-based nursing.

[c]MI: motivational interviewing.

Discussion

Challenges and Strategies

Thus far, we have described VP simulation development methods and the deliverables of each phase and subphase, including the theory-informed content. We now discuss the main challenges and strategies used to overcome them.

The most salient challenges and strategies are discussed in this section. The second phase (ie, designing the content, sequence, and format of the VP simulation) was the most challenging because of the difficulty in determining realistic clinical content in a virtual learning environment. We had to preserve the philosophy and values of SBN and the collaborative partnership that shaped the nurse-patient relationship. We were also concerned with translating relational skills informed by MI within the simulated nurse-patient dialogue while ensuring that the functionalities of the VP platform could support the content. *We also had to find a balance between a nursing-driven approach and a technology-driven approach while optimizing the learning experience.*

Challenge 1: Starting the Project With Divergent Conceptions of Virtual Patient Simulation Approaches

At the beginning of the project, the VP simulation team suggested a generic template that could help structure and organize the clinical content into pre-established *categories* consistent with virtual simulators already developed by the company. However, the 2 clinical inventors (GR and JP) felt that such a template would not be helpful because it reflected a problem-solving [47,49] or procedural [6] simulation approach. This is appropriate for teaching, for instance, clinical reasoning and diagnosis. To perform these tasks, learners must, for example, collect data and make diagnoses and treatment decisions based on their anamneses. In such simulations, learners are given a set of information from which they must draw conclusions (eg, in patient-facing problems with adherence to antiretroviral therapy, the nurse's identification of solutions drives the direction of the VP simulation script). However, this approach did not mesh with a collaborative partnership as a form of the nurse-patient relationship, and thus with the nature of relational skills that are consistent with a narrative approach to simulation [47,49]. The ways of seeing the approach to the

XSL•FO
RenderX

VP simulation diverged between the simulation team (who perceived it as a procedural one) and the clinical team (who perceived it as a narrative one).

The narrative approach, also called *situational simulation*, [6] is generally found in a personal storyline that progresses over time around a logic of cause and effect and that involves a decision-making process that yields different *outcomes* (or effects). The script is anchored in a strengths-based approach, consistent with MI, in which the virtual nurse's role is that of a facilitator, using relational skills to open up the dialogue with the patient (eg, where a patient is facing problems with adherence to antiretroviral therapy, the identification of solutions *by the patient* drives the direction of the VP simulation script). At the time we started the VP simulation project, it would have been helpful to rely on the paper of Peddle et al [43] because it describes the development of VP to support undergraduate nursing students in learning nontechnical skills such as communication. The authors give an overview of the narrative approach and the *choose-your-own-adventure* game structure used to design the VP simulation.

Strategies: Getting a Mutual Understanding of the VP Simulation Approach That is Aligned With the Philosophy and Values of a Collaborative Nurse-Patient Relationship

We first consulted some virtual clinical cases on the MedicActiV platform to familiarize ourselves with the possibilities and limitations of the virtual learning environment and to understand how it works. It helped to understand how the VP simulation team perceived the simulated clinical situations and then pinpointed the differences between our different VP simulation approaches (procedural vs narrative). We collaborated closely with the project manager of the VP simulation team from the outset of the project. It was essential to make sure that the narrative approach, the SBN, the relational skills informed by MI, the principles of adult learning theories, and the functionalities of the simulation platform all meshed together. Holding regular meetings with the working committee allowed us to gain a better understanding of each other's roles, responsibilities, and evolving perspectives on the VP simulation.

Challenge 2: Struggling With the Complexity of Designing Realistic Relational Skills Into Preprogrammed and Simulated Nurse-Patient Dialogue While Preserving an Immersive Learning Experience

This challenge was three-fold: (1) translating complex actions (uptake of relational skills) within fully automated and preprogrammed nurse-patient dialogue, (2) having insufficient guidance in integrating such skills in this form of virtual simulation, and (3) designing the proper immersive and realistic 2D learning environment.

Writing the scripts with high-quality motivational responses is not just about wording. It is about creating a natural flow of interactions that involves great sensitivity and attention to verbal and nonverbal aspects of communication while respecting values such as empathy, collaborative partnership, acceptance, affirmation, and so on. Such an endeavor is easier to put into action spontaneously in real-time interactions. This challenge is well summarized by Villaume et al [50]: "While the processes and skills of MI are theoretically understandable, using them in individual utterances requires a considerable adjustment of vocabulary, grammar, emotional tone, and rhetorical strategy. Trying to work through these adjustments in real time with a standardized patient is difficult."

Adding to the complexity of the actions to be performed is the lack of guidance on how to create and translate preprogrammed interpersonal nurse-patient interactions. It is easy to become quickly overwhelmed by the multiplicity of alternative scenarios and the growing decision tree (arborescence) if we do not limit the number of points at which learners can make choices and the number of options they have.

At the same time, as we were writing the script, SimforHealth had to design immersive and realistic nurse-patient interactions in a virtual learning environment, which was challenging. The decision tree is a kind of backbone for the nursing content, but it needs to be enhanced with graphic design, such as representing the *real appearance* of the VP's digital facial expression and behavior. This is key to achieving a better commitment from the learners.

Strategies: Creating a Collaborative and Work-in-Progress Writing Template

It is certain that we cannot make the nature of relational skills less complex, be they consistent with MI or not. However, the VP simulation's clinical inventors, as educators, were accountable for being skilled with using MI to be able to transfer theoretical knowledge in an understandable way within the VP simulation and allow for an optimal learning experience. The quality of the VP simulation depends partly on the inventors' expertise. In the working committee, the *codevelopment* involved knowledge transfer and an opportunity for discussion between these 2 members, given that one (JP) has expertise in MI whereas the other (GR) was new to this theory.

We had to build up our own method and find our own guidance to develop the VP simulation. The writing journey required a high level of creativity and inductivism. The flow of the storyline evolved over time, with a concern for making connections with theory and practice, so that nurses would benefit from a constructive and positive learning experience. In doing so, teamwork was the most important success element. To begin with, the 2 members of the working committee (GR and JP) worked in close collaboration to write the content for the nurse-patient consultation. We created a *writing template* (Multimedia Appendix 4) that evolved over time and represented a guide for both the clinical and the VP team. We wrote the script as a *choose-your-own-adventure* book, that is, by starting to write the green pathway, and then adding the red pathways. In addition, the role of learners' emotions and feelings was taken into account. *From our perspective, the conception of the clinical content (scenario) had to depict the nursing practice (ie, what nurses actually do) without falling into stereotypes or judgments. The clinical content inventors themselves had to be consistent with MI principles and values in their way of translating the educational content to the simulation. Learners had to recognize themselves in the whole scenario (including traps and roadblocks) to be engaged with and immersed in the*

XSL•FO
RenderX

learning experience (fidelity). This principle is comparable to what Yardley et al [51] call "promotion of a learner's positive experience and relatedness."

Strengths of VP Simulation

We believe our VP simulation has a long life span because the action of adopting relational skills consistent with MI is *stable*, and does not need regular updates, as it would if the simulation focused on pharmacological treatment, where scientific breakthroughs occur frequently. Behavioral change counseling techniques (eg, asking open-ended questions, summarizing) are timeless. Another strength is that communication is a transversal skill used by all nurses, from novice to expert, and by all other health care professionals. Relational skills are then applicable independently of the health care professionals' roles, the context of practice or level of experience, and of the population they care for. Focusing on this clinical behavior is a good avenue to ensure the adaptability, sustainability, and efficiency of this VP simulation. We ensured the integrity and quality control [52] of the MI intervention by relying on original work by the authors on MI [17] and by involving nurses who are experts in this approach: one nurse was involved in the design of the French content and a certified trainer in MI revised the English version. Furthermore, the involvement of a patient-partner added value at the early stage of the project, adding insight into the credibility and fidelity of the storyline and the graphic design of the VP simulation. It is a theory-informed intervention, and thus, we are confident in the quality of the clinical content and its potential to induce a practice change.

Conclusions

The knowledge gained from our experience of VP simulation codevelopment has the potential to optimize the development process of other VP simulations, particularly those using a narrative approach. VP simulation is the only means to support learning; it is not the technology per se that generates learning [53]. Many elements were put together to create favorable conditions for generating a positive learning experience. We strongly believe that the duration of the codevelopment process can be shorter if developers have a clear idea of the theories to use to structure the VP simulation (eg, SBN, MI, adult learning theories) and if they know which approach fits (eg, narrative, problem-based) with their perspective of the clinical behavior or action to address. This paper offers concrete examples of how to translate behavior change counseling techniques derived from MI into an asynchronous and preprogrammed nurse-patient dialogue. The inductive approach used in codeveloping the content of the VP simulation was a transformative learning experience for the working committee.

With a view to professional development, nurses will have the opportunity to try this VP simulation, informed by experiential, theoretical, and empirical knowledge, and thereby help evaluate it. The standardized nature of the intervention is a strength of this approach that could be helpful for evaluation purposes. If the nurses' relational competencies are enhanced, then the quality of the therapeutic relationship between nurse and patient may benefit and, ultimately, this can have positive repercussions on the health of people living with HIV.

Acknowledgments

The authors would like to thank the following people for their contributions at various stages of the project: Dominique Debay, Hélène Morin, Alexandra de Kiewit, Catherine Dufault, Kenneth Monteith, Julien Bouix-Picasso, John Kayser, and Production Multimédia du Centre hospitalier de l'Université de Montréal. The authors are grateful to Zozita Translation and Editing for their participation in editing this manuscript.

GR is funded by doctoral fellowships from the partnership between Fonds de recherche du Québec Santé (FRQS) and the Quebec Strategy for Patient-Oriented Research Support Unit, the Canadian Institutes of Health Research (CIHR), Quebec Network on Nursing Intervention Research, and Quebec's Ministry of Higher Education. JC holds the research chair in innovative nursing practices. MPG holds the Tier 2 Canadian Research Chair on Technologies and Practices in Health. VML is funded by Clinical Research Scholars–Junior 1 from the FRQS. The authors thank the Quebec Population Health Research Network for its contribution to financing this publication. GF was supported by the Vanier Canada Graduate Scholarship (CIHR), a doctoral fellowship from the FRQS, multiple scholarships from the Canadian Nurses Foundation, a doctoral scholarship from the Montreal Heart Institute Foundation, a doctoral scholarship from Quebec's Ministry of Higher Education, and multiple scholarships from the Faculty of Nursing at Université de Montréal.

The development of this VP simulation was funded by a student project grant from Quebec Network on Nursing Intervention Research, by a planning and dissemination grant from CIHR (144030) and by a doctoral scholarship allocated to GR from CIHR (337439). The collaborators of SimForHealth at Bordeaux, France, are Clémentine Dupouy, Olivier Gardinetti, Nathalie Piérard, Stéphane Petit, Eric Méresse, Anne-Sophie Milcent, Xavier Abadie, and Jérôme Leleu.

Conflicts of Interest

JC, the University of Montréal Hospital Research Centre, and SimforHealth are linked by a partnership contract. SimforHealth is the owner of the MedicActiV platform that supported the VP simulation and was involved throughout the codevelopment process. GR and JP codeveloped the clinical content of the VP simulation and JC validated it.

Multimedia Appendix 1

XSL•FO
RenderX

Definitions and examples consistent with motivational interviewing.
[DOCX File , 20 KB - jmir_v22i7e18225_app1.docx]

Multimedia Appendix 2
Definitions and examples inconsistent with motivational interviewing.
[DOCX File , 21 KB - jmir_v22i7e18225_app2.docx]

Multimedia Appendix 3
Key elements of the VP simulation.
[DOCX File , 22 KB - jmir_v22i7e18225_app3.docx]

Multimedia Appendix 4
Excerpt of the final writing template.
[DOCX File , 21 KB - jmir_v22i7e18225_app4.docx]

Multimedia Appendix 5
Table of contents glossary.
[DOCX File , 13 KB - jmir_v22i7e18225_app5.docx]

Multimedia Appendix 6
Prebriefing video (an excerpt).
[MP4 File (MP4 Video), 150207 KB - jmir_v22i7e18225_app6.mp4]

Multimedia Appendix 7
Demonstration of the virtual patient simulation (video).
[MP4 File (MP4 Video), 33765 KB - jmir_v22i7e18225_app7.mp4]

References

1. The Importance of Evaluating the Impact of Continuing Nursing Education on Outcomes: Professional Nursing Practice and Patient Care. American Nurses Credentialing Center: ANCC. 2014. URL: http://www.nursecredentialing.org/Accreditation/ResourcesServices/Evaluating-the-Impact-CNE-Outcomes.pdf [accessed 2020-05-16]
2. Fleet LJ, Kirby F, Cutler S, Dunikowski L, Nasmith L, Shaughnessy R. Continuing professional development and social accountability: a review of the literature. J Interprof Care 2008;22(Suppl 1):15-29. [doi: 10.1080/13561820802028360] [Medline: 19005951]
3. Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2009 Apr 15(2):CD003030 [FREE Full text] [doi: 10.1002/14651858.CD003030.pub2] [Medline: 19370580]
4. Nickerson M, Morrison B, Pollard M. Simulation in nursing staff development: a concept analysis. J Nurses Staff Dev 2011;27(2):81-89. [doi: 10.1097/NND.0b013e3181a68abd] [Medline: 21430479]
5. Berndt A, Murray CM, Kennedy K, Stanley MJ, Gilbert-Hunt S. Effectiveness of distance learning strategies for continuing professional development (CPD) for rural allied health practitioners: a systematic review. BMC Med Educ 2017 Jul 12;17(1):117 [FREE Full text] [doi: 10.1186/s12909-017-0949-5] [Medline: 28701199]
6. Cant RP, Cooper SJ. Simulation in the internet age: the place of web-based simulation in nursing education. An integrative review. Nurse Educ Today 2014 Dec;34(12):1435-1442. [doi: 10.1016/j.nedt.2014.08.001] [Medline: 25156144]
7. Ellaway R, Candler C, Greene P, Smothers V. An Architectural Model for MedBiquitous Virtual Patients. MedBiquitous. 2006. URL: http://groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture [accessed 2019-10-31]
8. Cant RP, Cooper SJ. Simulation-based learning in nurse education: systematic review. J Adv Nurs 2010 Jan;66(1):3-15. [doi: 10.1111/j.1365-2648.2009.05240.x] [Medline: 20423432]
9. Cant RP, Cooper SJ. The time is right for web-based clinical simulation in nursing education. J Nurs Educ Pract 2015 Aug 26;5(11):113. [doi: 10.5430/jnep.v5n11p113]
10. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, International Network for Simulation-Based Pediatric Innovation, Research,Education (INSPIRE) Reporting Guidelines Investigators. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. Simul Healthc 2016 Aug;11(4):238-248. [doi: 10.1097/SIH.0000000000000150] [Medline: 27465839]

XSL•FO
RenderX

11. Kaplonyi J, Bowles K, Nestel D, Kiegaldie D, Maloney S, Haines T, et al. Understanding the impact of simulated patients on health care learners' communication skills: a systematic review. Med Educ 2017 Dec;51(12):1209-1219. [doi: 10.1111/medu.13387] [Medline: 28833360]
12. Bracq M, Michinov E, Jannin P. Virtual reality simulation in nontechnical skills training for healthcare professionals: a systematic review. Simul Healthc 2019 Jun;14(3):188-194. [doi: 10.1097/SIH.0000000000000347] [Medline: 30601464]
13. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jul 2;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
14. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
15. Peddle M, Bearman M, Nestel D. Virtual patients and nontechnical skills in undergraduate health professional education: an integrative review. Clin Simul Nurs 2016 Sep;12(9):400-410. [doi: 10.1016/j.ecns.2016.04.004]
16. Rouleau G, Richard L, Côté J, Gagnon M, Pelletier J. Nursing practice to support people living with HIV with antiretroviral therapy adherence: a qualitative study. J Assoc Nurses AIDS Care 2019;30(4):e20-e37 [FREE Full text] [doi: 10.1097/JNC.0000000000000103] [Medline: 31241513]
17. Miller W, Rollnick S. Motivational Interviewing, Third Edition: Helping People Change. New York, USA: The Guilford Press; 2013.
18. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
19. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
20. Colquhoun HL, Squires JE, Kolehmainen N, Fraser C, Grimshaw JM. Methods for designing interventions to change healthcare professionals' behaviour: a systematic review. Implement Sci 2017 Mar 4;12(1):30 [FREE Full text] [doi: 10.1186/s13012-017-0560-5] [Medline: 28259168]
21. Llewellyn CD, Ayers S, McManus C, Newman S, Petrie KJ. Health Care Practioner Behavior. In: Cambridge Handbook of Psychology, Health and Medicine. Cambridge: Cambridge University Press; 2019.
22. Cheng A, Auerbach M, Hunt EA, Chang TP, Pusic M, Nadkarni V, et al. Designing and conducting simulation-based research. Pediatrics 2014 Jun;133(6):1091-1101. [doi: 10.1542/peds.2013-3267] [Medline: 24819576]
23. Bartholomew EL, Markham C, Ruiter R, Fernandez M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco, CA: Jossey-Bass A Wiley Brand; 2016.
24. Olszewski AE, Wolbrink TA. Serious gaming in medical education: a proposed structured framework for game development. Simul Healthc 2017 Aug;12(4):240-253. [doi: 10.1097/SIH.0000000000000212] [Medline: 28027076]
25. -. INACSL standards of best practice: simulationSM simulation design. Clin Simul Nurs 2016 Dec;12:S5-12. [doi: 10.1016/j.ecns.2016.09.005]
26. Gottlieb LN, Gottlieb B. Strengths-based nursing: a process for implementing a philosophy into practice. J Fam Nurs 2017 Aug;23(3):319-340. [doi: 10.1177/1074840717717731] [Medline: 28795856]
27. Gottlieb LN. Strengths-based nursing. Am J Nurs 2014 Aug;114(8):24-32; quiz 33,46. [doi: 10.1097/01.NAJ.0000453039.70629.e2] [Medline: 25036663]
28. Gottlieb L. Strengths-Based Nursing Care: Health And Healing For Person And Family. New York, USA: Springer; 2013.
29. Dragomir AI, Julien CA, Bacon SL, Boucher VG, Lavoie KL, Canadian Network for Health Behavior Change and Promotion (CAN-Change). Training physicians in behavioural change counseling: a systematic review. Patient Educ Couns 2019 Jan;102(1):12-24. [doi: 10.1016/j.pec.2018.08.025] [Medline: 30172573]
30. Fontaine G, Cossette S, Maheu-Cadotte M, Mailhot T, Heppell S, Roussy C, et al. Behavior change counseling training programs for nurses and nursing students: a systematic descriptive review. Nurse Educ Today 2019 Nov;82:37-50. [doi: 10.1016/j.nedt.2019.08.007] [Medline: 31425931]
31. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
32. Beach MC, Laws MB, Rose G, Roter D, Lee Y, Chander G, et al. Effects of minimal versus intensive intervention to enhance motivational interviewing in HIV care. AIDS Behav 2018 Jan;22(1):276-286 [FREE Full text] [doi: 10.1007/s10461-017-1794-6] [Medline: 28578544]
33. Hill S, Kavookjian J. Motivational interviewing as a behavioral intervention to increase HAART adherence in patients who are HIV-positive: a systematic review of the literature. AIDS Care 2012;24(5):583-592. [doi: 10.1080/09540121.2011.630354] [Medline: 22292452]
34. Palacio A, Garay D, Langer B, Taylor J, Wood BA, Tamariz L. Motivational interviewing improves medication adherence: a systematic review and meta-analysis. J Gen Intern Med 2016 Aug;31(8):929-940 [FREE Full text] [doi: 10.1007/s11606-016-3685-3] [Medline: 27160414]

XSL•FO
RenderX

35. Zomahoun HT, Guénette L, Grégoire JP, Lauzier S, Lawani AM, Ferdynus C, et al. Effectiveness of motivational interviewing interventions on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. Int J Epidemiol 2017 Apr 1;46(2):589-602. [doi: 10.1093/ije/dyw273] [Medline: 27864410]
36. Clapper TC. Beyond Knowles: what those conducting simulation need to know about adult learning theory. lin Simul Nurs 2010 Jan;6(1):e7-14. [doi: 10.1016/j.ecns.2009.07.003]
37. Kolb D. Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice-Hall; 1984.
38. Kolb D. Experiential Learning: Experience as the Source of Learning and Development. Boston, MA: Prentice-Hall; 2014.
39. Merriam S. Andragogy and self-directed learning: pillars of adult learning theory. New Dir Adult Contin Educ 2001;2001(89):3. [doi: 10.1002/ace.3]
40. Mezirow J. Learning to think like an adult. Core concepts of transformation theory. In: Learning as Transformation: Critical Perspectives on a Theory in Progress. San Francisco, CA: Jossey-Bass; 2000.
41. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005 Jan;27(1):10-28. [doi: 10.1080/01421590500046924] [Medline: 16147767]
42. MedicActiV: Virtual Simulation Platform for the Training of Health Professionals. 2020. URL: https://www.medicactiv.com/en/ [accessed 2020-06-16]
43. Peddle M, Mckenna L, Bearman M, Nestel D. Development of non-technical skills through virtual patients for undergraduate nursing students: an exploratory study. Nurse Educ Today 2019 Feb;73:94-101. [doi: 10.1016/j.nedt.2018.11.008] [Medline: 30610960]
44. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. Simul Healthc 2007;2(3):183-193. [doi: 10.1097/SIH.0b013e3180f637f5] [Medline: 19088622]
45. Rudolph JW, Raemer DB, Simon R. Establishing a safe container for learning in simulation. Simul Healthc J Soc Simul Healthc 2014;9(6):339-349. [doi: 10.1097/sih.0000000000000047]
46. Slater M, Sanchez-Vives MV. Enhancing our lives with immersive virtual reality. Front Robot AI 2016 Dec 19;3:-. [doi: 10.3389/frobt.2016.00074]
47. Bearman M. Is virtual the same as real? Medical students' experiences of a virtual patient. Acad Med 2003 May;78(5):538-545. [doi: 10.1097/00001888-200305000-00021] [Medline: 12742794]
48. Rudolph JW, Simon R, Raemer DB. Which reality matters? Questions on the path to high engagement in healthcare simulation. Simul Healthc 2007;2(3):161-163. [doi: 10.1097/SIH.0b013e31813d1035] [Medline: 19088618]
49. Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. Med Educ 2001 Sep;35(9):824-832. [doi: 10.1046/j.1365-2923.2001.00999.x] [Medline: 11555219]
50. Villaume WA, Berger BA, Barker BN. Learning motivational interviewing: scripting a virtual patient. Am J Pharm Educ 2006 Apr 15;70(2):33 [FREE Full text] [doi: 10.5688/aj700233] [Medline: 17149413]
51. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
52. Lavoie KL, Moullec G, Lemiere C, Blais L, Labrecque M, Beauchesne MF, et al. Efficacy of brief motivational interviewing to improve adherence to inhaled corticosteroids among adult asthmatics: results from a randomized controlled pilot feasibility trial. Patient Prefer Adherence 2014;8:1555-1569 [FREE Full text] [doi: 10.2147/PPA.S66966] [Medline: 25422587]
53. Clark R, Mayer R. e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. Hoboken, New Jersey: Wiley; 2016.

Abbreviations

2D: two-dimensional
CIHR: Canadian Institutes of Health Research
CPD: continuing professional development
FRQS: Fonds de Recherche du Québec–Santé
MI: motivational interviewing
SBN: strengths-based nursing
VP: virtual patient

XSL•FO
RenderX

Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by M Sobocan; comments to author 10.03.20; revised version received 01.05.20; accepted 03.06.20; published 15.07.20.

Please cite as:
Rouleau G, Pelletier J, Côté J, Gagnon MP, Martel-Laferrière V, Lévesque R, SimforHealth, Fontaine G
Codeveloping a Virtual Patient Simulation to Foster Nurses' Relational Skills Consistent With Motivational Interviewing: A Situation of Antiretroviral Therapy Nonadherence
J Med Internet Res 2020;22(7):e18225
URL: http://www.jmir.org/2020/7/e18225/
doi:10.2196/18225
PMID:32672679

©Geneviève Rouleau, Jérôme Pelletier, José Côté, Marie-Pierre Gagnon, Valérie Martel-Laferrière, Rock Lévesque, SimforHealth, Guillaume Fontaine. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Unmuting Medical Students' Education: Utilizing Telemedicine During the COVID-19 Pandemic and Beyond

Ariella Magen Iancu[1], MSED; Michael Thomas Kemp[2], MD; Hasan Badre Alam[2], MD

[1]University of Michigan Medical School, Ann Arbor, MI, United States

[2]Department of Surgery, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Hasan Badre Alam, MD
Department of Surgery
University of Michigan
2920 Taubman Center
1500 E Medical Center Dr
Ann Arbor, MI, 48109
United States
Phone: 1 734 936 5823
Email: alamh@med.umich.edu

Abstract

Due to the coronavirus disease (COVID-19) pandemic, medical schools have paused traditional clerkships, eliminating direct patient encounters from medical students' education for the immediate future. Telemedicine offers opportunities in a variety of specialties that can augment student education during this time. The projected growth of telemedicine necessitates that students learn new skills to be effective providers. In this viewpoint, we delineate specific telehealth opportunities that teach core competencies for patient care, while also teaching telemedicine-specific skills. Schools can further augment student education through a variety of telemedicine initiatives across multiple medical fields. The explosion of telemedicine programs due to the pandemic can be a catalyst for schools to integrate telemedicine into their current curricula. The depth and variety of telemedicine opportunities allow schools to continue providing high-quality medical education while maintaining social distancing policies.

(J Med Internet Res 2020;22(7):e19667) doi:10.2196/19667

KEYWORDS

telemedicine; medical education; medical school; COVID-19; education, medical, undergraduate; curriculum

Medical Education on Hold

On March 23, 2020, the Association of American Medical Colleges (AAMC) issued guidance on medical student involvement during the coronavirus disease (COVID-19) pandemic, recommending that medical students not participate in direct patient care, unless there is a critical workforce need, and only on a voluntary basis [1]. The COVID-19 pandemic has generated unprecedented stress for our national health care system. Consequently, it has created practical and logistical challenges for the learning environment that can have a lasting impact on medical education [2]. The widely varying predictions, regarding how long the pandemic will continue, raise questions as to when and how traditional clerkships will resume. In light of this, academic centers must consider new measures to continue their mission of training future physicians. There are potential technological solutions to the current challenge of clinical education during a time of social distancing.

The cornerstone of medical education is patient care. While direct patient care has been appropriately limited at most schools, the lessons learned from these critical patient interactions cannot be fully replaced by readings, lectures, case studies, or online modules. Certainly, direct in-person care has important educational value to the medical student. Although not an exact replacement for in-person care, institutions have a unique opportunity to bridge this current gap in clinical education with telemedicine. This viewpoint contributes to the ongoing discussion by offering specific guidance on how schools can incorporate medical students into the massive expansion of telehealth initiatives during and beyond the COVID-19 pandemic.

The Expansion of Telemedicine

In recent years, telemedicine has grown exponentially. Telemedicine applications have expanded from phone triage and ambulatory electronic visits (e-visits) to include mental

XSL•FO
RenderX

health, postsurgical operation follow-ups, and specialty consults [3]. In support of the efforts to socially distance medical care during the COVID-19 crisis, health systems across the United States are ramping up their telehealth programs [4]. As both patients and providers grow more accustomed to these virtual interfaces during the pandemic, many will likely choose to continue using telemedicine going forward [5]. Even before the pandemic, the telemedicine market was predicted to grow from $38.3 billion in 2018 to $130.5 billion by 2025 [3]. Therefore, while telemedicine has seen an acute spike in response to the pandemic, it is anticipated that it will continue to be an important vehicle for health care.

A Unique Skillset for the Telemedicine Provider

With the imminent growth of telemedicine, teaching providers this unique skillset is essential for the success of future telemedicine programs. Practitioners who have not used telemedicine may find it difficult to translate their traditional interview and physical exam skills to the virtual sphere [6]. Furthermore, technical challenges are cited as one of the most common reasons for failure to implement telemedicine by providers [7]. These barriers are compounded by patient discomfort, possible technological illiteracy, and the stress associated with adapting to a new environment or system [6]. Additionally, a computer screen or phone makes for more detached patient interactions and distant provider-patient relationships [6]. When providers do not have the relevant skills, it becomes harder to concurrently adopt new technologies and maintain high-quality patient care. One study has found that trainees inexperienced with telemedicine struggled with gaining detailed histories or providing appropriate next steps—jumping straight to requesting unnecessary in-person visits [8]. A systematic review has found that courses have been successful in teaching telemedicine skills through dedicated curricula [9]. With the proper resources, these skills can be taught and therefore should become a key component of contemporary medical school curricula.

Telemedicine Curriculum Considerations

Education is critical to the future success of telemedicine implementation. Integration of telemedicine into undergraduate medical education allows for all future physicians to have access to this type of training. The American Medical Association (AMA) has similarly articulated the value of telemedicine curricula in medical schools and residency programs [10]. In addition, the United States Medical Licensing Exam incorporates telephone encounters into its Step 2 Clinical Skills exam [11]. Multiple curricular resources already exist that have identified core competencies for telemedicine for physicians [6,12]. Other health professions have even initiated discussions around establishing competencies associated with virtual care [13].

There are a number of suggested skills associated with optimal telemedicine care. These include but are not limited to the following: communication, physical examination, professionalism, and technological literacy. Communication skills include clear enunciation and the minimization of body motions, gestures, and colloquial speech [12]. These measures help to account for clarity that may otherwise be lost through online platforms [6]. Unique physical exam techniques that can be learned involve functional physical exams [6], application of remote monitoring devices [12], and collaboration with on-site providers. Virtual evaluations can extend to include home assessments, such as in-home mobility barriers [6]. Training for professionalism in a virtual domain can cover education on privacy concerns, electronic prescribing (e-prescribing), and reporting of practices [6,12]. Trainees should also learn technological skills such as screen sharing to discuss diagnostic findings [6], assessing for technological literacy [6], and coaching patients through the use of virtual health platforms [12].

Some programs have already started to institute courses to teach these telemedicine communication and evaluation skills. Among the most common curricular elements includes standardized patients [8,9]. However, creating simulations, training standardized patients, and implementing new technology may prove to be insurmountable barriers for medical schools in the current setting. Instead of recreating new technology solely for medical students, schools can engage students in clinical care through the same telemedicine technology that hospital systems are currently employing. Already, a select number of schools successfully expose students to telemedicine during traditional clinical rotations [9]. With the acute expansion of telemedicine, more schools could provide enriching telemedicine opportunities for their students.

Envisioning Remote Clinical Education During COVID-19

Telemedicine curricula could mimic traditional curricula that medical educators already employ with minimal adjustments, allowing for the continuation of clinical education during the pandemic. Students can easily access and get involved with telemedicine initiatives through technological platforms. For example, some electronic health record (EHR) platforms already integrate third-party platforms that would allow three-person calls [14]. Other options could be to implement Health Insurance Portability and Accountability Act (HIPAA)–compliant group-video options [15], or a three-way call led by the student. Third-party platforms often require institution licenses; however, with the dramatic growth in the last month, many of these platforms have become more readily available and have expanded their services [15]. Future access may require advocacy from medical schools for student licenses, or for clinician educators to share licenses with their students.

Educating students through telehealth initiatives could parallel a majority of the 13 Core Entrustable Professional Activities for entering residency (EPAs) (Table 1). These EPAs were created by the AAMC to provide medical schools with curricular guidelines to ensure a uniform set of skills for all medical school graduates [16]. We identified several curricular activities for schools to integrate now and during the transition back to regular clerkships.

XSL•FO
RenderX

Table 1. Telemedicine curricular activities that align with each of the Association of American Medical Colleges' Core Entrustable Professional Activities (EPAs) [16].

EPA	Telemedicine curricular activity
EPA 1: Gather a history and perform a physical exam	• Clinical e-visit[a] • Virtual consult
EPA 2: Prioritize a differential diagnosis	• Clinical e-visit • Virtual consult • Pathology/radiology cases
EPA 3: Diagnostic and screening tests	• Clinical e-visit • Virtual consult • Pathology/radiology cases
EPA 4: Enter and discuss orders and prescriptions	• Clinical e-visit • Virtual consult
EPA 5: Document a clinical encounter	• Clinical e-visit • Virtual consult • Pathology/radiology cases
EPA 6: Provide an oral presentation of a clinical encounter	• Clinical e-visit • Virtual consult • Pathology/radiology cases • Student-led patient education project
EPA 7: Clinical questions to advance patient care	• Post e-visit reflection • Virtual journal clubs • Student-led inquiry projects
EPA 8: Give or receive a patient hand-off	• Virtual standardized patients and Objective Structured Clinical Examinations [8,9]
EPA 9: Collaborates as a member of an interprofessional team	• Interdisciplinary rounds • Teleconsults [4] • Tumor board [17] • Group discussions with other health-professional schools
EPA 10: Recognize urgent or emergent situations	• Clinical e-visit • Telestroke team [18] • Teletrauma team [19] • COVID-19 call centers, forward triage response team [4]
EPA 11: Obtain informed consent for tests and/or procedures	• Clinical e-visits in surgery, surgical subspecialties, obstetrics and gynecology, etc [20]
EPA 12: General procedures of a physician	• Online procedure courses, augmented and virtual reality simulations, including CPR[b] training and ultrasound techniques [20] • Participation in live-streamed surgical theaters [20]
EPA 13: Identify system failures and contribute to culture of safety and improvement	• Post e-visit reflection • Quality improvement training • Student reflections

[a]e-visit: electronic visit.

[b]CPR: cardiopulmonary resuscitation.

By integrating multiple EPAs, a single e-visit could provide students exposure to skills that they will employ throughout their careers. A telemedicine student visit could be similar to most other ambulatory student clinic visits. To ease patient burden and reduce confusion, ideally the patient, the medical student, and the physician would all be present concurrently at the e-visit. The attending could observe the student lead as much of the visit as possible, with a seamless transition when the attending takes lead of the conversation. Using telehealth technology, students could gather a history and perform a virtual physical exam (EPA 1), focusing on communication skills, rapport building, and functional physical exam maneuvers. Students could summarize their findings (EPA 6) to both the patient and attending in patient-friendly language. Depending

XSL•FO
RenderX

on the comfort level of the patient, the physician could lead the student through a discussion of the possible diagnoses (EPA 2), potential diagnostic tests (EPA 3), or orders (EPA 4). This style would mimic "family-centered rounds," providing the student the opportunity to think through an assessment and plan with direct patient input. Alternatively, the physician and student could discuss the clinical reasoning and key academic concepts after the patient's call. Following the visit, the medical student would then document the encounter (EPA 5). To augment these visits, medical students could research topics inspired by patient cases (EPA 7) and discuss the impacts of telemedicine on health care (EPA 13). These examples show how the e-visit can provide in-depth, high-quality education. A single e-visit can provide the opportunity to teach medical students invaluable skills that they would otherwise not be able to obtain from reading case studies.

With the expansion of novel telemedicine applications into more fields, including procedural fields, students can continue to learn foundational knowledge through engagement in telemedicine even beyond the stereotypical telemedicine visit. In fact, involvement in virtual opportunities exist for all EPAs (Table 1), thus making telemedicine a powerful supplement to the traditional inpatient and outpatient clinical curriculum. In surgery and obstetrics and gynecology, students could assist with pre- and postoperative e-visits (EPA 11) [20]. Students can become educators for patients, by researching a topic such as diabetes nutrition strategies (EPA 7) and then leading a group virtual session with patients (EPA 6). Within the context of inpatient encounters, the aforementioned outpatient recommendations can be further adapted to the inpatient setting. For example, students could participate in virtual consults by talking to admitted patients via a tablet or other similar equipment following the above format (EPAs 1-6). In pathology and radiology, screen-sharing technologies could allow students to become involved in review of slides and imaging remotely (EPAs 2 and 3). Unique experiences such as opportunities to triage urgent cases may exist for hospitals that have telestroke [18] or teletrauma teams (EPA 10) [19,20], or to participate in virtual interdisciplinary rounds (EPA 9) [17]. Students could benefit from participating in as few as one opportunity to augment their online learning. Furthermore, with so many clinicians going online, the use of telemedicine in education is not dependent on a single department. With telemedicine opportunities now available in almost every field, students have the opportunity to learn a variety of skills in multiple specialties.

Even if students are to return to the clinical setting soon, traditional clerkships may look different for them. For example, some students may have continued needs for self-quarantining while others may return to the hospital. For students in hospitals, there will likely still be patients admitted for COVID-19. In this context, telemedicine may provide a way to resume medical student education while maintaining a safe environment. For example, students can participate in consults or rounds without stepping into patient rooms through the use of tablets [4,5]. In fact, this is already a strategy for our institution's burn center during multidisciplinary rounds. As previously stated, even as clerkships return and social distancing measures are relaxed, the high use of telemedicine will likely persist. Therefore, students could continue to participate in these e-visits during their clerkships. The vast array of telemedicine initiatives can be utilized to augment traditional clerkships, providing students with broader access to diverse learning opportunities.

Concluding Thoughts

As telemedicine is further implemented, clinical skills unique to remote care have become necessary. The rapid adoption of telemedicine due to the COVID-19 outbreak can serve as an opportunity to augment medical education curricula and to continue to provide medical students with critical educational opportunities when in-person encounters are limited. If hospital systems are already building and expanding videoconferencing and e-visit tools for providers, why not include students as well?

Authors' Contributions

All authors contributed substantially to the formulation, drafting, and critical revision of this article. All authors have approved the final manuscript.

Conflicts of Interest

AMI is a member of the United States Air Force in the Health Professions Scholarship Program. The views expressed in this material are those of the authors, and do not reflect the official policy or position of the US Government, the Department of Defense, the Department of the Air Force, or the University of Michigan. There are no other disclosures to report for the remaining authors.

References

1. Interim guidance on medical students? Participation in direct patient contact activities: principles and guidelines. Association of American Medical Colleges. 2020 Mar 23. URL: https://www.aamc.org/system/files/2020-03/Interim%20Guidance%20on%20medical%20students%20voluntary%20participation%203.23%20Final.pdf [accessed 2020-03-23]
2. Rose S. Medical Student Education in the Time of COVID-19. JAMA 2020 Mar 31:e. [doi: 10.1001/jama.2020.5227] [Medline: 32232420]
3. Ugalmugale S, Swain R. Telemedicine market size by services (tele-consulting, tele-monitoring, tele-education/training), by type (telehospital, telehome), by specialty (cardiology, gynecology, neurology, orthopedics, dermatology, mental health),

XSL•FO
RenderX

by delivery mode (web/mobile {telephonic, visualized}, call centers), industry analysis report, regional outlook, type potential, competitive market share & forecast, 2019-2025. Global Market Insights. 2019. URL: https://www.gminsights.com/industry-analysis/us-telemedicine-market [accessed 2020-03-30]
4. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
5. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 Pandemic, Lessons for the Future. Telemedicine and e-Health 2020 May 01;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb]
6. DeJong C, Lucey CR, Dudley RA. Incorporating a New Technology While Doing No Harm, Virtually. JAMA 2015 Dec 08;314(22):2351-2352. [doi: 10.1001/jama.2015.13572] [Medline: 26647252]
7. Scott KC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
8. Sartori DJ, Olsen S, Weinshel E, Zabar SR. Preparing trainees for telemedicine: a virtual OSCE pilot. Med Educ 2019 May 11;53(5):517-518. [doi: 10.1111/medu.13851] [Medline: 30859605]
9. Waseh S, Dicker AP. Telemedicine Training in Undergraduate Medical Education: Mixed-Methods Review. JMIR Med Educ 2019 Apr 08;5(1):e12515 [FREE Full text] [doi: 10.2196/12515] [Medline: 30958269]
10. Kelly J. AMA encourages telemedicine training for medical students, residents. American Medical Association. 2016 Jun 15. URL: https://www.ama-assn.org/press-center/press-releases/ama-encourages-telemedicine-training-medical-students-residents [accessed 2020-03-23]
11. STEP 2 Clinical Skills (CS) Content Description and General Information. United States Medical Licensing Examination. 2019 Aug. URL: https://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf [accessed 2020-06-14]
12. Sharma R, Nachum S, Davidson KW, Nochomovitz M. It's not just FaceTime: core competencies for the Medical Virtualist. Int J Emerg Med 2019 Mar 12;12(1):8 [FREE Full text] [doi: 10.1186/s12245-019-0226-y] [Medline: 31179921]
13. van Houwelingen CTM, Moerman AH, Ettema RGA, Kort HSM, Ten Cate O. Competencies required for nursing telehealth activities: A Delphi-study. Nurse Educ Today 2016 Apr;39:50-62 [FREE Full text] [doi: 10.1016/j.nedt.2015.12.025] [Medline: 27006033]
14. Getting started with epic telemedicine integration. The HCI Group. 2020 Jan 28. URL: https://blog.thehcigroup.com/telemedicine-integration-with-mychart-powered-by-epic [accessed 2020-04-07]
15. Siwicki B. A guide to telehealth vendors in the age of COVID-19. Healthcare IT news. 2020 Mar 30. URL: https://www.healthcareitnews.com/news/guide-telehealth-vendors-age-covid-19 [accessed 2020-04-07]
16. Core Entrustable Professional Activities (EPAs) for entering residency. Association of American Medical Colleges. 2014. URL: https://www.aamc.org/what-we-do/mission-areas/medical-education/cbme/core-epas [accessed 2020-03-26]
17. Shea CM, Teal R, Haynes-Maslow L, McIntyre M, Weiner BJ, Wheeler SB, et al. Assessing the feasibility of a virtual tumor board program: a case study. J Healthc Manag 2014;59(3):177-193 [FREE Full text] [Medline: 24988672]
18. Klein BC, Busis NA. COVID-19 is catalyzing the adoption of teleneurology. Neurology 2020 Apr 01;94(21):903-904. [doi: 10.1212/wnl.0000000000009494]
19. Langabeer JR, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-Enabled Emergency Medical Services Program Reduces Ambulance Transport to Urban Emergency Departments. West J Emerg Med 2016 Nov;17(6):713-720 [FREE Full text] [doi: 10.5811/westjem.2016.8.30660] [Medline: 27833678]
20. Huang EY, Knight S, Guetter CR, Davis CH, Moller M, Slama E, et al. Telemedicine and telementoring in the surgical specialties: A narrative review. Am J Surg 2019 Oct;218(4):760-766. [doi: 10.1016/j.amjsurg.2019.07.018] [Medline: 31350010]

Abbreviations

AAMC: Association of American Medical Colleges
AMA: American Medical Association
COVID-19: coronavirus disease
CPR: cardiopulmonary resuscitation
EHR: electronic health record
e-prescribing: electronic prescribing
EPA: Entrustable Professional Activity
e-visit: electronic visit
HIPAA: Health Insurance Portability and Accountability Act

XSL•FO
RenderX

Edited by G Eysenbach; submitted 27.04.20; peer-reviewed by A Dicker, M Singh, S Sabarguna; comments to author 12.06.20; revised version received 22.06.20; accepted 01.07.20; published 20.07.20.

Please cite as:
Iancu AM, Kemp MT, Alam HB
Unmuting Medical Students' Education: Utilizing Telemedicine During the COVID-19 Pandemic and Beyond
J Med Internet Res 2020;22(7):e19667
URL: http://www.jmir.org/2020/7/e19667/
doi:10.2196/19667
PMID:32614776

©Ariella Magen Iancu, Michael Thomas Kemp, Hasan Badre Alam. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Story of #GetMePPE and GetUsPPE.org to Mobilize Health Care Response to COVID-19 : Rapidly Deploying Digital Tools for Better Health Care

Shuhan He[1,2], MD; Ayotomiwa Ojo[3], BSc; Adam L Beckman[3], BSc; Suhas Gondi[3], BA; Megan Ranney[4], MD, MPH; Marian Betz[5], MD, MPH; Jeremy S Faust[6], MD, MSc; Esther Choo[7], MD, MPH; Dara Kass[8], MD; Ali S Raja[2], MD, MBA, MPH

[1]Center for Innovation in Digital HealthCare, Massachusetts General Hospital, Boston, MA, United States
[2]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States
[3]Harvard Medical School, Boston, MA, United States
[4]Department of Emergency Medicine, Warren Alpert Medical School, Brown University, Providence, RI, United States
[5]Department of Emergency Medicine, School of Medicine, University of Colorado, Denver, CO, United States
[6]Department of Emergency Medicine, Brigham and Women's Hospital, Department of Emergency Medicine, Boston, MA, United States
[7]Department of Emergency Medicine, Oregon Health and Science University, Portland, OR, United States
[8]Department of Emergency Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, United States

Corresponding Author:
Shuhan He, MD
Center for Innovation in Digital HealthCare
Massachusetts General Hospital
50 Staniford street
suite 750
Boston, MA, 02114
United States
Phone: 1 8477074695
Email: She8@partners.org

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2023/1/e46028/

Abstract

Physicians, nurses, and other health care providers initiated the #GetMePPE movement on Twitter to spread awareness of the shortage of personal protective equipment (PPE) during the coronavirus disease (COVID-19) pandemic. Dwindling supplies, such as face masks, gowns and goggles, and inadequate production to meet increasing demand have placed health care workers and patients at risk. The momentum of the #GetMePPE Twitter hashtag resulted in the creation of a petition to urge public officials to address the PPE shortage through increased funding and production. Simultaneously, the GetUsPPE.org website was launched through the collaboration of physicians and software engineers to develop a digital platform for the donation, request, and distribution of multi-modal sources of PPE. GetUsPPE.org and #GetMePPE were merged in an attempt to combine public engagement and advocacy on social media with the coordination of PPE donation and distribution. Within 10 days, over 1800 hospitals and PPE suppliers were registered in a database that enabled the rapid coordination and distribution of scarce and in-demand materials. One month after its launch, the organization had distributed hundreds of thousands of PPE items and had built a database of over 6000 PPE requesters. The call for action on social media and the rapid development of this digital tool created a productive channel for the public to contribute to the health care response to COVID-19 in meaningful ways. #GetMePPE and GetUsPPE.org were able to mobilize individuals and organizations outside of the health care system to address the unmet needs of the medical community. The success of GetUsPPE.org demonstrates the potential of digital tools as a platform for larger health care institutions to rapidly address urgent issues in health care. In this paper, we outline this process and discuss key factors determining success.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e20469) doi:10.2196/20469

KEYWORDS

digital health; getusppe; getmeppe; COVID-19; personal protective equipment; protection; Twitter; pandemic; health care worker

Introduction

On March 17, 2020, the Twitter hashtag #GetMePPE went viral [1-3] as thousands of nurses, physicians, and other health care providers posted pictures of themselves with personal protective equipment (PPE). The pictures were accompanied by attestations of dwindling—or in some cases completely absent—stocks of face masks, N95 respirators, and other supplies, even as providers began to evaluate and treat patients with coronavirus disease (COVID-19). The hashtag drew an outpouring of public support and media attention [4] amplifying the PPE shortage (Figure 1); a website, GetMePPE.org, obtained over 62,000 signatures for its online petition for increased government support of, and funding for, PPE [5].

Figure 1. Example tweet from an individual who increased the viewership of #GetMePPE [4].



Simultaneously, a separate group launched GetUsPPE.org [6] with the goal of connecting people who needed PPE with individuals or organizations who had some available. The two groups came together on March 21, with the GetUsPPE.org webpage taking on the #GetMePPE petition as well as the coordination of PPE donation delivery. Only 10 days later, the site had over 120,000 views, the petition had over 10,000 signatures, and over 1800 hospitals and PPE suppliers had entered their information into the continuously updated database. Currently, one month after the merger, a web application with automated attribution and a last-mile system has facilitated the delivery of hundreds of thousands of items of PPE in three trial cities (getusppe.org/data), and partnerships have been formed with the American Medical Association, the American Hospital Association, and a number of specialty societies [7]. The GetUsPPE.org website was rapidly launched to collect available supply and demand requests. Textbox 1 shows a list of the types of facilities that can request PPE on the website.

XSL•FO
RenderX

Textbox 1. Types of health care facility designations for the Request Donations page on the GetUsPPE.org website.

- Acute care hospital
- Freestanding emergency room
- Field hospital
- Hospital overflow facility
- Emergency medical services
- Urgent care clinic
- Nonacute care hospital
- Inpatient psychiatric facility
- Inpatient rehabilitation facility
- Residential substance treatment facility
- Nursing home or skilled nursing facility
- Assisted living facility or group home
- Homeless shelter
- Correctional facility or detention center
- Hospice
- Dialysis center
- Outpatient clinic (primary care or specialist)
- Public health clinic or community health center
- Indian or Tribal healthcare facility (any kind)
- Outpatient rehabilitation facility
- Dental or oral surgery clinic (providing urgent/emergency care)
- Home health agency
- COVID-19 test site
- Ophthalmic provider
- Clinical laboratory
- Pharmacy
- Mortuary or coroner's office
- Social services organization
- Other

Much like just-in-time manufacturing [8], we propose that this type of rapid website deployment will be a solution in the future when urgent health care scenarios arise.

Needs

At the time of writing, 56/60 US jurisdictions (93.3%) have reported cases of COVID-19 infection [9]; most states are experiencing sustained community spread. While analysis of US statistics is ongoing, in analyses from Wuhan, China, 26% of patients received ICU care, and the mortality rate was 4.3% [10]. PPE is critical to protect health care workers at the front lines of caring for these patients. According to US Centers for Disease Control guidelines for use of PPE in US hospitals [11], both N-95 filtering facemask respirators and powered air-purifying respirators should be accompanied by the use of a full-face shield, helmet, or headpiece. Prior experience with the Ebola virus and severe acute respiratory syndrome (SARS) affirm that PPE provides “the last physical barrier” between health care workers and infectious body fluids [12,13].

Despite this, health care providers across the United States are lacking basic PPE, including gowns and masks, as they care for sick patients with COVID-19 [14]. Evidence from past epidemics suggests that clinical workers are at much higher risk than the general population of being infected [15,16]. PPE shortages not only compromise the safety of health care workers but also put patients at risk. As health care providers fall sick, fewer of us are available to care for patients. The lack of PPE also results in higher infection rates within hospitals as staff move from patient to patient without proper protection. In the Wuhan case analysis, hospital-related COVID-19 transmission occurred in 41% of patients [10].

XSL•FO
RenderX

Application of Digital Tools

In less than 12 hours over the course of a weekend, a group of medical students, doctors, software engineers, and volunteers from around the US built and launched a web-based platform, GetUsPPE.org, to enable the public to answer the call for help. On the website, organizations requesting PPE report their name, organization type, and address, followed by their PPE needs (such as PPE types and conditions of acceptance). People or organizations interested in donating PPE provide their contact details and information about their supplies, and they indicate if they are available to deliver or ship them. We were immediately flooded with heartwarming support and offers to help, not only from a variety of existing sources of PPE but also from novel additive manufacturing facilities and homemade makers. More importantly, the site enabled us to collect signatures to push for government action, process PPE donations from a public now keenly aware of the problem, and facilitate distribution of the PPE to health care workers in need. The website now supports global affiliates, including India as a prominent partner [17].

We attribute the success of our digital platform to three key components. First, there was unity around a clear and common cause. Importantly, the effort was not limited to physicians but was extended to all staff impacted by the shortage of PPE. Through firsthand stories of paramedics facing exposure in the field, nurses treating patients who were coughing in their faces, housekeeping staff cleaning rooms without adequate protection, pharmacists assisting with intensive care unit medication dosing without masks, and concerned friends and family, it was clear that our platform was needed for health care workers of all types.

Second, as health care providers in the "trenches," we leveraged social media to share what we were witnessing and give all stakeholders a voice. Using the #GetUsPPE hashtag, physicians shared photographs of themselves without proper PPE, video tutorials teaching each other how to create homemade PPE [18], and images of store shelves fully stocked with boxes of N95 masks [19] while hospitals had none [20]. Stories circulated quickly on Twitter and on other social media platforms, including stories of emergency physicians hospitalized after contracting COVID-19 on the job [21]. Later, the stories became more uplifting, with updates of community members coming together to support health care workers and donate PPE to hospitals and other health care organizations.

Finally, the digital platform itself is intuitive and easy for users to negotiate, and it was not difficult to develop. The website, while initially primitive, was created literally overnight with the aid of the increasingly large number of tools that facilitate building and maintaining websites for the uninitiated; after its quick launch, the site experienced rapid growth. The live launch and traffic catalyzed further development and collaboration with subject matter experts to quickly iterate multimodal technological solutions, including databases (eg, to store information regarding which hospitals are most in need of PPE), matching algorithms (eg, to match PPE donations with the closest hospital in need), and logistics capabilities (eg, to coordinate the manufacturing of PPE and delivery to hospitals).

Lessons

Our experience with #GetMePPE and GetUsPPE.org exemplifies the untapped potential of digital platforms to mobilize and enable a ready and willing public to alleviate some of the largest barriers faced by health care workers. The COVID-19 pandemic has illuminated the deficiencies of our health care supply chain, rendering traditional sources of support inadequate. Fortunately, digital literacy is increasing rapidly worldwide, and the general public is increasingly embracing digital media. However, the medical field has yet to take full advantage of this opportunity.

Social media is just one example of a digital opportunity for health care providers to help increase the quality and efficiency of health care [22]. Current use of social media by physicians is mostly one-way rather than interactive [23]. However, it is critical for health care providers and organizations to create and participate in websites. While social media is essential for raising public awareness and excitement, a website enables health care providers to action this momentum to manifest tangible benefits for themselves and their patients.

Conclusion

GetUsPPE.org is just one example of the successful use of digital tools not only to voice needs, such as those of people working on the front lines of a pandemic, but to mobilize available resources at a speed we have rarely seen before. This initiative has demonstrated the efficacy, reach, and speed of digital platforms; however, it by no means represents the potential the medical community could achieve with full support and buy-in from larger national organizations or the federal government. To date, we have taken a bottom-up, grassroots approach, operating in entrepreneurial fashion, often donating our personal time and resources. It is interesting to imagine what we could accomplish with formal support. In this time of unprecedented crisis and need, we health care professionals have a responsibility to do everything we can to offer the best possible care, including ensuring that we can continue serving our patients without sacrificing either our safety or theirs.

Conflicts of Interest

All authors are volunteers for GetUsPPE, a subsidiary of AFFIRM Research. SH and MLR are directors of GetUsPPE. ALB reports receiving consulting fees from Aledade.

References

XSL•FO
RenderX

1. Gondi S, Beckman AL, Deveau N, Raja AS, Ranney ML, Popkin R, et al. Personal protective equipment needs in the USA during the COVID-19 pandemic. Lancet 2020 May 23;395(10237):e90-e91 [FREE Full text] [doi: 10.1016/S0140-6736(20)31038-2] [Medline: 32416784]
2. Glenza J. The Guardian. 2020 Mar 24. Medical staff beg for masks on social media amid coronavirus crisis URL: https://www.theguardian.com/world/2020/mar/24/coronavrirus-medical-staff-beg-for-masks-social-media [accessed 2020-06-15]
3. @nytimes. 2020 Mar 19. Doctors, nurses and others are rallying on social media with the hashtag #GetMePPE — a reference to personal protective equipment — to put pressure on elected leaders to get them more gear to guard against the coronavirus URL: https://twitter.com/nytimes/status/1240634143266013185?lang=en [accessed 2020-06-06]
4. @ChelseaClinton. 2020 Mar 24. Healthcare workers are on the frontlines of #COVID19 every day & they don't have the protective equipment they need to keep themselves, or us safe. If you are able to support the effort to #GetMePPE or make masks, please visit: http://getusppe.org & http://weneedmasks.org URL: https://twitter.com/ChelseaClinton/status/1242528991485087744 [accessed 2020-06-15]
5. GetUsPPE. GetMePPE Petition URL: https://getusppe.org/getmeppe/ [accessed 2020-06-15]
6. GetUsPPE. URL: https://getusppe.org/ [accessed 2020-06-15]
7. GetUsPPE. Partners URL: https://getusppe.org/partners [accessed 2020-06-15]
8. Tack P, Victor J, Gemmel P, Annemans L. 3D-printing techniques in a medical setting: a systematic literature review. Biomed Eng Online 2016 Oct 21;15(1):115. [doi: 10.1186/s12938-016-0236-4] [Medline: 27769304]
9. US Centers for Disease Control and Prevention. CDC COVID Data Tracker URL: https://www.cdc.gov/covid-data-tracker/ [accessed 2020-06-05]
10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Feb 07:1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
11. US Centers for Disease Control and Prevention. 2019. Strategies to Optimize the Supply of PPE and Equipment URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/ [accessed 2020-04-06]
12. Fischer WA, Weber D, Wohl DA. Personal Protective Equipment: Protecting Health Care Providers in an Ebola Outbreak. Clin Ther 2015 Nov 01;37(11):2402-2410 [FREE Full text] [doi: 10.1016/j.clinthera.2015.07.007] [Medline: 26452427]
13. Weber DJ, Rutala WA, Schaffner W. Lessons learned: Protection of healthcare workers from infectious disease risks. Crit Care Med 2010 Aug;38:S306-S314. [doi: 10.1097/ccm.0b013e3181e69ebd]
14. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
15. Chan-Yeung M. Severe acute respiratory syndrome (SARS) and healthcare workers. Int J Occup Environ Health 2004;10(4):421-427. [doi: 10.1179/oeh.2004.10.4.421] [Medline: 15702757]
16. Beckman AL, Gondi S, Forman HP. Health Affairs. 2020 Mar 18. How To Stand Behind Frontline Health Care Workers Fighting Coronavirus URL: https://www.healthaffairs.org/do/10.1377/hblog20200316.393860/full/ [accessed 2020-03-18]
17. GetUsPPE. GetUsPPE India URL: https://in.getusppe.org/ [accessed 2020-06-06]
18. @mystifide. 2020 Mar 23. It is terrible to say this, but if it comes to using sewn masks, might as well make one with extra filters. #covid4MDs #GetUsPPE URL: https://twitter.com/mystifyde/status/1242130913254760448?s=20 [accessed 2020-03-30]
19. @drdanchoi. 2020 Mar 21. Wow @Target please set an example and redirect your ENTIRE national stock of N95 masks to hospitals for them to purchase Or if you want to generate some great PR, donate them! URL: https://twitter.com/drdanchoi/status/1241562744807133185?s=20 [accessed 2020-03-30]
20. @choo_ek. 2020 Mar 20. From a friend in the emergency department at a major academic hospital. #GetMePPE #GetUsPPE URL: https://twitter.com/choo_ek/status/1240909433032421376?s=20 [accessed 2020-03-30]
21. American College of Emergency Physicians. 2020 Mar 14. Two Emergency Physicians in Critical Condition URL: https://www.acep.org/corona/covid-19-articles/a-statement-from-acep-president-william-jaquis-md-facep [accessed 2020-03-30]
22. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
23. Campbell L, Evans Y, Pumper M, Moreno MA. Social media use by physicians: a qualitative study of the new frontier of medicine. BMC Med Inform Decis Mak 2016 Jul 15;16:91 [FREE Full text] [doi: 10.1186/s12911-016-0327-y] [Medline: 27418201]

Abbreviations

COVID-19: coronavirus disease
PPE: personal protective equipment
SARS: severe acute respiratory syndrome

XSL•FO
RenderX

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by P Chai, M Alvarez de Mon; comments to author 03.06.20; revised version received 04.06.20; accepted 09.06.20; published 20.07.20.

Please cite as:
He S, Ojo A, Beckman AL, Gondi S, Ranney M, Betz M, Faust JS, Choo E, Kass D, Raja AS
The Story of #GetMePPE and GetUsPPE.org to Mobilize Health Care Response to COVID-19 : Rapidly Deploying Digital Tools for Better Health Care
J Med Internet Res 2020;22(7):e20469
URL: https://www.jmir.org/2020/7/e20469
doi:10.2196/20469
PMID:32530813

©Shuhan He, Ayotomiwa Ojo, Adam L Beckman, Suhas Gondi, Megan Ranney, Marian Betz, Jeremy S Faust, Esther Choo, Dara Kass, Ali S Raja. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges

Huailiang Wu[1,2*], MBBS; Weiwei Sun[1,2*], MBBS; Xinyu Huang[1,2*], MBBS; Shinning Yu[2], MBBS; Hao Wang[3], BBA; Xiaoyu Bi[4], BA; Jie Sheng[4], BA; Sihan Chen[1], BA; Babatunde Akinwunmi[5,6], MD, MMSc; Casper J P Zhang[7], MPH, PhD; Wai-Kit Ming[1,2], MD, PhD, MPH

[1]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[2]Faculty of Medicine, International School, Jinan University, Guangzhou, China

[3]Department of Culture and Tourism, Xiamen University of Technology, Xiamen, China

[4]College of Economics, Jinan University, Guangzhou, China

[5]Maternal-Fetal Medicine Unit, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[6]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, MA, United States

[7]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, PhD, MPH
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
No 602 Huangpu Avenue West, Tianhe District
Guangzhou, 510000
China
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Abstract

People across the world have been greatly affected by the ongoing coronavirus disease (COVID-19) pandemic. The high infection risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in hospitals is particularly problematic for recently delivered mothers and currently pregnant women who require professional antenatal care. Online antenatal care would be a preferable alternative for these women since it can provide pregnancy-related information and remote clinic consultations. In addition, online antenatal care may help to provide relatively economical medical services and diminish health care inequality due to its convenience and cost-effectiveness, especially in developing countries or regions. However, some pregnant women will doubt the reliability of such online information. Therefore, it is important to ensure the quality and safety of online services and establish a stable, mutual trust between the pregnant women, the obstetric care providers and the technology vis-a-vis the online programs. Here, we report how the COVID-19 pandemic brings not only opportunities for the development and popularization of online antenatal care programs but also challenges.

(J Med Internet Res 2020;22(7):e19916) doi:10.2196/19916

KEYWORDS

coronavirus disease; COVID-19; SARS-CoV-2; online prenatal education; pregnancy; online education; antenatal; telehealth; convenience; inequality

Introduction

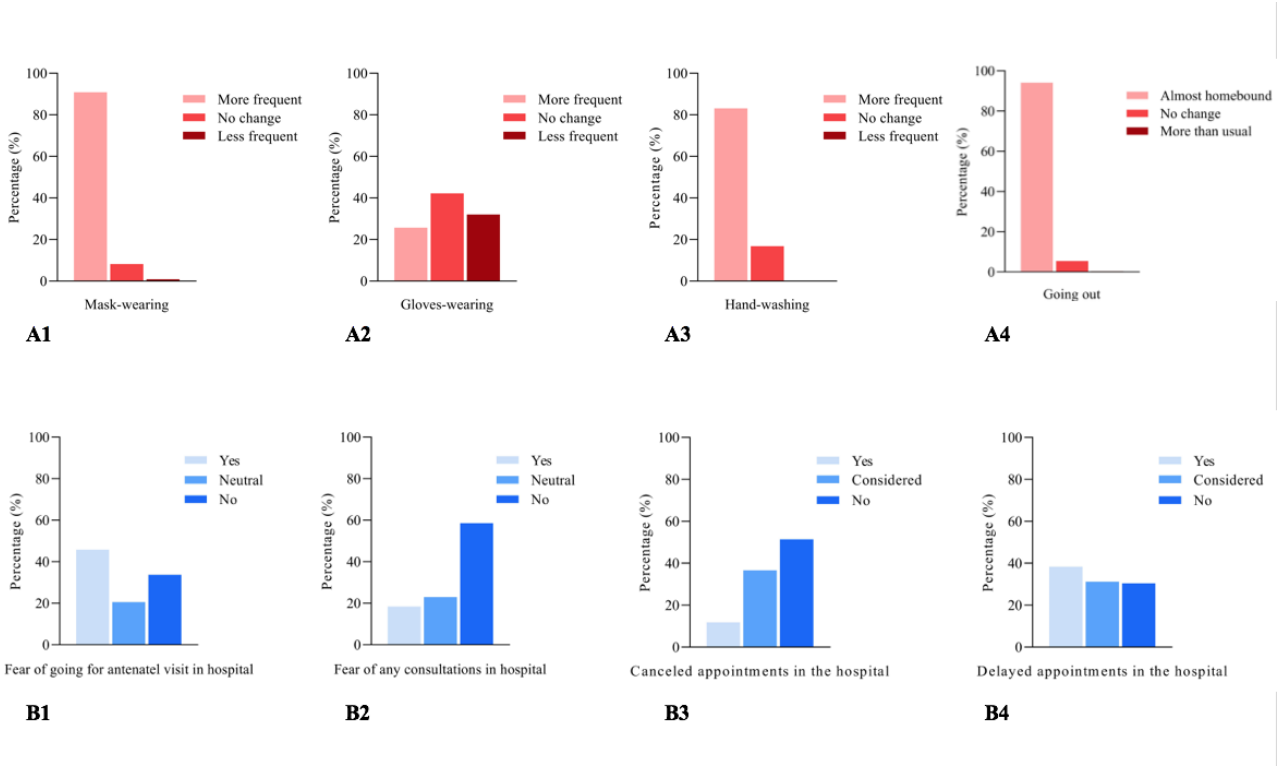
The coronavirus disease (COVID-19) outbreak has spread globally and caused a pandemic that has led to almost 10,000,000 diagnosed cases and 500,000 deaths as of June 28, 2020 [1]. The impact of COVID-19 can be greater in vulnerable populations. Pregnant women, for example, tend to be more physiologically and psychologically susceptible to infectious diseases, putting them at higher risk of maternal complications such as preterm birth, gestational hypertension, gestational diabetes, and miscarriage [2-4]. Appropriate antenatal education can be beneficial to them in many ways including reducing cesarean section rates, maternal and infant mortality, and anxiety

XSL•FO
RenderX

problems as well as improve their general reproductive health outcomes [5]. In the context of the pandemic, pregnant women face an additional dilemma—they need professional antenatal care, there is a potential risk for cross-infection if they choose to visit a hospital to receive this service [6].

Recently, we performed a web-based survey among Chinese pregnant women via a national online platform (Banmi National Online Maternity School) to investigate their self-protection behaviors and attitudes toward antenatal care during the pandemic. A total of 983 Chinese pregnant women completed the questionnaire, and it was found that more than 80% had taken self-protection actions, such as wearing a face mask, handwashing, and home quarantine, to avoid being infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Figure 1). Our findings indicate that about 20% of respondents were afraid of any type of consultation at a hospital, while over 40% feared in-hospital antenatal visits. Moreover, more than half considered or decided to cancel their in-hospital antenatal care visits and postponed their appointments. These behaviors and attitudes indicate that pregnant women were anxious and worried about potential infection in the hospital environment. Considering the dilemma mentioned above and the fear of some other unknowns from hospital visits, online antenatal care might be a preferable choice for pregnant women during this pandemic [7].

Figure 1. Pregnant women's self-protection behaviors and attitudes toward antenatal care in hospitals.



Online Antenatal Care Programs Before the COVID-19 Pandemic

Online antenatal care programs can take various forms: online courses to deliver pregnancy-related information, mobile phone apps to motivate healthy behaviors during pregnancy, and mobile health (mHealth) apps to provide mental consultations [8,9]. Numerous studies have reported that a great majority of pregnant women search for pregnancy-related information on the internet [10,11]. However, the popularity of online antenatal care programs remains low.

Opportunities for Online Antenatal Care During the Pandemic

In the midst of the COVID-19 pandemic, it is advisable that pregnant women should stay at home and receive necessary antenatal care via online antenatal care programs. For example, online antenatal care programs may guide pregnant women to perform blood and urine glucose tests at regular intervals, especially for those with or at high risk of gestational diabetes mellitus (GDM). The obstetric doctor can then closely monitor patients' glucose levels and provide appropriate dietary suggestions and medications. This would be helpful in decreasing the incidence and negative impact of GDM. Moreover, women can upload their daily blood pressure and simple home urine dipsticks results to an online system. This would allow monitoring for serious pregnancy complications such as preeclampsia. General gestational education and mental health consultation can be done through an online education program via voice or video calls. Pregnant women can also be guided to study pregnancy and labor instructions through books, and receive antenatal care education by online conferencing. All of these approaches would likely contribute to better labor preparedness, maternal experience, postpartum adjustment and outcomes, fewer infection risks, and more economic benefits to the health system and the women themselves. Therefore, pregnant women without any serious issues should use online antenatal care programs as an alternative to routine antenatal

XSL•FO
RenderX

care in a hospital at least to some extent. Learning to self-monitor while in lockdown or home quarantine is essential to prevent viral infection or spread [12]; this is especially important to pregnant women as they are more likely to have severe complications if infected with SARS-CoV-2 [13-15].

During the pandemic, daily necessities were scarce and the cost of medical appointments was substantially higher than usual. Many families lost employment and suffered a heavy economic burden. People who were unable to obtain or pay for sufficient medical resources were most vulnerable [16]. Online antenatal care can provide relatively less expensive medical services and diminish inequalities in health care due to its convenience and cost-effectiveness, especially in developing countries or regions. This could contribute to a reduction in medical resources inequality and help pregnant women resolve various health problems during the pandemic. Medical experts and institutions should strengthen the quality of online antenatal care in terms of skilled professional services, technology, and availability.

Challenges of Online Antenatal Care During the Pandemic

There are some challenges and areas for improvement associated with online antenatal care. A study in China found that many pregnant women had concerns about the reliability of online gestational information [17]. Therefore, it is important to establish close collaboration between hospitals and professional institutes to improve the quality of online programs, ensuring the reliability of their information. In the meantime, this can be combined with information technology products, such as using electronic devices with remote monitoring functions to monitor basic indicators such as fetal heart rate and movement, to serve as a proxy for some routine obstetric examinations. However, in certain circumstances, antenatal care in hospitals is irreplaceable (eg, high-risk pregnant women with or at risk of vaginal bleeding, abdominal pain, or other serious discomforts) [18]. In such cases, further obstetric examinations and consultations in a hospital are essential. Moreover, some specific antenatal examinations such as the Nuchal Translucency test, Down syndrome screening, and the Oral Glucose Tolerance Test should be completed in a hospital. The future of maternal care is likely to include technological innovations to address the above challenges. Due to this, more extensive, optimized maternal care services should be applied to reduce overall maternal morbidity and mortality if online antenatal care are to be further developed, popularized, and adopted as an alternative path to health care services for pregnant women.

Furthermore, it has been mentioned that less than one-third of the population in Africa and the Middle East use the internet, with a global usage rate of 51% in 2018 [19]. Therefore, in addition to improving online antenatal care, the popularization of both the use of the internet and mobile electronic devices is crucial to allow more pregnant women to receive online education and care. Network operators need to be improved to guarantee the widespread use of internet services during the COVID-19 outbreak. Governments should use their financial budget to support the popularization of modern electronic devices and internet service, which are essential for online antenatal care programs. Additionally, governments need to legislate relevant laws to regulate and protect the privacy of pregnant women when using online antenatal care services.

Conclusion

In summary, online antenatal care can be a useful, alternative option for pregnant women in need of basic antenatal care and mental health consultation. It can reduce unnecessary hospital visits and limit potential risks of infection among this vulnerable group during the COVID-19 pandemic. Efforts to implement online care is likely to result in multiple innovations and revolutionize antenatal care services both in China and globally. This will contribute to reducing maternal morbidity and mortality by providing opportunities for wider coverage. The popularization of online antenatal care programs is likely to have an economic benefit to both the health care system and to women in terms of cost, time, and manpower. This can improve overall maternal and reproductive health services and family life.

Authors' Contributions

HW, WS, and XH contributed equally to this work. HW, WS, and XH contributed to the conception and design of this viewpoint, drafted the primary version of manuscript, edited the manuscript, and reviewed the final version. SY, HW, XB, JS, SC, BA, and CJPZ edited and reviewed the final version of manuscript. WM contributed to the conception and design of this viewpoint, edited the manuscript, and reviewed the final version. All authors discussed and agreed on the implications of the study findings and approved the final version to be published.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) Situation Report - 160. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200628-covid-19-sitrep-160.pdf?sfvrsn=2fe1c658_2 [accessed 2020-06-29]

XSL•FO
RenderX

2. Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. J Matern Fetal Neonatal Med 2020 May 18:1-7. [doi: 10.1080/14767058.2020.1763946] [Medline: 32419558]
3. Dashraath P, Wong JLJ, Lim MXK, Lim LM, Li S, Biswas A, et al. Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. Am J Obstet Gynecol 2020 Jun;222(6):521-531 [FREE Full text] [doi: 10.1016/j.ajog.2020.03.021] [Medline: 32217113]
4. Murphy HR. Managing Diabetes in Pregnancy Before, During, and After COVID-19. Diabetes Technol Ther 2020 Jun 01;22(6):454-461. [doi: 10.1089/dia.2020.0223] [Medline: 32396397]
5. Ferguson S, Davis D, Browne J. Does antenatal education affect labour and birth? A structured review of the literature. Women Birth 2013 Mar;26(1):e5-e8 [FREE Full text] [doi: 10.1016/j.wombi.2012.09.003] [Medline: 23063931]
6. Ho PL, Tang XP, Seto WH. SARS: hospital infection control and admission strategies. Respirology 2003 Nov;8 Suppl(s1):S41-S45 [FREE Full text] [doi: 10.1046/j.1440-1843.2003.00523.x] [Medline: 15018133]
7. Du L, Gu Y, Cui M, Li WX, Wang J, Zhu LP, et al. [Investigation on demands for antenatal care services among 2 002 pregnant women during the epidemic of COVID-19 in Shanghai]. Zhonghua Fu Chan Ke Za Zhi 2020 Mar 25;55(3):160-165. [doi: 10.3760/cma.j.cn112141-20200218-00112] [Medline: 32268713]
8. Knight-Agarwal C, Davis DL, Williams L, Davey R, Cox R, Clarke A. Development and Pilot Testing of the Eating4two Mobile Phone App to Monitor Gestational Weight Gain. JMIR Mhealth Uhealth 2015 Jun 05;3(2):e44 [FREE Full text] [doi: 10.2196/mhealth.4071] [Medline: 26048313]
9. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Nov 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
10. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. Midwifery 2013 Jul;29(7):730-735. [doi: 10.1016/j.midw.2012.07.003] [Medline: 22958935]
11. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
12. National Health Commission. Notice on strengthening maternal disease treatment and safe midwifery during the prevention and control of new coronavirus pneumonia (in Chinese). People's Republic of China. URL: http://www.nhc.gov.cn/fys/s3581/202002/4f80657b346e4d6ba76e2cfc3888c630.shtml [accessed 2020-02-08]
13. Coronavirus infection and pregnancy: Information for pregnant women and their families. Royal College of Obstetricians and Gynaecologists. URL: https://www.rcog.org.uk/en/guidelines-research-services/guidelines/coronavirus-pregnancy/covid-19-virus-infection-and-pregnancy/ [accessed 2020-04-12]
14. People Who Need Extra Precautions: Pregnancy and Breastfeeding. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/pregnancy-breastfeeding.html [accessed 2020-04-12]
15. Online free consultation, online pregnant women's school, scheduled appointments for consultations during the epidemic prevention and control period in Henan to ensure the safety of mothers and infants (in Chinese). National Working Committee on Children and Women under State Council. URL: http://www.nwccw.gov.cn/2020-02/11/content_278883.htm [accessed 2020-02-11]
16. Ahmed F, Ahmed N, Pissarides C, Stiglitz J. Why inequality could spread COVID-19. The Lancet Public Health 2020 May;5(5):e240 [FREE Full text] [doi: 10.1016/s2468-2667(20)30085-2]
17. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
18. N/A. Covid-19 and pregnancy. BMJ 2020 May 04;369:m1672. [doi: 10.1136/bmj.m1672] [Medline: 32366505]
19. Meeker M. Internet Trends 2019. Bondcap. URL: https://www.bondcap.com/report/itr19/ [accessed 2020-05-05]

Abbreviations

COVID-19: coronavirus disease
GDM: gestational diabetes mellitus
mHealth: mobile health
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

XSL•FO
RenderX

Edited by G Eysenbach, G Fagherazzi; submitted 10.05.20; peer-reviewed by D Groselj, A Al-Hasan; comments to author 22.06.20; revised version received 02.07.20; accepted 13.07.20; published 22.07.20.

Please cite as:
Wu H, Sun W, Huang X, Yu S, Wang H, Bi X, Sheng J, Chen S, Akinwunmi B, Zhang CJP, Ming WK
Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges
J Med Internet Res 2020;22(7):e19916
URL: http://www.jmir.org/2020/7/e19916/
doi:10.2196/19916
PMID:32658860

©Huailiang Wu, Weiwei Sun, Xinyu Huang, Shinning Yu, Hao Wang, Xiaoyu Bi, Jie Sheng, Sihan Chen, Babatunde Akinwunmi, Casper J P Zhang, Wai-Kit Ming. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

An Integrative Model for the Effectiveness of Biofeedback Interventions for Anxiety Regulation: Viewpoint

Joanneke Weerdmeester[1], MSc; Marieke MJW van Rooij[1], IR, PhD; Rutger CME Engels[2], PhD; Isabela Granic[1], PhD

[1]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

[2]Erasmus University Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Joanneke Weerdmeester, MSc
Behavioural Science Institute
Radboud University
Montessorilaan 3
Nijmegen, 6525 HR
Netherlands
Phone: 31 243612076
Email: j.weerdmeester@psych.ru.nl

Abstract

Biofeedback has shown to be a promising tool for the treatment of anxiety; however, several theoretical as well as practical limitations have prevented widespread adaptation until now. With current technological advances and the increasing interest in the use of self-monitoring technology to improve mental health, we argue that this is an ideal time to launch a new wave of biofeedback training. In this viewpoint paper, we reflect on the current state of biofeedback training, including the more traditional techniques and mechanisms that have been thought to explain the effectiveness of biofeedback such as the integration of operant learning and meditation techniques, and the changes in interoceptive awareness and physiology. Subsequently, we propose an integrative model that includes a set of cognitive appraisals as potential determinants of adaptive trajectories within biofeedback training such as growth mindset, self-efficacy, locus of control, and threat-challenge appraisals. Finally, we present a set of detailed guidelines based on the integration of our model with the mechanics and mechanisms offered by emerging interactive technology to encourage a new phase of research and implementation using biofeedback. There is a great deal of promise for future biofeedback interventions that harness the power of wearables and video games, and that adopt a user-centered approach to help people regulate their anxiety in a way that feels engaging, personal, and meaningful.

(J Med Internet Res 2020;22(7):e14958) doi:10.2196/14958

KEYWORDS

biofeedback; neurofeedback; anxiety; appraisal; mechanisms; mental health; eHealth; video games; wearable technology; review; mobile phone

Introduction

Background

The era of the quantified self is upon us [1]. Anywhere and at any time, we have access to an abundance of information about how our bodies are adapting to the world around us, be it through the internet, apps on our smartphones, or wearable devices. With these current technological advances, we are able to track, monitor, and regulate various aspects of our physiological and behavioral activation, ranging from heart rate to sleep patterns, to number of steps taken, and calories burned. The popularity of these devices and apps is increasing, and they are becoming increasingly simple to integrate into our daily lives, providing us with the means to take on a more active role in the management of our health and well-being [2-9].

Monitoring our physiological state informs us about not only our physical health but also our psychological well-being, as our physiology is intimately linked with our psychology [10-12]. In particular, our ability to monitor and modulate our emotional arousal, feelings, and expressions—our capacity for emotion regulation—is an integral part of mental health [13,14]. One fundamental aspect of emotion regulation is interoceptive awareness, which is the ability to sense and interpret internal physiological signals [15-18]. Furthermore, physiological dysregulation underpins breakdowns in mental health and psychopathology (eg, anxiety disorders) [19-24]. For some diagnoses such as generalized anxiety disorder and posttraumatic

XSL•FO
RenderX

stress disorder, physiological symptoms and dysregulation also serve as important inclusion criteria for diagnosis [19]. However, the majority of current interventions for anxiety and related disorders mainly focus on cognition and behavior, with physiology being seemingly undervalued.

The Promise of Biofeedback

Together with psychopharmacology, cognitive behavioral therapy (CBT) is currently viewed as the gold standard for the treatment of anxiety [25-27]. Although some components of CBT involve attending to the physiological signals of anxiety, such as self-monitoring and relaxation training, the core of CBT programs is aimed at modifying the maladaptive cognitive and behavioral components of anxiety instead [28-30]. This is an important limitation, as somatic symptoms are strongly related to the severity and level of impairment associated with anxiety [22,31,32]. Therefore, integrating training that facilitates the awareness, monitoring, and regulation of physiological signals might strengthen the effectiveness of anxiety treatment.

In recent years, there has been a rising interest in incorporating self-monitoring technology in the field of medicine and treatment of mental health [7,9,33,34]. In addition to fostering self-regulation, this type of technology also aligns with the current treatment needs as individuals seem to increasingly gravitate toward holistic forms of therapy that emphasize the mind-body connection and allow them to take on an active role in managing their health [7,35]. One promising form of self-monitoring technology is biofeedback. Biofeedback is the process of measuring an individual's physiological states and feeding that information back to them so that they can learn how to change their physiological activity for the purpose of health improvement [35]. Biofeedback has an extensive history in science [36] and has been shown to be efficacious as a treatment for a large variety of physical [37-39] as well as mental health issues, including stress and anxiety [40-43]. Surprisingly, however, biofeedback has not yet been widely implemented in standard anxiety treatment. Instead, it is generally regarded as an alternative form of treatment [25]. The reported use of alternative treatment is fairly high among people with anxiety, with the most recent numbers indicating that 1 out of 6 (16%) patients with anxiety received alternative care in addition to conventional care [44]. Earlier studies reported numbers ranging between 110 out of 193 (56.7%) [45] and 430 out of 1004 (43%) [26]. However, the available estimate for the specific use of biofeedback among people with anxiety seems fairly small, specifically 3 out of 193 (1.6%) [45].

Part of the reason why biofeedback has not yet been incorporated in standard treatment programs is our general lack of understanding of the mechanisms by which biofeedback training actually works. Although some factors that can explain the effectiveness of biofeedback have been proposed, there is no existing model that encompasses the possible mediating or moderating factors that contribute to positive outcomes of biofeedback training and what should be implemented to strengthen its effects. Understanding the processes that lead to improvement is essential in intervention research, even though they are often overlooked [46]. In addition, practical limitations including hardware requirements, costs, time commitment, and lack of engagement may have prevented wide adaptation [47,48]. However, there have been some promising advancements in recent years that may address some of these limitations, such as the development of biofeedback-based video games, which may increase accessibility and engagement, especially for youth [49-57]. With these current technological advancements in biofeedback training and the accompanying surge in self-monitoring apps, we propose that biofeedback could have a second run in the treatment of anxiety.

Objective

This viewpoint paper reflects on the current state of biofeedback training for the treatment of anxiety. The theoretical and scientific background regarding the working mechanisms of biofeedback are reviewed. In addition, several key gaps in the empirical literature are highlighted, and reasons for these oversights are addressed. Subsequently, we propose an integrative theoretical model that combines factors that have been traditionally linked to the effectiveness of biofeedback with cognitive appraisals as determinants of the effectiveness of biofeedback training. Finally, a set of guidelines will be presented for future research and the design and implementation of a new wave of biofeedback training. These guidelines will be based on the integration of our theoretical model with the mechanics and mechanisms offered by emerging interactive technologies such as video games and wearables. Using these guidelines, we aim to encourage a new phase of research and implementation using biofeedback.

Techniques Used in Biofeedback Training

Operant Learning

One core technique that is integrated in biofeedback training is operant learning. In biofeedback training, feedback is given about physiological changes that occur. The complexity of biofeedback can vary from presenting raw signals of physiological activity (eg, heart rate variability [HRV], respiration, and electroencephalography [EEG] signals) to a moving frequency analysis [58]. In clinical and educational settings, additional aids are often provided by presenting feedback in the form of graphs, images, or sounds [59]. However, simply showing individuals their activity is insufficient. Instead, it is vital that individuals are actively taught, using operant (or instrumental) learning, how to change their physiological state, specifically by providing real-time feedback and reinforcement while regulation attempts are made [60-64].

In operant learning, positive reinforcement is used to shape an individual's behavior by strengthening adaptive behavior through the use of rewards, which makes the original behavior more likely to occur [65]. In a typical biofeedback paradigm, an individual is placed in front of a computer screen on which feedback regarding their physiological state is presented. When the appropriate activity increases or inappropriate activity decreases, this change is followed by a pleasant response (eg, a pleasant tone). As sessions are repeated, the thresholds for receiving a reward are gradually modified, thereby stimulating the display of healthy physiological activity [66,67].

XSL•FO
RenderX

Meditation Techniques

In addition to operant learning techniques, biofeedback training often incorporates self-regulation techniques that are similar to meditation and mindfulness practices. Meditation is a practice consisting of various exercises that are meant to shift one's attention primarily to internal stimuli to achieve better well-being and emotional balance [31,68]. Mindfulness meditation likewise aims to shift one's attention to present experiences, but puts an additional emphasis on regarding experiences in a nonjudgmental and accepting manner [69,70].

Furthermore, similar mechanisms of change can be identified in both biofeedback and meditation. For instance, similar physiological patterns of change are found in response to biofeedback as well as mediation [71], such as a coherent cardiac rhythm [72,73]. Another similarity is that both practices encourage individuals to shift their attention to internal experiences. In biofeedback, this is achieved by using biosensors to provide individuals with feedback on changes in their physiological activity, whereas in mindfulness training, this is achieved by guided mediation exercises where individuals are prompted to shift their attention to the present moment and present experiences [69,70,74]. Both approaches also often train individuals to use slow, diaphragmatic breathing [58,75-79]. However, although both approaches try to enhance individuals' internal attention, biofeedback specifically focuses on changes in physiological activity, whereas meditation aims for a broader focus on the present moment, including all present experiences. Furthermore, mindfulness meditation explicitly encourages individuals to regard observed sensations in a nonjudgmental manner [69,70], whereas biofeedback training does not. The core difference between meditation and biofeedback, however, is that meditation training does not provide feedback on how the person is doing in terms of regulating their stress. In contrast, biofeedback training gives participants continuous feedback on how they are doing in terms of physiologically regulating their stress. Thus, although there are various similarities between mediation practices and biofeedback, we argue that the feedback provided in biofeedback training is crucial to help people gauge their progress and to keep them engaged in the training.

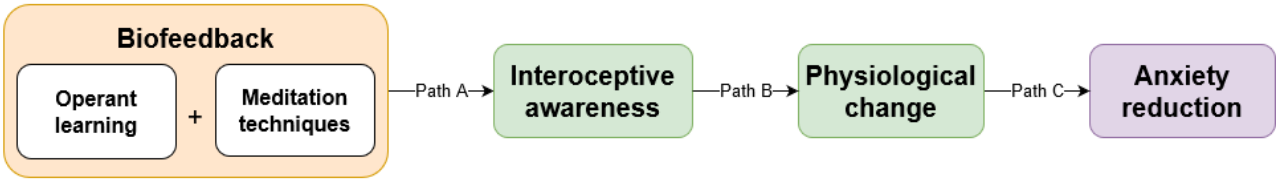
Working Mechanisms in Biofeedback Training

Interoceptive Awareness

Various mechanisms have been proposed to explain the effectiveness of biofeedback (Figure 1). One of the core assumptions is that biofeedback leads to increases in interoceptive awareness, which in turn helps individuals to better regulate their physiology (Figure 1, paths A and B) [80-82]. Interoceptive awareness is our ability to sense and interpret internal and physiological signals, which is an important part of emotion regulation [13,15,36,83,84]. Scientific research on interoceptive awareness started in the field of psychophysiology back in the 70s and 80s, followed by a new wave of interest in the 90s sparked by the introduction of the somatic marker hypothesis [36,83]. According to this hypothesis, the interpretation of bodily sensations is closely associated with emotional processing and decision making [10,85]. These early days of interoceptive awareness research also gave rise to the idea that biofeedback training could help individuals improve their interoceptive awareness [36], and this assumption is still at the core of biofeedback research today [35,38,41,43,47,59,60,86,87]. The support for this claim, however, seems to be mostly theoretical in nature, as there is very little direct empirical evidence showing that biofeedback training systematically improves awareness of physiological states (Figure 1, path A). Equally important is that it is unclear if and when this internal acuity is adaptive [36,88,89].

Interoceptive awareness has been positively related to decision making [90,91] as well as various forms of emotional processing, such as emotional memory [92], emotion recognition [93], and emotion regulation [15-18]. Importantly, it seems that increased interoceptive awareness is not always better. For example, heightened interoceptive awareness has been linked to increases in anxiety [94,95] and has also been linked to various anxiety disorders such as panic disorder and social anxiety disorder [89,96-99]. In individuals with panic disorder, an increased focus on and the subsequent misinterpretation of physiological sensations is the main cause of panic attacks [96,100]. For individuals with social anxiety disorder, this vigilant attention to physiological sensations likewise impairs their ability to process information from their immediate social environment. Socially anxious individuals often interpret bodily sensations of anxiety as a confirmation that they cannot function well in social situations. Furthermore, these anxious individuals believe that their bodily sensations are clearly noticeable by others, which plays into their fear of being humiliated [89,97,101,102]. Thus, whether or not high levels of interoceptive awareness are adaptive seems to, at least in part, rely on how internal sensations are interpreted.

Figure 1. Traditional techniques and mechanisms thought to explain the effectiveness of biofeedback training.



XSL•FO
RenderX

Physiological Changes

Biofeedback training teaches individuals how to regulate their physiological activity [35], and the benefit of biofeedback is often attributed to the acquisition of these self-regulation skills. This belief is rooted in the large body of research dedicated to linking specific emotional states to specific physiological changes [36]. The strong physiological ties of anxiety make it a particularly fitting target for biofeedback, as biofeedback training addresses a wide array of physiological processes. The most prominent forms of biofeedback that are used for anxiety treatment and stress management are HRV biofeedback and neurofeedback.

Heart Rate Variability Biofeedback

The primary aim of HRV biofeedback is to create a balance in the autonomic nervous system (ANS). The ANS is divided into 2 branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS prepares individuals for action by increasing their heart rate, blood pressure, and cardiac output, whereas the PNS slows these processes as part of the restorative process [103,104]. Anxiety has mostly been tied to hyperarousal of the SNS, which is associated with physiological responses such as increases in breathing rate and intensity, heightened blood pressure, increases in electrodermal activity, increases in heart rate, and decreases in HRV [42,105]. However, anxiety can also be tied to disrupted PNS function, as is reflected in the low vagal tone that is exhibited by individuals with anxiety disorders [105-109]. Vagal tone is the inhibitory action exerted by the vagus nerve. Specifically, an increase in vagal tone lowers the heart rate and a decrease in vagal tone elevates the heart rate [108]. Vagal tone is used as a marker for emotion regulation, with a higher vagal tone reflecting effective emotion regulation and a lower vagal tone reflecting emotion regulation deficits [110-112].

As the name indicates, HRV biofeedback targets the variability in time (R-R) intervals between heart beats [113,114]. HRV is an important marker for emotion regulation as it reflects the interplay between the PNS and the SNS [110,114-117]. HRV has shown to be particularly sensitive to changes in emotional state, with positive and negative emotions often being readily distinguishable [72]. During negative emotions such as anxiety, there is less synchronization between the PNS and the SNS [118]. In particular, low HRV has been linked to a lack of emotional flexibility [119,120] and has been found to be present in depression [121,122] and anxiety disorders [105,106,109,123].

The primary aim of HRV biofeedback is to create autonomic balance by enhancing cardiac coherence, which is a heart rhythm oscillating at a frequency of approximately 0.1 Hz [39,58,72]. In common HRV biofeedback paradigms, this rhythm is achieved by training individuals to increase the amplitude of respiratory sinus arrhythmia (RSA) [39,58]. RSA is the cyclical change in heart rate occurring when the heart rate increases during inhalation and decreases during exhalation. RSA is commonly used as an index of vagal control of the heart, particularly parasympathetic control [124]. Low amplitudes of RSA are found in anxious and depressed individuals [125,126], and HRV biofeedback aims to increase these amplitudes [58]. In HRV biofeedback training, participants are often shown raw signals of their HRV or a moving frequency analysis. Participants are then instructed to pace their breathing to around 6 breaths per minute to maximize their RSA amplitude [58,75].

Neurofeedback

In addition to heart function, emotion regulation deficits in anxious individuals have also been linked to irregularities in electrical activity in various brain regions such as the amygdala and the prefrontal cortex [127,128]. Neurofeedback uses a brain-computer interface to provide feedback to an individual about their brain's electrical activity, which is measured by EEG electrodes placed on the scalp. The aim of neurofeedback training is to maintain the level of brain activity within a specified frequency range (eg, Theta 4-8 Hz) [61,129,130]. Effective neurofeedback training protocols for the treatment of anxiety have mostly focused on enhancing alpha, increasing higher theta activity, and inhibiting beta frequencies [66,67,131,132]. However, quantitative EEG is increasingly used to identify additional patterns of brain activity related to anxiety and to improve neurotherapy (for a detailed overview, refer to the study by Price and Budzynski [132]).

Limitations and Gaps in Biofeedback Research

Despite the fact that teaching individuals to change their physiology is central to both HRV biofeedback as well as neurofeedback, there is no convincing evidence as to whether these changes are indeed essential for positive treatment outcomes (Figure 1, path C). In fact, early evidence suggests that physiological changes may only account for a small percentage of variance in clinical outcomes [133,134]. In current efficacy studies, physiological changes are treated as secondary to clinical changes [39], with many studies failing to check whether subjects have been trained to criterion, thus lacking a demonstration that physiological responses have truly been altered [43]. When changes are monitored in neurofeedback, many individuals seem unable to effectively control their physiological activity [135]. However, even in the absence of significant physiological changes, or even when the change was in the wrong direction, significant clinical improvements can still be found [131]. In HRV biofeedback training, similar findings indicated improvements in symptoms of depression and anxiety, although no significant changes in HRV could be established [136]. In addition, improvements in anxiety were found even when sham feedback was given [137]. This could indicate that certain expectations or perceptions of control in relation to biofeedback training may be important determinants for the treatment outcomes of biofeedback [62,135]. However, empirical evidence regarding specific working mechanisms and possible influencing factors related to physiological changes in biofeedback training is lacking.

In the current biofeedback literature, the efficacy of biofeedback is attributed to the implementation of operant learning reinforcement and meditation techniques as well as changes in interoceptive awareness and physiology (Figure 1). Although there is indeed a wealth of theoretical speculation about the importance of these factors, the empirical evidence linking these processes to biofeedback outcomes remains scarce. Equally important is the fact that there are likely important mechanisms

XSL•FO
RenderX

that mediate the effectiveness of biofeedback that have not yet been investigated. Specifically, we propose that there is a set of cognitive appraisals that acts as a mediator between previously identified processes (eg, interoceptive awareness and physiological change) and anxiety reduction from biofeedback training. In the section *Cognitive Mechanisms Underlying Biofeedback Processes*, we introduce this set of cognitive mechanisms and discuss the importance of including them in a more comprehensive framework for understanding biofeedback training.

Cognitive Mechanisms Underlying Biofeedback Processes

Regulating physiology (such as heart rate) is only one aspect of emotion regulation. Emotion regulation also includes changing subjective experience (affective states) as well as cognitions [13]. Cognitive processes, including how we interpret and evaluate certain situations or feelings, play a major role in the onset and regulation of emotion as well as how we respond to and interact with our environment [11,13,138,139]. Evaluations and interpretations of certain events or situations are commonly referred to as appraisals, which are central to our current understanding of emotion [104,140]. Appraisals drive our entire emotional experience, including how our body reacts (eg, sweating or heart racing), how we feel (eg, anxious or calm), and how we take action (eg, avoid or approach) [139,141-144]. Furthermore, dysfunctional interpretations and attributions mediate recovery from mental health problems [145,146]. Appraisals may, therefore, be an important factor explaining or contributing to the effectiveness of biofeedback interventions that aim to help individuals with stress and anxiety.

The broad categories of appraisals include valence (eg, whether something is viewed as positive or negative); causal agency; evaluations regarding one's potential to cope with or control a situation; and compatibility and relevance with regard to one's goals, norms, and expectations. Different appraisals of the same situation or stressor can therefore lead to an array of emotional responses [140]. We suggest that there are varying levels of appraisals that are relevant to understanding the mechanisms by which biofeedback works: appraisals of the self, such as self-efficacy and locus of control, and situational appraisals, such as threat-challenge appraisals.

Self-Efficacy

When considering appraisal mechanisms by which biofeedback may work to regulate anxiety, appraisals about the self are particularly relevant [140,147]. For example, someone's uncertainty regarding their ability to cope with a situation is highly predictive of anxiety [148]. Whether someone succeeds in self-regulation may therefore strongly depend on their self-efficacy, in other words, the belief that they can do it. Various meta-analyses have shown that self-efficacy is important for self-development, adaptation, and change [149]. Moreover, high self-efficacy is linked with better emotion regulation skills and general psychosocial functioning [150]. In contrast, low levels of self-efficacy are accompanied by high levels of anxiety [151-153]. In youth, self-efficacy has been found to be predictive of the development and maintenance of affective disorders [154,155]. Furthermore, self-efficacy is an important predictor of treatment outcomes for panic disorder [156]. Given the importance of self-efficacy in anxiety regulation, it is likely a key mechanism for effective biofeedback training.

Locus of Control

An additional way in which self-efficacy may be tied to biofeedback is by changing the locus of control. Locus of control refers to the degree to which individuals believe that they themselves have control over the outcome of events in their lives (internal locus) as opposed to forces outside of their control (external locus) [157,158]. People with high self-efficacy generally have an internal locus of control, believing that their own actions and decisions shape outcomes. In contrast, people with low self-efficacy have an external locus of control, often viewing their lives as being beyond their control [149]. Control is an important factor that determines distress and anxiety in a given situation. Specifically, the less someone feels like they are in control, the more anxious they become [159-163]. Moreover, perceived control has been shown to predict outcomes of mental health therapy, with stronger feelings of control being linked to better outcomes [164]. The amount of control someone feels in biofeedback paradigms may therefore influence the success of the training [60,162,164].

Over the course of biofeedback training, with repeated practice and continued feedback on physiological changes, participants in training may increasingly believe that they themselves can influence the outcome of the intervention, in turn leading to a decrease in distress and anxiety. If biofeedback training indeed leads to an increased internal locus of control and locus of control directly influences anxiety, this may also explain why positive therapeutic outcomes of biofeedback can still be present even in the absence of significant physiological changes or when sham feedback is given.

Threat and Challenge

Appraisals of the self often interact with more specific appraisals regarding the situation at hand to determine how anxious someone is and whether they are able to regulate their anxiety. For instance, when personal resources are perceived as greater than the situational demands posed by a stressor, the situation is likely to be appraised as a *challenge*. However, when situational demands are perceived as exceeding personal resources, the situation is more likely to be appraised as a *threat* [36,165-167]. Therefore, someone with high self-efficacy is more likely to view difficult tasks as something to be mastered (ie, a challenge) rather than something that should be avoided (ie, a threat) [168].

Challenge and threat appraisals are both related to activation in the SNS. However, they differ in how they prepare the body for action. When individuals enter a threat state, the body responds by activating the hypothalamic-pituitary-adrenal axis, leading to increases in cortisol production. In addition, downstream vascular resistance increases and cardiac efficiency decreases. These changes prepare the body for damage of a physical or social nature. In contrast, when individuals are in a challenge state, there is an increased activation of the sympathetic-adrenal-medullary axis, which results in increased oxygenation of the bloodstream to the brain and peripheral sites

XSL•FO
RenderX

and vasodilation, leading to increased cardiac efficiency. These changes prepare the body for approach-oriented behavior [166].

In biofeedback training, the goal is to change a person's physiological activity from an erratic state to one where there is increased cardiac efficiency and synchrony between the parasympathetic and sympathetic branches of the ANS, that is, a state of autonomic balance or homeostasis [39,58,72,118]. However, the extent to which homeostasis is maintained may not solely depend on practice and feedback but may also be determined or moderated by how physiological activity is appraised over the course of biofeedback training. According to the threat-challenge model, these appraisals are key to changing physiology, cognition, and behavior in stressful situations [166,169]. Indeed, reappraising stress arousal as helpful rather than harmful has been shown to effectively reduce attention bias to threat cues and improve physiological functioning, resulting in decreased vasoconstriction and increased cardiac efficiency [170]. Thus, another way in which biofeedback may work is by helping individuals to shift from interpreting their physiological arousal as being indicative of a challenge rather than a threat.

An Integrative Model of Biofeedback Training

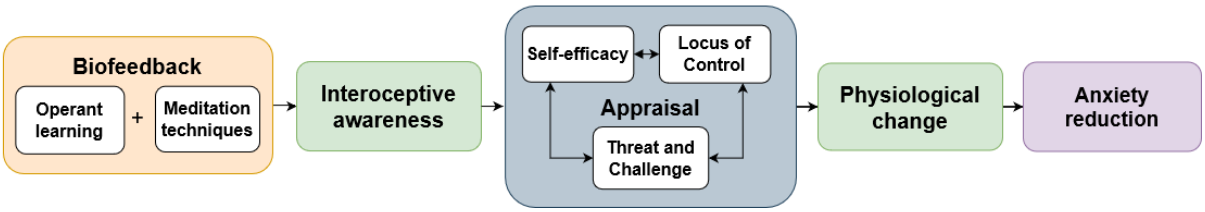
We began by reviewing the main factors to which the effectiveness of biofeedback has conventionally been attributed, including the implementation of operant learning and meditation techniques and changes in interoceptive awareness and physiology (Figure 1). We suggested that cognitive processes have been largely neglected in biofeedback research even though they play an important role in emotion regulation and adaptive trajectories of anxiety treatment [11,13,138,139,145,146]. Several cognitive appraisal dimensions were identified that may be important determinants of adaptive trajectories within biofeedback training. Specifically, we identified appraisals varying from appraisals of the self to situation-specific appraisals. In the section Real Time Change we propose an integrative model that combines and causally links these appraisals with previously identified mechanisms to explain the effectiveness of biofeedback. In this model, we pay particular attention to both the real-time changes that occur within single biofeedback sessions and the developmental changes that may happen as a result of repeated exposure and training.

Real Time Change

Figure 2 summarizes our integrative model of biofeedback effectiveness on a real-time scale. We propose that the relationship between interoceptive awareness and anxiety regulation could be mediated by changes in a person's moment-to-moment appraisals (Figure 2). Focusing on these real-time processes could potentially address conflicting results in past research on biofeedback. On the one hand, increasing interoceptive awareness through biofeedback is thought to facilitate anxiety regulation. On the other hand, we also know that an increase in physiological awareness can lead to an increase in anxiety [94,95]. Examining how cognitive appraisals may interact in a feedback process in real time with interoceptive awareness helps us make sense of these contradictory findings.

At the start of biofeedback training, participants with heightened anxiety symptoms may begin with low interoceptive awareness, and they may not yet be able to effectively self-regulate. However, as biofeedback training continues, awareness of physiological signals may increase as participants begin to detect changes in their physiology. Importantly, however, this increase in awareness may only have a positive effect on self-regulation when certain appraisals are elicited. For instance, when a person becomes aware of physiological sensations and changes related to their anxiety but they have low self-efficacy, their ability to effectively self-regulate may be impeded and give rise to further anxiety. Alternatively, a person with higher self-efficacy who becomes increasingly aware of changes in their physiology may feel more competent to focus on the training and regulate that physiology. Thus, appraisals such as self-efficacy and locus of control may influence the real-time effects of biofeedback training within individual sessions. However, we argue that for lasting improvements in anxiety to be realized through biofeedback, repeated sessions are needed for locus of control, self-efficacy, and threat-challenge appraisals to change and stabilize into more resilient patterns.

Figure 2. Integrative model of changes in real time including appraisal mechanisms.

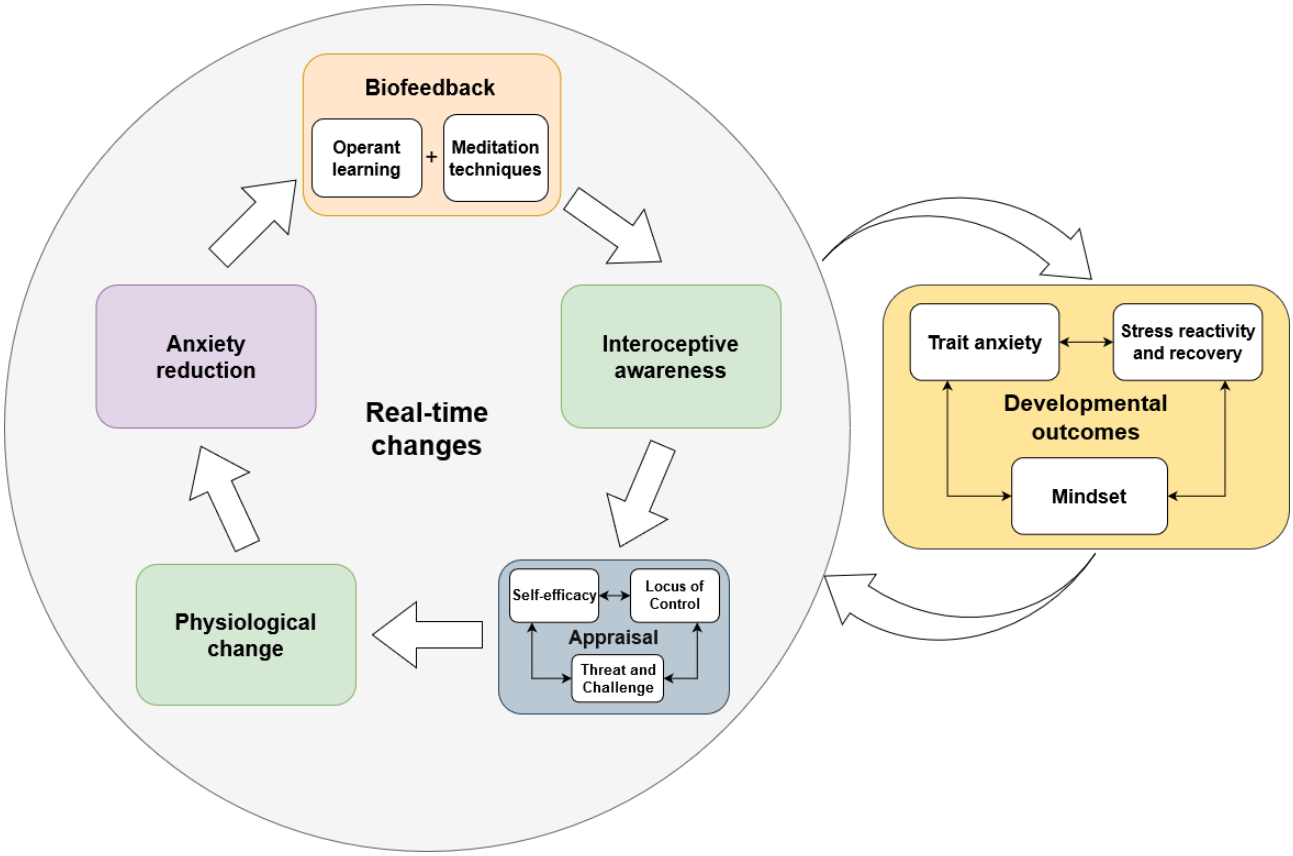


Developmental Change

Some mechanisms of change may already occur within a single biofeedback session, but it is likely that repeated exposure is essential for changes to become automated and internalized. Specifically, we suggest that only when practice is provided in such a way that the proposed changes are optimized will changes generalize to outside the context of the training and influence developmental outcomes such as mindset, stress reactivity and recovery, and trait anxiety (Figure 3).

XSL•FO
RenderX

Figure 3. Model of developmental change showing repeated practice leading to changes in mindset, stress reactivity and recovery, and a trait proclivity to feel anxiety.



Mindset

Before someone engages in a biofeedback training, they already have certain expectations regarding the efficacy of the training and their ability to change. These expectations are tied to a general belief system regarding the malleability of traits. Someone with a growth mindset believes that traits are malleable and change as a result of circumstances and situations. In contrast, someone with a fixed mindset assumes that traits are stable across time and situations [171,172]. Having a fixed mindset compared with a growth mindset has been found to predict psychopathology [173,174]. Furthermore, a recent study found that students who received a growth mindset intervention viewed stressors as less threatening and showed more adaptive physiological responses, which resulted in better performance outcomes, such as better grades [175]. Differences in individual performance and improvement in biofeedback paradigms may likewise be influenced by someone's initial mindset with regard to the malleability of self-regulation skills.

Someone with a growth mindset may have a stronger belief that, through practice, they will be able to effectively exert control over their physiology and will regard the training as a positive challenge. This is because individuals with a growth mindset believe change can be obtained through effortful engagement and they regard failure as a sign to remain engaged and bolster one's efforts [176]. Training as a positive challenge may improve not only someone's performance in the training but also their physiological functioning [170]. Furthermore, we argue that biofeedback training may lead to changes in people's mindset. In initial training sessions, participants may start out with a fixed mindset, low self-efficacy, and a low internal locus of control. However, as participants practice repeatedly and observe their progress in terms of self-regulation, they begin to learn that they are indeed able to effectively change their physiology; this awareness, in turn, may influence their mindset. Specifically, someone who previously had a more fixed mindset with regard to their ability to self-regulate could shift toward a more growth-oriented mindset as the training continues.

Stress Reactivity and Recovery

Over the course of biofeedback training, participants are also likely to change how they physiologically and psychologically respond to stressful situations (ie, their *stress reactivity*) [177] as well as the speed at which they are able to recover after the stressor has ended (ie, their *stress recovery*) [177]. In particular, anxious individuals exhibit increased stress reactivity and slower stress recovery [177-179], but with repeated practice, they may be able to change their reactivity and increase the speed at which they recover. Specifically, as participants receive practice in changing their physiological activity, they are likely to increase the speed at which they are able to do so. Furthermore, as they receive feedback on their progress, they may change the way they perceive changes in physiological activity and their ability to control this change. For instance, instead of being hypervigilant toward bodily cues and changes in physiology, which often results in increases in anxiety, they may tune their attention to a more balanced, optimal level of engagement. It is likely that *riding* this more balanced form of attention to physiology feels more like a *flow* state [180,181] and results in more positive, agentive, and challenge-based appraisals of the

XSL•FO
RenderX

situation. In turn, these challenge-based appraisals will improve stress reactivity and may help individuals to recover more quickly from being exposed to stressors [166,169,170].

Trait Anxiety

With repeated practice, as the self-regulation techniques and changes in cognitive mechanisms become more internalized, people may become more adept at appraising and regulating their physiological and psychological responses to stressful situations. Over development and repeated training sessions, we propose that an increased growth mindset and more adaptive stress reactivity and recovery will develop in concert with lower day-to-day levels of anxiety (ie, their trait anxiety).

Implications for Research

Studies testing the effectiveness of biofeedback training have primarily focused on outcome measures. However, to update theoretical models that link physiology to emotional states, and to improve outcomes from biofeedback interventions, it is vital to study how individuals change with repeated practice and exposure and how outcomes may be tied to certain mediators and moderators. In particular, there seems to be a great deal of promise in studying the appraisal processes outlined in Figures 2 and 3 as mediators within individual sessions as well as across sessions to determine how they influence developmental outcomes.

To capture the mechanisms of change, experimental studies need to be set up in such a way that measurements of interest are repeated within and across multiple training sessions. For instance, training at the very beginning may make individuals feel like they do not have a lot of control over their physiology and the way that they feel, but this perceived control may increase over the course of the training. By assessing locus of control at multiple times during the training, the extent to which changes in these appraisals, in real time, predict or perhaps mediate the effect of the training on anxiety symptoms can be examined. Subsequently, by examining at what point in the training and which particular circumstances these changes occur, future training programs can be designed to amplify these changes and subsequently maximize positive outcomes.

Although it is important to capture change, it is also important to examine how certain beliefs or characteristics present before starting the training (ie, cognitive moderators) may influence performance or how quickly someone learns to self-regulate. For example, someone with a growth mindset who believes that they are able to change their ability to self-regulate through practice may be more receptive to feedback about their physiology, viewing the training as an opportunity to better themselves. In contrast, a fixed mindset may initially inhibit progress in the training, as physiological sensations of anxiety and feedback on physiology may be interpreted as threatening. Thus, in studies focused on evaluating the effect of a biofeedback training, it seems important to include mindset as a moderator in the analyses. Alternatively, participants can be divided into groups using a particular cutoff of the characteristic of interest. The extent to which individuals with one particular trait (such as growth mindset) improve more than those with a fixed mindset can be examined. It would also be interesting to track the trajectories of change over the course of the various training sessions to determine whether they differ between groups. Information about the influence of certain traits and characteristics can inform us about who is an ideal participant for biofeedback training and why, such that future interventions can be improved and better targeted. For instance, if the results indicate that individuals with a growth mindset respond better to biofeedback training, then it may be more effective if the training is adapted in such a way that expectations related to this type of mindset are primed early on in the intervention.

Implications for Designing New Biofeedback Interventions

Thus far, we have proposed a new integrative model that presents several cognitive appraisals as possible moderators or mediators of adaptive outcomes in biofeedback training and have suggested ways in which these appraisals can be assessed in future experimental studies. Furthermore, we emphasized that adaptive and lasting developmental changes can only be achieved by repeated practice. Our design recommendations follow from these points, emphasizing 2 main sets of suggestions. First, biofeedback interventions should work optimally if they enforce all factors in our model, including interoceptive awareness and physiological change as well as the newly proposed cognitive mechanisms. Second, these interventions need to be designed to keep individuals intrinsically motivated and engaged enough to keep practicing over several repeated training sessions. In the following sections, we describe how specific design aspects of interactive technology, such as video games and wearables, could be integrated with biofeedback training to optimize the factors described in our integrative model.

Interoceptive Awareness

Even if someone may become increasingly aware of their internal state while engaging in biofeedback training, it may still be difficult to notice changes in physiology in someone's daily life. Although continued practice may increase the likelihood that changes and skills transfer to outside the training context, this process may be further enhanced by using wearable self-monitoring technology. *Wearable* technology is becoming increasingly easy to integrate in our daily lives [9]. There are already a large number of mobile phone apps available that are either linked to internal sensors (eg, global positioning system or pedometer) or external sensors (eg, heart rate sensors), which can generate real-time information or provide daily, weekly, or even yearly overviews of a person's activities, mood, and physiology. Using these wearables may not only increase someone's interoceptive awareness but may also provide information regarding the contexts in which certain internal changes are most likely to occur. Furthermore, providing an overview of how someone has changed over a certain period may foster a growth mindset as it emphasizes their malleability. Finally, reports provided by wearables may also be valuable for clinicians as they can more closely monitor their clients' improvement or deterioration and identify contexts in which certain exercises or interventions may be most effective.

At this point, it is important to emphasize that although interoceptive awareness is an important aspect of emotion

XSL•FO
RenderX

regulation, merely increasing this awareness may actually result in an increase in anxiety [94,95]. Continuously providing feedback on whether a desired level of physiological activity is reached may lead to extreme attention vigilance and be counterproductive for those anxious people who already focus too much on their internal states. Therefore, fostering a balanced awareness may be a better approach for biofeedback training for anxiety reduction. To achieve this balance, the training should be designed in such a way that other elements in the environment provide opportunities to disengage from attending to physiological activity. For instance, in the biofeedback video game Dojo [49-51], there is a figure of a seated human with a heart displayed in the corner of the computer screen that changes color based on the player's HRV. Specifically, the heart changes in color from red to orange to green, with green being the most optimal level of HRV. Using this type of stimulus clearly communicates changes in the physiology of the player; however, having it continuously present may result in a singular focus on their internal state. Thus, making the training context more engaging in different parts of the game and leading participants' attention away from the heart representation for short periods is likely a more optimal strategy. It may actually be most effective to move away from explicit representations of heart rate, pulse, and other bodily representations of stress and arousal.

If the digital context where self-regulation is practiced is engaging enough (eg, by offering thrilling or beautiful environments to explore where stimuli respond to changes in physiology), these design features may craft a more balanced attention landscape. Practicing in this type of landscape may change from a vigilant form of interoceptive awareness to a state of concentrated relaxation in which participants are aware of their arousal, but instead of fighting it or losing themselves in it, they are in control and move along with it. Gamers have described this state of *concentrated relaxation* as feeling completely focused and concentrated but at the same time feeling like they are able to let go of both [182,183]. For example, the game Nevermind [52], in addition to using explicit representations of HRV (a heart changing in color), also uses atmospheric changes that reflect the player's level of arousal. Players find themselves in an eerie environment that becomes more or less unsettling depending on how well the player is able to self-regulate their arousal. The screen becomes more distorted or ominous noises become even more pronounced as the person feels more anxious. Moreover, specific scenic elements are used, such as the kitchen that floods with milk, the more anxious the player becomes. In addition, the game also includes challenging puzzles that require the player's attention, which means they cannot singularly focus on changes in their HRV [53]. Another exemplary biofeedback app is DEEP, which integrates atmospheric forms of biofeedback [54]. In DEEP, players receive feedback on their breathing by means of a circle that enlarges and shrinks in accordance with the player's breath. Although this circle is always directly in the player's line of sight, the game environment is designed in such a way that it grabs the attention and invites players to explore the beautiful underwater world with its vibrant flora and fauna that respond to players' breath. For example, some of the plants mirror players' in and out breath by growing and shrinking or becoming more or less bright [55,56]. As these examples demonstrate, there is a great deal of promise in integrating biofeedback into engaging environments where participants are made aware of their internal states but feel competent and able to focus on the task at hand.

Self-Efficacy and Locus of Control

One of the foremost sources of information from which we derive our sense of self-efficacy is the genuine experience of mastery [184]. Specifically, our feelings of competence and confidence are largely based on earned failures and successes [184,185]. A sense of mastery can be felt when a task's level of challenge is aligned with our perceived level of competence. In the developmental literature, this ideal balance between challenge and guidance is known as the zone of proximal development [186]. Putting players in this sweet spot of challenge to keep them motivated and engaged is something that video game designers excel at, as is demonstrated by the large number of individuals that play games and by the number of hours that they play them [187-189].

This balance of challenge and difficulty experienced in games is also referred to as a state of flow. Flow is experienced when individuals are immersed in a task, with energized focus, full involvement, and enjoyment in the process of the activity. This flow state coincides with a high sense of control and a loss of self-consciousness [181]. Notably, flow has also been linked to increased self-esteem and decreased anxiety [190]. To achieve a flow state, clear goals must be set beforehand and there must be a possibility to monitor progress. In addition, clear and immediate feedback needs to be given with the aim of balancing the perceived challenge and one's perceived skills [180]. Some of these conditions are already an inherent part of biofeedback, such as immediate feedback and the ability to monitor progress. However, most biofeedback training programs are not specifically designed with flow states in mind. Video games are designed in such a way that flow states are likely to occur as the difficulty level dynamically adapts to the player's current level of mastery, ramping up whenever the skill of the player increases [191]. Although some biofeedback programs do in fact dynamically adjust the level of difficulty, participants may still not experience mastery. To accomplish a true sense of mastery, it is not only important to balance the level of challenge but also to balance rewards and acknowledgments of success. Providing rewards when appropriate instils a sense of accomplishment and pride. However, if challenges are too easily overcome and rewards are disproportionately doled out, the same sense of accomplishment is not felt due to ill-gotten gains [192]. When designing biofeedback interventions, it may therefore be beneficial to focus on facilitating flow and genuine mastery experiences, thereby creating experiences where participants feel competent, in control, and focused on the goals in the training but are still aware of their physiological arousal.

Threat and Challenge

Clinical interventions often focus on decreasing physiological arousal as increases in arousal (eg, quickening of the breath and an increase in heart rate) are often associated with negative emotional states such as anxiety. Interventions that focus on changing appraisals of stress from a threat to a challenge do not focus on reducing arousal or inducing relaxation per se, but

XSL•FO
RenderX

instead aim to change the way the arousal is interpreted and experienced [169,170,193]. Interventions aimed at changing threat-challenge appraisals either directly target arousal interpretations or target higher-level belief systems such as growth mindset. When directly targeting threat-challenge appraisals, participants are often provided with information in a written, oral, and sometimes video form, which demonstrates that arousal is not harmful but rather a functional response of the body that aids performance in challenging and stressful situations. Interventions focused on fostering growth mindset use similar strategies but have an additional emphasis on people's ability to change (for a complete review of threat-challenge manipulation strategies, refer to the study by Jamieson et al [193]). Appraising arousal in a positive manner has been shown to lead to more adaptive stress responses [166] and even better academic performance [169,175].

To encourage challenge appraisals, feedback messages could be implemented in biofeedback training before and throughout biofeedback training, which frames arousal as adaptive. However, a downside to integrating these types of messages is that they are fairly explicit, which may clash with the more implicit, bottom-up type of feedback that is delivered through biofeedback. Therefore, if feedback messages are integrated, they should nicely blend into the environment. One successful example of this type of integration can be found in the phone-based companion app called #SelfCare [194]. In this app, which aims to provide relaxation and mindfulness training, users are invited to stay inside their virtual bedroom where they can interact with objects in the environment and play some minigames. When interacting with the app, messages pop up on screen that encourage the user to observe how they feel, such as "Is there tension somewhere?," "Where does this tension come from?," and "What can we learn from it?." These types of messages frame bodily sensations and arousal as informative, similar to threat-challenge manipulation strategies [193]. Furthermore, the user is addressed in a comforting and encouraging way, such as "There's no hurry, we can stay as long as we'd like." However, the messages do not feel patronizing or forced because they are phrased in a personal and relational way and because they blend in with the soothing and safe atmosphere of the environment. Thus, although biofeedback training may benefit from integrating feedback messages that frame arousal as functional and encourage a growth mindset, we argue that caution must be given so that it is well balanced with the design.

In previous sections, we posited that merely interacting with biofeedback could already help individuals to appraise increasing arousal from a perspective of challenge as they witness that arousal as something they can control. However, the environments in which participants train are often fairly neutral, and participants are rarely given the opportunity to practice in situations that increase their arousal [57]. We argue that especially when it comes to anxiety, it is important that individuals practice in contexts where feelings of anxiety are actually triggered, similar to exposure exercises in anxiety treatment [195]. When practicing self-regulation within contexts that evoke arousal, the likelihood increases that skills are transferred outside of the training to stressful situations where these skills are needed most. Moreover, when practicing within these contexts, reappraisals of arousal are more likely to occur than practicing in a neutral or relaxing setting [196,197]. Some biofeedback-based games such as Nevermind [53] and Mindlight [198] have already adopted the use of anxiety-inducing settings. The latter game, Mindlight, has also been shown to be as effective in relieving anxiety as CBT [199]. Ideally, stressful training environments are designed to increase arousal, but they also allow individuals to directly interact with and exert control over the environment so that they feel immersed, in control, and motivated to continue.

Mindset

Although biofeedback training may already implicitly change participants' mindset as feedback with regard to change and progress is constantly provided, incorporating design elements that explicitly facilitate a growth mindset may further strengthen the training's effects. Various approaches have been used in previous interventions to foster a growth mindset, such as providing individuals with excerpts from scientific texts, showing them educational videos, or giving them writing exercises. All these strategies focus on priming individuals with or explicitly displaying information that emphasizes people's ability to change [175,200]. These mindset interventions have shown positive results in improving physiological responses, work performance, affective responses, and health outcomes (for overview, refer to the study by Jamieson et al [193]).

Although biofeedback training may already foster a growth mindset by providing feedback with regard to physiological change, the training may further benefit from including elements that are explicitly designed to emphasize a person's malleability. For instance, feedback messages in initial biofeedback sessions could be adapted so that they not only indicate how someone is currently doing but also highlight their rate of improvement and that they have the potential to further improve. Furthermore, participants can be provided with regular opportunities to check their progress throughout the course of the training. For instance, they could be shown a visual representation of all changes that they went through from the first session onward. Specifically, feedback messages in the initial session can be used to demonstrate the possibility of change, and showing explicit evidence of a person's progress can serve as a more concrete representation of improvement. Furthermore, for these growth mindset microinterventions, it is essential to structure the specific feedback messages such that they reinforce and focus on the participant's efforts instead of emphasizing that the participant is a competent person [176]. For instance, messages should not focus on personal traits such as "You're awesome!" or "You're so smart!" Instead, these messages should highlight progress and effort such as "You're doing great, it's clear that you are improving!" or "Keep at it, you'll figure it out!" Adding these types of feedback messages could result in someone adopting a growth mindset regarding their ability to self-regulate.

Engagement as a Prerequisite for Change

In the previous section, we have shown how future biofeedback interventions can be designed in such a way that they optimize all factors in our integrative model (Figure 3). Several of these

XSL•FO
RenderX

suggestions focused on harnessing the power of new technologies such as video games and wearables. These forms of technology can directly target some of the suggested mechanisms and are also more likely to keep users engaged and motivated. However, although we have proposed that using these forms of technology is promising in terms of accessibility and engagement, we also suggest that merely integrating game elements or wearables into a biofeedback intervention is not sufficient to guarantee user engagement. For instance, the use of wearables has increased in recent years, and one-third of all adopters stop using their devices after a couple months [201]. A similar lack of adherence and uptake has been observed for the use of digital mental health apps, including game-based interventions [202-205]. For wearables, the observed drop in usage is attributed to the fact that many of these apps suffer from poor user research (or none at all), which results in unsatisfactory user experiences [7,206]. A similar lack of user-centered approaches and design principles has been posited for digital mental health interventions [202,204].

One possible solution to avoid the same pitfalls in the design of the next generation of biofeedback interventions may be to adopt design thinking principles and practices. A recent paper by Scholten and Granic [204] outlined how design thinking could be used to improve digital mental health interventions. Design thinking focuses on empathy, multidisciplinary ideation, and experimentation. The design thinking approach puts the needs of the user at the center of the development process. Furthermore, there is cross-disciplinary teamwork and collaboration, and the development process consists of rapid prototyping and iterative testing of services [204]. Although this is just one suggested approach, its emphasis on user engagement integrated with a strong design foundation that tailors the intervention or app to the user's needs seems particularly promising.

Conclusions

With current technological advances and the increasing interest in the use of self-monitoring technology to improve mental health, we argue that this is an ideal time to launch a new wave of biofeedback training. Our hope for this paper was to inspire a new phase of research and implementation of biofeedback training. We reviewed the more traditional techniques and mechanisms thought to explain the effectiveness of biofeedback: operant learning and meditation techniques, interoceptive awareness, and physiological change. We then identified several cognitive appraisal dimensions as potential determinants of adaptive trajectories within biofeedback training, including self-efficacy, locus of control, and threat-challenge appraisals. Subsequently, we proposed a new comprehensive model addressing real-time as well as developmental processes of change. Specifically, we posited that the relationship between changes in interoceptive awareness and anxiety regulation in individual sessions may be mediated by changes in a person's moment-to-moment appraisals such as self-efficacy and locus of control. Furthermore, we highlighted the importance of repeated exposure and practice to achieve adaptive and lasting developmental changes in growth mindset, stress reactivity and recovery, and trait levels of anxiety. Finally, we presented guidelines for the design of future experimental studies as well as new biofeedback training programs and apps that are in line with our integrative model. In summary, there is a great deal of promise for future biofeedback interventions that harness the power of wearables and video games and that adopt a user-centered approach to help people regulate their anxiety in a way that feels engaging, personal, and meaningful.

Acknowledgments

This work was supported by the Netherlands Organisation of Scientific Research Creative Industry grant (314-99-115). In addition, the authors would like to acknowledge Owen Harris, Niki Smit (Monobanda), Andy Mooney (Paradoxical Recordings), and Bryan Duggan (Dublin Institute of Technology School of Computing) for developing the game DEEP, which inspired this work.

Conflicts of Interest

None declared.

References

1. Lupton D. The Quantified Self: A Sociology of Self-Tracking. Malden, Massachusets: Polity Press; 2016.
2. Barick U, Gowda A, Mohanty R, Dutt A, Somanath M, Mittal S, et al. Harnessing real world data from wearables and self-monitoring devices: feasibility, confounders. MEFANET J 2016;4(1):44-49 [FREE Full text]
3. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
4. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
5. Kim KJ, Shin D. An acceptance model for smart watches. Internet Res 2015 Aug 3;25(4):527-541 [FREE Full text] [doi: 10.1108/IntR-05-2014-0126]
6. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the future patient trial, phase I. Sensors (Basel) 2017 Jan 22;17(1):211 [FREE Full text] [doi: 10.3390/s17010211] [Medline: 28117736]

XSL•FO
RenderX

7. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
8. Swan M. Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. J Sens Actuator Netw 2012 Nov 8;1(3):217-253 [FREE Full text] [doi: 10.3390/jsan1030217]
9. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in medicine. Adv Mater 2018 Jun 11:e1706910 [FREE Full text] [doi: 10.1002/adma.201706910] [Medline: 29893068]
10. Damasio AR. The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 1996 Oct 29;351(1346):1413-1420. [doi: 10.1098/rstb.1996.0125] [Medline: 8941953]
11. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]
12. Niedenthal PM. Embodying emotion. Science 2007 May 18;316(5827):1002-1005. [doi: 10.1126/science.1136930] [Medline: 17510358]
13. Gross JJ. Emotion regulation: affective, cognitive, and social consequences. Psychophysiology 2002 May;39(3):281-291. [doi: 10.1017/s0048577201393198] [Medline: 12212647]
14. Thompson RA. Emotion regulation: a theme in search of definition. Monogr Soc Res Child Dev 1994;59(2-3):25-52. [Medline: 7984164]
15. Füstös J, Gramann K, Herbert BM, Pollatos O. On the embodiment of emotion regulation: interoceptive awareness facilitates reappraisal. Soc Cogn Affect Neurosci 2013 Dec;8(8):911-917 [FREE Full text] [doi: 10.1093/scan/nss089] [Medline: 22933520]
16. Kanbara K, Fukunaga M. Links among emotional awareness, somatic awareness and autonomic homeostatic processing. Biopsychosoc Med 2016;10:16 [FREE Full text] [doi: 10.1186/s13030-016-0059-3] [Medline: 27175214]
17. Kever A, Pollatos O, Vermeulen N, Grynberg D. Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. Pers Individ Dif 2015 Dec;87:20-23 [FREE Full text] [doi: 10.1016/j.paid.2015.07.014]
18. Wiens S. Interoception in emotional experience. Curr Opin Neurol 2005 Aug;18(4):442-447. [doi: 10.1097/01.wco.0000168079.92106.99] [Medline: 16003122]
19. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Arlington, Virgana: American Psychiatric Publications; 2013.
20. Crawley SA, Caporino NE, Birmaher B, Ginsburg G, Piacentini J, Albano AM, et al. Somatic complaints in anxious youth. Child Psychiatry Hum Dev 2014 Aug;45(4):398-407 [FREE Full text] [doi: 10.1007/s10578-013-0410-x] [Medline: 24129543]
21. Eisen A, Engler L. Chronic anxiety. In: Eisen AR, Kearney CA, Schaefer CE, editors. Clinical Handbook of Anxiety Disorders in Children and Adolescents. New Jersey, USA: Jason Aronson; 1995:223-250.
22. Ginsburg GS, Riddle MA, Davies M. Somatic symptoms in children and adolescents with anxiety disorders. J Am Acad Child Adolesc Psychiatry 2006 Oct;45(10):1179-1187. [doi: 10.1097/01.chi.0000231974.43966.6e] [Medline: 17003663]
23. Last CG. Somatic complaints in anxiety disordered children. J Anxiety Disord 1991 Jan;5(2):125-138 [FREE Full text] [doi: 10.1016/0887-6185(91)90024-N]
24. Silverman W, Kurtines W. Anxiety and Phobic Disorders: A Pragmatic Approach. New York, USA: Plenum Press; 1996.
25. Bandelow B, Michaelis S, Wedekind D. Treatment of anxiety disorders. Dialogues Clin Neurosci 2017 Jun;19(2):93-107 [FREE Full text] [Medline: 28867934]
26. Bystritsky A, Hovav S, Sherbourne C, Stein MB, Rose RD, Campbell-Sills L, et al. Use of complementary and alternative medicine in a large sample of anxiety patients. Psychosomatics 2012;53(3):266-272 [FREE Full text] [doi: 10.1016/j.psym.2011.11.009] [Medline: 22304968]
27. Silverman WK, Pina AA, Viswesvaran C. Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. J Clin Child Adolesc Psychol 2008 Jan;37(1):105-130. [doi: 10.1080/15374410701817907] [Medline: 18444055]
28. Beck J. Cognitive Behavior Therapy: Basics and Beyond. New York, USA: Guilford Press; 2011.
29. Kendall PC. Treating anxiety disorders in children: results of a randomized clinical trial. J Consult Clin Psychol 1994 Feb;62(1):100-110. [doi: 10.1037//0022-006x.62.1.100] [Medline: 8034812]
30. Podell JL, Mychailyszyn M, Edmunds J, Puleo CM, Kendall PC. The coping cat program for anxious youth: the FEAR plan comes to life. Cogn Behav Pract 2010 May;17(2):132-141 [FREE Full text] [doi: 10.1016/j.cbpra.2009.11.001]
31. Astin JA, Shapiro SL, Eisenberg DM, Forys KL. Mind-body medicine: state of the science, implications for practice. J Am Board Fam Pract 2003;16(2):131-147 [FREE Full text] [doi: 10.3122/jabfm.16.2.131] [Medline: 12665179]
32. Storch EA, Merlo LJ, Keeley ML, Grabill K, Milsom VA, Geffken GR, et al. Somatic symptoms in children and adolescents with obsessive-compulsive disorder: associations with clinical characteristics and cognitive-behavioral therapy response. Behav Cogn Psychother 2008 May 14;36(3):283-297 [FREE Full text] [doi: 10.1017/s1352465808004335]
33. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

XSL•FO
RenderX

34. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
35. Gilbert C, Moss D. Biofeedback and biological monitoring. In: Moss D, McGrady A, Davies T, Wickramasekera I, editors. Handbook of Mind-Body Medicine for Primary Care. Thousand Oaks, CA: Sage Publications; 2003:109-122.
36. Blascovich J, Mendes W. Social psychophysiology and embodiment. In: Fiske SK, Gilbert DT, Lindzey G, editors. Handbook of Social Psychology. Hoboken, New Jersey: John Wiley & Sons; 2010:194-227.
37. Gevirtz R. The promise of heart rate variability biofeedback: evidence-based applications. Biofeedback 2013 Sep;41(3):110-120 [FREE Full text] [doi: 10.5298/1081-5937-41.3.01]
38. Kaushik RM. Biofeedback in Medicine. Semantic Scholar. 2009. URL: https://www.semanticscholar.org/paper/3-Biofeedback-in-Medicine-Kaushik/c50d870b4f4d77a55da7c10e86bb5faf26c38332 [accessed 2019-06-04] [WebCite Cache ID 78saed6fU]
39. Wheat AL, Larkin KT. Biofeedback of heart rate variability and related physiology: a critical review. Appl Psychophysiol Biofeedback 2010 Sep;35(3):229-242. [doi: 10.1007/s10484-010-9133-y] [Medline: 20443135]
40. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol Med 2017 Nov;47(15):2578-2586. [doi: 10.1017/S0033291717001003] [Medline: 28478782]
41. Lantyer AD, Viana MD, Padovani RD. Biofeedback no tratamento de transtornos relacionados ao estresse e à ansiedade: uma revisão crítica. Psico-USF 2013 Apr;18(1):131-140 [FREE Full text] [doi: 10.1590/S1413-82712013000100014]
42. Schoenberg PL, David AS. Biofeedback for psychiatric disorders: a systematic review. Appl Psychophysiol Biofeedback 2014 Jun;39(2):109-135. [doi: 10.1007/s10484-014-9246-9] [Medline: 24806535]
43. Yucha C, Montgomery D. Evidence-based Practice in Biofeedback and Neurofeedback. Immune Stimulation. 2008. URL: http://downloads.imune.net/medicalbooks/Evidence_Based_Practice_in_Biofeedback_and_Neurofeedback_2008_0_8_26_2011_9_15_24_AM_0_.pdf [accessed 2019-06-04]
44. de Jonge P, Wardenaar KJ, Hoenders HR, Evans-Lacko S, Kovess-Masfety V, Aguilar-Gaxiola S, et al. Complementary and alternative medicine contacts by persons with mental disorders in 25 countries: results from the world mental health surveys. Epidemiol Psychiatr Sci 2018 Dec;27(6):552-567 [FREE Full text] [doi: 10.1017/S2045796017000774] [Medline: 29283080]
45. Kessler RC, Soukup J, Davis RB, Foster DF, Wilkey SA, Van Rompay MI, et al. The use of complementary and alternative therapies to treat anxiety and depression in the United States. Am J Psychiatry 2001 Feb;158(2):289-294. [doi: 10.1176/appi.ajp.158.2.289] [Medline: 11156813]
46. Kazdin AE. Evidence-based treatment research: advances, limitations, and next steps. Am Psychol 2011 Nov;66(8):685-698. [doi: 10.1037/a0024975] [Medline: 22082384]
47. Dedeepya P, Nuvvula S, Kamatham R, Nirmala SV. Behavioural and physiological outcomes of biofeedback therapy on dental anxiety of children undergoing restorations: a randomised controlled trial. Eur Arch Paediatr Dent 2014 Apr;15(2):97-103. [doi: 10.1007/s40368-013-0070-3] [Medline: 23907736]
48. Parnandi A, Gutierrez-Osuna R. Physiological modalities for relaxation skill transfer in biofeedback games. IEEE J Biomed Health Inform 2017 Mar;21(2):361-371. [doi: 10.1109/JBHI.2015.2511665] [Medline: 28055927]
49. Scholten H, Malmberg M, Lobel A, Engels RC, Granic I. A randomized controlled trial to test the effectiveness of an immersive 3D video game for anxiety prevention among adolescents. PLoS One 2016;11(1):e0147763 [FREE Full text] [doi: 10.1371/journal.pone.0147763] [Medline: 26816292]
50. Schuurmans AA, Nijhof KS, Engels RC, Granic I. Using a videogame intervention to reduce anxiety and externalizing problems among youths in residential care: an initial randomized controlled trial. J Psychopathol Behav Assess 2018;40(2):344-354 [FREE Full text] [doi: 10.1007/s10862-017-9638-2] [Medline: 29937625]
51. Schuurmans AA, Nijhof KS, Vermaes IP, Engels RC, Granic I. A pilot study evaluating 'dojo,' a videogame intervention for youths with externalizing and anxiety problems. Games Health J 2015 Oct;4(5):401-408. [doi: 10.1089/g4h.2014.0138] [Medline: 26287930]
52. Nevermind. 2015. URL: https://nevermindgame.com [accessed 2020-06-17]
53. Lobel A, Gotsis ME, Annetta M, Engels RC, Granic I. Designing and Utilizing Biofeedback Games for Emotion Regulation: The Case of Nevermind. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2851581.2892521]
54. Harris O, Smith N. Explore Deep. 2015. URL: http://exploredeep.com [accessed 2020-06-17]
55. van Rooij M, Lobel A, Harris O, Smit N, Granic I. DEEP: A Biofeedback Virtual Reality Game for Children At-risk for Anxiety. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2851581.2892452]
56. Weerdmeester J, van Rooij M, Harris O, Smit N, Engels R, Granic I. Exploring the Role of Self-Efficacy in Biofeedback Video Games. In: Extended Abstracts: Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play. 2017 Presented at: CHI PLAY'17; October 15-18, 2017; Amsterdam, Netherlands. [doi: 10.1145/3130859.3131299]

XSL•FO
RenderX

57. Parnandi A, Ahmed B, Shipp E, Gutierrez-Osuna R. Chill-Out: Relaxation Training through Respiratory Biofeedback in a Mobile Casual Game. In: Proceedings of the International Conference on Mobile Computing, Applications, and Services. 2013 Presented at: MobiCASE'13; November 7-8, 2013; Paris, France. [doi: 10.1007/978-3-319-05452-0_18]
58. Lehrer PM, Vaschillo E, Vaschillo B. Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. Appl Psychophysiol Biofeedback 2000 Sep;25(3):177-191. [doi: 10.1023/a:1009554825745] [Medline: 10999236]
59. Peper E, Harvey R, Takebayashi N. Biofeedback an evidence based approach in clinical practice. Jpn J Biofeed Res 2009:3-10 [FREE Full text]
60. McKee MG. Biofeedback: an overview in the context of heart-brain medicine. Cleve Clin J Med 2008 Mar;75(Suppl 2):S31-S34 [FREE Full text] [doi: 10.3949/ccjm.75.suppl_2.s31] [Medline: 18540143]
61. Simkin DR, Thatcher RW, Lubar J. Quantitative EEG and neurofeedback in children and adolescents: anxiety disorders, depressive disorders, comorbid addiction and attention-deficit/hyperactivity disorder, and brain injury. Child Adolesc Psychiatr Clin N Am 2014 Jul;23(3):427-464. [doi: 10.1016/j.chc.2014.03.001] [Medline: 24975621]
62. Ninaus M, Kober SE, Witte M, Koschutnig K, Stangl M, Neuper C, et al. Neural substrates of cognitive control under the belief of getting neurofeedback training. Front Hum Neurosci 2013;7:914 [FREE Full text] [doi: 10.3389/fnhum.2013.00914] [Medline: 24421765]
63. Coben R, Evans J. Neurofeedback and Neuromodulation Techniques and Applications. London, UK: Academic Press; 2010.
64. Lawrence EJ, Su L, Barker GJ, Medford N, Dalton J, Williams SC, et al. Self-regulation of the anterior insula: reinforcement learning using real-time fMRI neurofeedback. Neuroimage 2014 Mar;88:113-124. [doi: 10.1016/j.neuroimage.2013.10.069] [Medline: 24231399]
65. McLeod S. Skinner - Operant Conditioning. Simply Psychology: Study Guides for Psychology Students. 2018. URL: https://www.simplypsychology.org/operant-conditioning.html [accessed 2019-06-05] [WebCite Cache ID 78uDUlBtP]
66. Hammond DC. Neurofeedback treatment of depression and anxiety. J Adult Dev 2005 Aug;12(2-3):131-137 [FREE Full text] [doi: 10.1007/s10804-005-7029-5]
67. Hammond DC. What is neurofeedback? J Neurother 2007 Mar 29;10(4):25-36 [FREE Full text] [doi: 10.1300/J184v10n04_04]
68. Khalsa SS, Rudrauf D, Davidson RJ, Tranel D. The effect of meditation on regulation of internal body states. Front Psychol 2015;6:924 [FREE Full text] [doi: 10.3389/fpsyg.2015.00924] [Medline: 26217263]
69. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
70. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life. Conneticut, USA: Hyperion; 1994.
71. Dziembowska I, Izdebski P, Rasmus A, Brudny J, Grzelczak M, Cysewski P. Effects of heart rate variability biofeedback on EEG alpha asymmetry and anxiety symptoms in male athletes: a pilot study. Appl Psychophysiol Biofeedback 2016 Jun;41(2):141-150. [doi: 10.1007/s10484-015-9319-4] [Medline: 26459346]
72. McCraty R, Atkinson M, Tomasino D, Bradley R. The coherent heart heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. Transdiscip Transcult J New Thought Res Praxis 2009;5(2):11-115 [FREE Full text]
73. Bradley RT, McCraty R, Atkinson M, Tomasino D, Daugherty A, Arguelles L. Emotion self-regulation, psychophysiological coherence, and test anxiety: results from an experiment using electrophysiological measures. Appl Psychophysiol Biofeedback 2010 Dec;35(4):261-283. [doi: 10.1007/s10484-010-9134-x] [Medline: 20559707]
74. Baer R. Mindfulness-based Treatment Approaches: Clinician's Guide to Evidence Base and Applications. Burlington, MA: Elsevier; 2006.
75. Vaschillo EG, Vaschillo B, Lehrer PM. Characteristics of resonance in heart rate variability stimulated by biofeedback. Appl Psychophysiol Biofeedback 2006 Jun;31(2):129-142. [doi: 10.1007/s10484-006-9009-3] [Medline: 16838124]
76. Lehrer PM, Gevirtz R. Heart rate variability biofeedback: how and why does it work? Front Psychol 2014;5:756 [FREE Full text] [doi: 10.3389/fpsyg.2014.00756] [Medline: 25101026]
77. Cahn BR, Polich J. Meditation states and traits: EEG, ERP, and neuroimaging studies. Psychol Bull 2006 Mar;132(2):180-211. [doi: 10.1037/0033-2909.132.2.180] [Medline: 16536641]
78. Demos J. Getting Started with Neurofeedback. New York, USA: Norton & Company; 2005.
79. Khalsa SS, Rudrauf D, Damasio AR, Davidson RJ, Lutz A, Tranel D. Interoceptive awareness in experienced meditators. Psychophysiology 2008 Jul;45(4):671-677 [FREE Full text] [doi: 10.1111/j.1469-8986.2008.00666.x] [Medline: 18503485]
80. Cameron OG. Interoception: the inside story--a model for psychosomatic processes. Psychosom Med 2001;63(5):697-710. [doi: 10.1097/00006842-200109000-00001] [Medline: 11573016]
81. Chentsova-Dutton YE, Dzokoto V. Listen to your heart: the cultural shaping of interoceptive awareness and accuracy. Emotion 2014 Aug;14(4):666-678. [doi: 10.1037/a0036193] [Medline: 24749640]
82. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The multidimensional assessment of interoceptive awareness (MAIA). PLoS One 2012;7(11):e48230 [FREE Full text] [doi: 10.1371/journal.pone.0048230] [Medline: 23133619]

XSL•FO
RenderX

83. Garfinkel SN, Seth AK, Barrett AB, Suzuki K, Critchley HD. Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. Biol Psychol 2015 Jan;104:65-74 [FREE Full text] [doi: 10.1016/j.biopsycho.2014.11.004] [Medline: 25451381]
84. Craig AD. How do you feel? Interoception: the sense of the physiological condition of the body. Nat Rev Neurosci 2002 Aug;3(8):655-666. [doi: 10.1038/nrn894] [Medline: 12154366]
85. Damasio AR, Tranel D, Damasio HC. Somatic markers and guidance of behavior: theory and preliminary testing. In: Levin HS, Eisenberg HM, Benton AL, editors. Frontal Lobe Function and Dysfunction. New York, USA: Oxford Publishing; 1991:217-228.
86. Fourie P. An Evaluation of Functional Assessment and Biofeedback for the Treatment of Generalised Anxiety Disorder. Bond University. 2014. URL: https://pure.bond.edu.au/ws/portalfiles/portal/18243122/AN_EVALUATION_OF_FUNCTIONAL_ASSESSMENT_AND_BIOFEEDBACK_FOR_THE_TREATMENT_OF_GENERALISED_ANXIETY_DISORDER.pdf [accessed 2019-06-05]
87. Ratanasiripong P, Sverduk K, Prince J, Hayashino D. Biofeedback and counseling for stress and anxiety among college students. J Coll Stud Dev 2010;53(5):97-100 [FREE Full text] [doi: 10.1353/csd.2012.0070]
88. Mehling WE, Gopisetty V, Daubenmier J, Price CJ, Hecht FM, Stewart A. Body awareness: construct and self-report measures. PLoS One 2009;4(5):e5614 [FREE Full text] [doi: 10.1371/journal.pone.0005614] [Medline: 19440300]
89. Rachman S. Anxiety (Clinical Psychology: A Modular Course). New York, USA: Psychology Press; 2004.
90. Lux E, Hawlitschek F, Adam M, Pfeiffer J. Using Live Biofeedback for Decision Support: Investigating Influences of Emotion Regulation in Financial Decision Making. Association for Information Systems eLibrary. 2015. URL: https://aisel.aisnet.org/ecis2015_rip/50/ [accessed 2019-06-04] [WebCite Cache ID 78sbPWG6M]
91. Sütterlin S, Schulz SM, Stumpf T, Pauli P, Vögele C. Enhanced cardiac perception is associated with increased susceptibility to framing effects. Cogn Sci 2013 Jul;37(5):922-935 [FREE Full text] [doi: 10.1111/cogs.12036] [Medline: 23607678]
92. Pollatos O, Schandry R. Emotional processing and emotional memory are modulated by interoceptive awareness. Cogn Emot 2008 Feb;22(2):272-287 [FREE Full text] [doi: 10.1080/02699930701357535]
93. Sze JA, Gyurak A, Yuan JW, Levenson RW. Coherence between emotional experience and physiology: does body awareness training have an impact? Emotion 2010 Dec;10(6):803-814 [FREE Full text] [doi: 10.1037/a0020146] [Medline: 21058842]
94. Pollatos O, Traut-Mattausch E, Schroeder H, Schandry R. Interoceptive awareness mediates the relationship between anxiety and the intensity of unpleasant feelings. J Anxiety Disord 2007;21(7):931-943. [doi: 10.1016/j.janxdis.2006.12.004] [Medline: 17257810]
95. Pollatos O, Traut-Mattausch E, Schandry R. Differential effects of anxiety and depression on interoceptive accuracy. Depress Anxiety 2009;26(2):167-173. [doi: 10.1002/da.20504] [Medline: 19152366]
96. Clark D, Beck A. Defeating panic and avoidance. In: Clark DA, Beck AT, editors. The Anxiety and Worry Workbook: The Cognitive Behavioral Solution. New York, USA: Guilford Press; 2011:171-177.
97. Hope D, Heimberg R, Turk C. Managing Social Anxiety: A Cognitive-behavioral Therapy Approach. New York, USA: Oxford University Press; 2010.
98. Naragon-Gainey K. Meta-analysis of the relations of anxiety sensitivity to the depressive and anxiety disorders. Psychol Bull 2010 Jan;136(1):128-150. [doi: 10.1037/a0018055] [Medline: 20063929]
99. Domschke K, Stevens S, Pfleiderer B, Gerlach AL. Interoceptive sensitivity in anxiety and anxiety disorders: an overview and integration of neurobiological findings. Clin Psychol Rev 2010 Feb;30(1):1-11. [doi: 10.1016/j.cpr.2009.08.008] [Medline: 19751958]
100. Clark DM. Cognitive therapy for anxiety. Behav Cogn Psychother 2009 Jun 16;14(4):283-294 [FREE Full text] [doi: 10.1017/S0141347300014907]
101. Clark D, Wells A. A cognitive model of social phobia. In: Heimberg RG, Liebowitz MR, Hope DA, Schneier FR, editors. Social Phobia: Diagnosis, Assessment, and Treatment. New York, USA: Guilford Press; 1995:69-94.
102. McEwan KL, Devins GM. Is increased arousal in social anxiety noticed by others? J Abnorm Psychol 1983 Nov;92(4):417-421. [doi: 10.1037//0021-843x.92.4.417] [Medline: 6643820]
103. McCorry LK. Physiology of the autonomic nervous system. Am J Pharm Educ 2007 Aug 15;71(4):78 [FREE Full text] [doi: 10.5688/aj710478] [Medline: 17786266]
104. Oatley K, Keltner D, Jenkins J. Understanding Emotions. Malden, IL: Blackwell Publishing; 2006.
105. Friedman BH. An autonomic flexibility-neurovisceral integration model of anxiety and cardiac vagal tone. Biol Psychol 2007 Feb;74(2):185-199. [doi: 10.1016/j.biopsycho.2005.08.009] [Medline: 17069959]
106. Lyonfields JD, Borkovec T, Thayer JF. Vagal tone in generalized anxiety disorder and the effects of aversive imagery and worrisome thinking. Behav Ther 1995;26(3):457-466 [FREE Full text] [doi: 10.1016/S0005-7894(05)80094-2]
107. Miu AC, Heilman RM, Miclea M. Reduced heart rate variability and vagal tone in anxiety: trait versus state, and the effects of autogenic training. Auton Neurosci 2009 Jan 28;145(1-2):99-103. [doi: 10.1016/j.autneu.2008.11.010] [Medline: 19059813]
108. Porges SW, Doussard-Roosevelt JA, Portales AL, Greenspan SI. Infant regulation of the vagal 'brake' predicts child behavior problems: a psychobiological model of social behavior. Dev Psychobiol 1996 Dec;29(8):697-712. [doi: 10.1002/(SICI)1098-2302(199612)29:8<697::AID-DEV5>3.0.CO;2-O] [Medline: 8958482]

XSL•FO
RenderX

109. Thayer JF, Friedman BH, Borkovec TD. Autonomic characteristics of generalized anxiety disorder and worry. Biol Psychiatry 1996 Feb 15;39(4):255-266. [doi: 10.1016/0006-3223(95)00136-0] [Medline: 8645772]
110. Porges S. Vagal tone: an autonomic mediator of affect. In: Garber J, Dodge KA, editors. Cambridge Studies in Social and Emotional Development. The Development of Emotion Regulation and Dysregulation. New York, USA: Cambridge University press; 1991:111-128.
111. Porges SW. Vagal tone: a physiologic marker of stress vulnerability. Pediatrics 1992 Sep;90(3):498-504. [Medline: 1513615]
112. Porges SW. Cardiac vagal tone: a physiological index of stress. Neurosci Biobehav Rev 1995;19(2):225-233. [doi: 10.1016/0149-7634(94)00066-a] [Medline: 7630578]
113. Gruzelier JH, Thompson T, Redding E, Brandt R, Steffert T. Application of alpha/theta neurofeedback and heart rate variability training to young contemporary dancers: state anxiety and creativity. Int J Psychophysiol 2014 Jul;93(1):105-111. [doi: 10.1016/j.ijpsycho.2013.05.004] [Medline: 23684733]
114. Thayer JF, Ahs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012 Feb;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
115. Appelhans BM, Luecken LJ. Heart rate variability as an index of regulated emotional responding. Rev Gen Psychol 2006 Sep;10(3):229-240. [doi: 10.1037/1089-2680.10.3.229]
116. Thayer JF, Brosschot JF. Psychosomatics and psychopathology: looking up and down from the brain. Psychoneuroendocrinology 2005 Nov;30(10):1050-1058. [doi: 10.1016/j.psyneuen.2005.04.014] [Medline: 16005156]
117. Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. J Affect Disord 2000 Dec;61(3):201-216. [doi: 10.1016/s0165-0327(00)00338-4] [Medline: 11163422]
118. Uy CC, Jeffrey IA, Wilson M, Aluru V, Madan A, Lu Y, et al. Autonomic mechanisms of emotional reactivity and regulation. Psychology 2013;4(8):669-675 [FREE Full text] [doi: 10.4236/psych.2013.48095]
119. Friedman BH, Thayer JF. Anxiety and autonomic flexibility: a cardiovascular approach. Biol Psychol 1998 Mar;47(3):243-263. [doi: 10.1016/s0301-0511(97)00027-6] [Medline: 9564452]
120. Gorman JM, Sloan RP. Heart rate variability in depressive and anxiety disorders. Am Heart J 2000 Oct;140(4):77-83. [doi: 10.1067/mhj.2000.109981] [Medline: 11011352]
121. Agelink MW, Boz C, Ullrich H, Andrich J. Relationship between major depression and heart rate variability. Clinical consequences and implications for antidepressive treatment. Psychiatry Res 2002 Dec 15;113(1-2):139-149. [doi: 10.1016/s0165-1781(02)00225-1] [Medline: 12467953]
122. Yeragani VK, Rao KA, Smitha MR, Pohl RB, Balon R, Srinivasan K. Diminished chaos of heart rate time series in patients with major depression. Biol Psychiatry 2002 May 1;51(9):733-744. [doi: 10.1016/s0006-3223(01)01347-6] [Medline: 11983187]
123. Watkins LL, Blumenthal JA, Carney RM. Association of anxiety with reduced baroreflex cardiac control in patients after acute myocardial infarction. Am Heart J 2002 Mar;143(3):460-466. [doi: 10.1067/mhj.2002.120404] [Medline: 11868052]
124. Berntson GG, Cacioppo JT, Quigley KS. Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications. Psychophysiology 1993 Mar;30(2):183-196. [doi: 10.1111/j.1469-8986.1993.tb01731.x] [Medline: 8434081]
125. Asmundson GJ, Stein MB. Vagal attenuation in panic disorder: an assessment of parasympathetic nervous system function and subjective reactivity to respiratory manipulations. Psychosom Med 1994;56(3):187-193. [doi: 10.1097/00006842-199405000-00002] [Medline: 8084962]
126. Rechlin T, Weis M, Spitzer A, Kaschka WP. Are affective disorders associated with alterations of heart rate variability? J Affect Disord 1994 Dec;32(4):271-275. [doi: 10.1016/0165-0327(94)90091-4] [Medline: 7897091]
127. Banks SJ, Eddy KT, Angstadt M, Nathan PJ, Phan KL. Amygdala-frontal connectivity during emotion regulation. Soc Cogn Affect Neurosci 2007 Dec;2(4):303-312 [FREE Full text] [doi: 10.1093/scan/nsm029] [Medline: 18985136]
128. Thomas KM, Drevets WC, Dahl RE, Ryan ND, Birmaher B, Eccard CH, et al. Amygdala response to fearful faces in anxious and depressed children. Arch Gen Psychiatry 2001 Nov;58(11):1057-1063. [doi: 10.1001/archpsyc.58.11.1057] [Medline: 11695953]
129. Cannon R. Editorial perspective: defining neurofeedback and its functional processes. NeuroRegulation 2015 Jul 15;2(2):60-69 [FREE Full text] [doi: 10.15540/nr.2.2.60]
130. Putman J. The effects of brief, eyes-open alpha brain wave training with audio and video relaxation induction on the EEG of 77 army reservists. J Neurother 2000 Jan;4(1):17-28 [FREE Full text] [doi: 10.1300/J184v04n01_03]
131. Moore NC. A review of EEG biofeedback treatment of anxiety disorders. Clin Electroencephalogr 2000 Jan;31(1):1-6. [doi: 10.1177/155005940003100105] [Medline: 10638346]
132. Price J, Budzynski T. Anxiety, EEG patterns, and neurofeedback. In: Budzynski TH, Budzynski HK, Evans JR, Abarbanel A, editors. Introduction to Quantitative EEG and Neurofeedback. London, UK: Academic Press; 2009:453-470.
133. Cohen HD, Graham C, Fotopoulos SS, Cook MR. A double-blind methodology for biofeedback research. Psychophysiology 1977 Nov;14(6):603-608. [doi: 10.1111/j.1469-8986.1977.tb01209.x] [Medline: 928613]

XSL•FO
RenderX

134. Wickramasekera I. How does biofeedback reduce clinical symptoms and do memories and beliefs have biological consequences? Toward a model of mind-body healing. Appl Psychophysiol Biofeedback 1999 Jun;24(2):91-105. [doi: 10.1023/a:1022201710323] [Medline: 10575537]
135. Witte M, Kober SE, Ninaus M, Neuper C, Wood G. Control beliefs can predict the ability to up-regulate sensorimotor rhythm during neurofeedback training. Front Hum Neurosci 2013;7:478 [FREE Full text] [doi: 10.3389/fnhum.2013.00478] [Medline: 23966933]
136. Pierini D, Tzavaris P, Sgromo D, Pruneti C, Cosentino C. Heart Rate Variability Biofeedback Reduces Symptoms of Depression and Anxiety in Depressed People. In: Biofeedback Federation of Europe 18th Annual Meeting. 2015 Presented at: BFE'15; March 23-28, 2015; Rome, Italy. [doi: 10.13140/RG.2.2.14098.45762]
137. Strunk KK, Sutton GW, Burns NS. Beneficial effects of accurate and false brief biofeedback on relaxation. Percept Mot Skills 2009 Dec;109(3):881-886. [doi: 10.2466/pms.109.3.881-886] [Medline: 20178287]
138. Lazarus RS. Cognition and motivation in emotion. Am Psychol 1991 Apr;46(4):352-367. [doi: 10.1037//0003-066x.46.4.352] [Medline: 2048794]
139. Moors A, Ellsworth PC, Scherer KR, Frijda NH. Appraisal theories of emotion: state of the art and future development. Emot Rev 2013 Mar 20;5(2):119-124 [FREE Full text] [doi: 10.1177/1754073912468165]
140. Ellsworth P, Scherer K. In: Davidson RJ, Scherer KR, Goldsmith HH, editors. Handbook of Affective Sciences. New York, USA: Oxford University Press; 2003.
141. Frijda N, Scherer KR. Emotion definitions (psychological perspectives). In: Sander D, Scherer KR, editors. Oxford Companion to Emotion and the Affective Sciences. New York, USA: Oxford University Press; 2009:142-144.
142. Scherer KR. The dynamic architecture of emotion: evidence for the component process model. Cogn Emot 2009 Nov;23(7):1307-1351 [FREE Full text] [doi: 10.1080/02699930902928969]
143. Moors A. Theories of emotion causation: a review. Cogn Emot 2009 Jun;23(4):625-662 [FREE Full text] [doi: 10.1080/02699930802645739]
144. Scherer K, Schorr A, Johnstone T. Appraisal Processes in Emotion: Theory, Methods, Research. New York, USA: Oxford University Press; 2001.
145. Driessen E, Hollon SD. Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. Psychiatr Clin North Am 2010 Sep;33(3):537-555 [FREE Full text] [doi: 10.1016/j.psc.2010.04.005] [Medline: 20599132]
146. Garratt G, Ingram RE, Rand KL, Sawalani G. Cognitive processes in cognitive therapy: evaluation of the mechanisms of change in the treatment of depression. Clin Psychol Sci 2007 Sep;14(3):224-239 [FREE Full text] [doi: 10.1111/j.1468-2850.2007.00081.x]
147. Beck AT. Cognitive Therapy and the Emotional Disorders. New York, USA: International Universities Press; 1979.
148. Roseman I, Smith C. Appraisal theory: overview, assumptions, varieties, controversies. In: Incherer KR, Schorr A, Johnstone T, editors. Appraisal Processes in Emotion: Theory, Methods, Research. New York, USA: Oxford University Press; 2001:3-19.
149. Bandura A. Guide for constructing self-efficacy scales. In: Urdan T, Pajares F, editors. Self-Efficacy Beliefs of Adolescents. Greenwich, Connecticut: Information Age Publishing; 2006:307-337.
150. Bandura A, Caprara GV, Barbaranelli C, Gerbino M, Pastorelli C. Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning. Child Dev 2003;74(3):769-782. [doi: 10.1111/1467-8624.00567] [Medline: 12795389]
151. Muris P. Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. Pers Individ Dif 2002 Jan;32(2):337-348 [FREE Full text] [doi: 10.1016/S0191-8869(01)00027-7]
152. Mathews BL, Koehn AJ, Abtahi MM, Kerns KA. Emotional competence and anxiety in childhood and adolescence: a meta-analytic review. Clin Child Fam Psychol Rev 2016 Jun;19(2):162-184. [doi: 10.1007/s10567-016-0204-3] [Medline: 27072682]
153. Niditch LA, Varela RE. Perceptions of parenting, emotional self-efficacy, and anxiety in youth: test of a mediational model. Child Youth Care 2011 Oct 1;41(1):21-35 [FREE Full text] [doi: 10.1007/s10566-011-9150-x]
154. O'Neal LJ, Cotten SR. Promotive factors and psychosocial adjustment among urban youth. Child Youth Care 2016 Jun 27;45(6):947-961 [FREE Full text] [doi: 10.1007/s10566-016-9364-z]
155. Bandura A, Pastorelli C, Barbaranelli C, Caprara GV. Self-efficacy pathways to childhood depression. J Pers Soc Psychol 1999 Feb;76(2):258-269. [doi: 10.1037//0022-3514.76.2.258] [Medline: 10074708]
156. Williams S. Self-efficacy, anxiety, and phobic disorders. In: Maddux JE, editor. Self-Efficacy, Adaptation, and Adjustment: Theory, Research, and Application. New York, USA: Plenum Press; 1995:69-107.
157. Rotter J. Social Learning and Clinical Psychology. Innglewood Cliffs, New Jersey: Prentice-Hall; 1954.
158. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 1966;80(1):1-28. [Medline: 5340840]
159. Bandura A. Self-efficacy conception of anxiety. Anxiety Res 1988 Jan;1(2):77-98. [doi: 10.1080/10615808808248222]
160. Bandura A. Self-Efficacy: The Exercise of Control. New York, USA: Freeman; 1997.
161. Ellsworth PC, Smith CA. From appraisal to emotion: differences among unpleasant feelings. Motiv Emot 1988 Sep;12(3):271-302 [FREE Full text] [doi: 10.1007/BF00993115]

XSL•FO
RenderX

162. Roseman I. A model of appraisal in the emotion system. In: Incherer KR, Schorr A, Johnstone T, editors. Appraisal Processes in Emotion: Theory, Methods, Research. New York, USA: Oxford University Press; 2001.
163. Schwarzer R. Self-Efficacy: Thought Control Of Action. New York, USA: Taylor & Francis; 2014.
164. Warmerdam L, van Straten A, Jongsma J, Twisk J, Cuijpers P. Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: exploring mechanisms of change. J Behav Ther Exp Psychiatry 2010 Mar;41(1):64-70. [doi: 10.1016/j.jbtep.2009.10.003] [Medline: 19913781]
165. Gross JJ. Emotion regulation: current status and future prospects. Psychol Inq 2015 Mar 9;26(1):1-26 [FREE Full text] [doi: 10.1080/1047840X.2014.940781]
166. Jamieson JP, Mendes WB, Nock MK. Improving acute stress responses. Curr Dir Psychol Sci 2013 Feb;22(1):51-56 [FREE Full text] [doi: 10.1177/0963721412461500]
167. Seery MD. The biopsychosocial model of challenge and threat: using the heart to measure the mind. Soc Personal Psychol Compass 2013 Sep 2;7(9):637-653 [FREE Full text] [doi: 10.1111/spc3.12052]
168. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
169. Jamieson JP, Mendes WB, Blackstock E, Schmader T. Turning the knots in your stomach into bows: reappraising arousal improves performance on the GRE. J Exp Soc Psychol 2010 Jan 1;46(1):208-212 [FREE Full text] [doi: 10.1016/j.jesp.2009.08.015] [Medline: 20161454]
170. Jamieson JP, Nock MK, Mendes WB. Changing the conceptualization of stress in social anxiety disorder. Clin Psychol Sci 2013 Apr 8;1(4):363-374 [FREE Full text] [doi: 10.1177/2167702613482119]
171. Chiu CY, Hong YY, Dweck CS. Lay dispositionism and implicit theories of personality. J Pers Soc Psychol 1997 Jul;73(1):19-30. [Medline: 9216077]
172. Yeager DS, Dweck CS. Mindsets that promote resilience: when students believe that personal characteristics can be developed. Educ Psychol 2012 Oct;47(4):302-314 [FREE Full text] [doi: 10.1080/00461520.2012.722805]
173. Miu AS, Yeager DS. Preventing symptoms of depression by teaching adolescents that people can change. Clin Psychol Sci 2014 Sep 15;3(5):726-743 [FREE Full text] [doi: 10.1177/2167702614548317]
174. Schleider JL, Abel MR, Weisz JR. Implicit theories and youth mental health problems: a random-effects meta-analysis. Clin Psychol Rev 2015 Feb;35:1-9. [doi: 10.1016/j.cpr.2014.11.001] [Medline: 25462109]
175. Yeager DS, Lee HY, Jamieson JP. How to improve adolescent stress responses: insights from integrating implicit theories of personality and biopsychosocial models. Psychol Sci 2016 Aug;27(8):1078-1091 [FREE Full text] [doi: 10.1177/0956797616649604] [Medline: 27324267]
176. Dweck C, Molden D. Self-theories. In: Inlliot AJ, Dweck CS, editors. Handbook of Competence and Motivation. New York, USA: Guilford Press; 2005:122-140.
177. Linden W, Earle TL, Gerin W, Christenfeld N. Physiological stress reactivity and recovery: conceptual siblings separated at birth? J Psychosom Res 1997 Feb;42(2):117-135. [doi: 10.1016/s0022-3999(96)00240-1] [Medline: 9076640]
178. Villada C, Hidalgo V, Almela M, Salvador A. Individual differences in the psychobiological response to psychosocial stress (trier social stress test): the relevance of trait anxiety and coping styles. Stress Health 2016 Apr;32(2):90-99. [doi: 10.1002/smi.2582] [Medline: 24916722]
179. Vitaliano PP, Russo J, Paulsen VM, Bailey SL. Cardiovascular recovery from laboratory stress: biopsychosocial concomitants in older adults. J Psychosom Res 1995 Apr;39(3):361-377. [doi: 10.1016/0022-3999(94)00144-t] [Medline: 7636779]
180. Csikszentmihalyi M, Abuhamdeh S, Nakamura J. Flow. In: Inlliot AJ, Dweck CS, editors. Handbook of Competence and Motivation. New York, USA: Guilford Press; 2005:598-608.
181. Sherry JL. Flow and media enjoyment. Commun Theory 2004 Nov;14(4):328-347 [FREE Full text] [doi: 10.1111/j.1468-2885.2004.tb00318.x]
182. Kottke J. Update to Alto's Adventure adds endless Zen Mode. Kottke. 2016. URL: https://kottke.org/16/06/update-to-altos-adventure-adds-endless-zen-mode [accessed 2019-06-05] [WebCite Cache ID 78uE0CaFK]
183. Turkle S. The Second Self: Computers and the Human Spirit. New York, USA: Simon & Schuhster; 1984:84-85.
184. Bandura A. The explanatory and predictive scope of self-efficacy theory. J Soc Clin Psychol 1986 Sep;4(3):359-373 [FREE Full text] [doi: 10.1521/jscp.1986.4.3.359]
185. Pajares F, Schunk D. Self-efficacy and self-concept beliefs. In: Marsh HW, Graven RG, McInerney DM, editors. New Frontiers for Self Research. Greenwich, Conneticut: IAP; 2005:95-123.
186. Vygotsky L. Interaction between learning and development. In: Gauvain M, Cole M, editors. Readings on the Development of Children. New York, USA: Scientific American Books; 2019:34-41.
187. Over Half of The 211.2 Million Video Gamers in The US Play Games Across Multiple Platforms, According to NPD. NPD Group - Market Research: Consumer Trends. 2018. URL: https://www.npd.com/wps/portal/npd/us/news/press-releases/2018/over-half-of-the-211-2-million-video-gamers-in-the-u-s--play-games-across-multiple-platforms--according-to-npd/ [accessed 2019-06-05] [WebCite Cache ID 78uJo8lCt]
188. Perrin A. 5 Facts About Americans and Video Games. Pew Research Center. 2018. URL: https://www.pewresearch.org/fact-tank/2018/09/17/5-facts-about-americans-and-video-games/ [accessed 2019-06-05] [WebCite Cache ID 78uEIVUpY]

XSL•FO
RenderX

189. Przybylski AK, Weinstein N. A large-scale test of the goldilocks hypothesis. Psychol Sci 2017 Feb;28(2):204-215. [doi: 10.1177/0956797616678438] [Medline: 28085574]
190. Csikszentmihalyi M, Rathunde K, Whalen S. Talented Teenagers: The Roots of Success and Failure. New York, USA: Cambridge University Press; 1993.
191. Sweetser P, Wyeth P. Gameflow: a model for evaluating player enjoyment in games. Comput Entertain 2005 Jul 1;3(3):3 [FREE Full text] [doi: 10.1145/1077246.1077253]
192. Bura S. Emotion Engineering in Videogames: Toward a Scientific Approach to Understanding the Appeal of Videogames. Stephane Bura. 2008. URL: http://www.stephanebura.com/emotion/version_1_0_2.html [accessed 2019-06-05] [WebCite Cache ID 78uAFBuKf]
193. Jamieson JP, Crum AJ, Goyer JP, Marotta ME, Akinola M. Optimizing stress responses with reappraisal and mindset interventions: an integrated model. Anxiety Stress Coping 2018 May;31(3):245-261. [doi: 10.1080/10615806.2018.1442615] [Medline: 29471669]
194. #SelfCare. #SelfCare on the App Store - Apple. 2018. URL: https://itunes.apple.com/us/app/selfcare/id1378384555?mt=8 [accessed 2020-06-17]
195. Parker Z, Waller G, Gonzalez SD, Dawson J. The role of exposure in treatment of anxiety disorders: a meta-analysis. Int J Psychol Psychol Ther 2018:111-141 [FREE Full text]
196. Driskell J, Johnston J. Stress exposure training. In: Cannon-Bowers JA, Salas E, editors. Making Decisions Under Stress: Implications for Individual and Team Training. Washington, DC: American Psychological Association; 1998:191-217.
197. Salkovskis PM, Hackmann A, Wells A, Gelder MG, Clark DM. Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: a pilot study. Behav Res Ther 2007 May;45(5):877-885. [doi: 10.1016/j.brat.2006.02.008] [Medline: 17296165]
198. Introducing MindLight. PlayNice Institute. URL: https://www.theplayniceinstitute.com/ [accessed 2020-06-17]
199. Schoneveld EA, Lichtwarck-Aschoff A, Granic I. Preventing childhood anxiety disorders: is an applied game as effective as a cognitive behavioral therapy-based program? Prev Sci 2018 Feb;19(2):220-232 [FREE Full text] [doi: 10.1007/s11121-017-0843-8] [Medline: 28956222]
200. Yeager DS, Trzesniewski KH, Dweck CS. An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. Child Dev 2013;84(3):970-988 [FREE Full text] [doi: 10.1111/cdev.12003] [Medline: 23106262]
201. Ledger D, McCaffrey D. Inside Wearables Part 1: How Behavior Change Unlocks Long-Term Engagement. Medium. 2014. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2019-06-05] [WebCite Cache ID 78uDHBydu]
202. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
203. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 1;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
204. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
205. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
206. Norman D. The Design of Everyday Things. New York, USA: Basic Books; 2013.

Abbreviations

ANS: autonomic nervous system
CBT: cognitive behavioral therapy
EEG: electroencephalography
HRV: heart rate variability
PNS: parasympathetic nervous system
RSA: respiratory sinus arrhythmia
SNS: sympathetic nervous system

XSL•FO
RenderX

Edited by G Eysenbach; submitted 06.06.19; peer-reviewed by N Zarifsanaiey, J Hunter; comments to author 09.11.19; revised version received 03.01.20; accepted 22.01.20; published 23.07.20.

Please cite as:
Weerdmeester J, van Rooij MMJW, Engels RCME, Granic I
An Integrative Model for the Effectiveness of Biofeedback Interventions for Anxiety Regulation: Viewpoint
J Med Internet Res 2020;22(7):e14958
URL: https://www.jmir.org/2020/7/e14958
doi:10.2196/14958
PMID:32706654

©Joanneke Weerdmeester, Marieke MJW van Rooij, Rutger CME Engels, Isabela Granic. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Using Digital Technology to Reduce the Prevalence of Mental Health Disorders in Populations: Time for a New Approach

C Barr Taylor[1], MD; Josef I Ruzek[2], PhD; Ellen E Fitzsimmons-Craft[3], PhD; Shiri Sadeh-Sharvit[4], PhD; Naira Topooco[5], PhD; Ruth Striegel Weissman[6], PhD; Daniel Eisenberg[7], PhD; David Mohr[8], PhD; Andrea Graham[8], PhD; Corinna Jacobi[9], PhD; Brian Oldenburg[10], PhD

[1]Stanford University, Palo Alto University, Stanford, CA, United States
[2]Palo Alto University, Palo Alto, CA, United States
[3]Division of Hematology Oncology, Department of Pediatrics, Washington University, St. Louis, MO, United States
[4]Baruch Ivcher School of Psychology, Herzliya, Israel
[5]Linköping University, Linköping, Sweden
[6]Wesleyan University, Middletown, CT, United States
[7]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[8]Northwestern University, Chicago, IL, United States
[9]Technical University of Dresden, Dresden, Germany
[10]The University of Melbourne, Melbourne, Australia

Corresponding Author:
C Barr Taylor, MD
Stanford University
Palo Alto University
850 Esplanada Way
Stanford, CA, 94305
United States
Phone: 1 4159871810
Email: btaylor@stanford.edu

Abstract

Digital technology, which includes the collection, analysis, and use of data from a variety of digital devices, has the potential to reduce the prevalence of disorders and improve mental health in populations. Among the many advantages of digital technology is that it allows preventive and clinical interventions, both of which are needed to reduce the prevalence of mental health disorders, to be feasibly integrated into health care and community delivery systems and delivered at scale. However, the use of digital technology also presents several challenges, including how systems can manage and implement interventions in a rapidly changing digital environment and handle critical issues that affect population-wide outcomes, including reaching the targeted population, obtaining meaningful levels of uptake and use of interventions, and achieving significant outcomes. We describe a possible solution, which is to have an outcome optimization team that focuses on the dynamic use of data to adapt interventions for populations, while at the same time, addressing the complex relationships among reach, uptake, use, and outcome. We use the example of eating disorders in young people to illustrate how this solution could be implemented at scale. We also discuss system, practitioner-related, and other issues related to the adaptation of such an approach. Digital technology has great potential for facilitating the reduction of mental illness rates in populations. However, achieving this goal will require the implementation of new approaches. As a solution, we argue for the need to create outcome optimization teams, tasked with integrating data from various sources and using advanced data analytics and new designs to develop interventions/strategies to increase reach, uptake, use/engagement, and outcomes for both preventive and treatment interventions.

(J Med Internet Res 2020;22(7):e17493) doi:10.2196/17493

KEYWORDS

internet; mental health, interventions; outcome; prevalence

XSL•FO
RenderX

Introduction

Digital technology, which includes the collection, analysis, and use of data from a variety of digital devices and other sources, has the potential to reduce the prevalence of mental health disorders and improve mental health in populations by integrating preventive interventions to reduce incidence with clinical interventions to reduce existing cases, both of which are needed to reduce the prevalence of mental health disorders, into health care and community delivery systems. Furthermore, digital technology has the advantage of collecting large amounts of information that can inform preventive and intervention processes. Such data need to be analyzed and, even more importantly, used to adapt and improve a variety of interventions, including digital ones, toward optimizing outcomes. The process of optimizing outcomes needs to be dynamic and responsive to the rapid changes in the use of digital technology, consumer interests and preferences, and regulations among other factors as well as to enable rapid changes in the content and process of intervention delivery (a study by Michie et al [1] provides an excellent overview of challenges in developing and evaluating digital interventions targeting behavior change as well as methods for doing so). At the same time, interventions need to be designed in partnership with stakeholders and consumers to increase the likelihood of subsequent implementation and dissemination [2-6].

The need to find novel solutions for improving mental health outcomes in populations is based on extensive data, showing that mental health disorders are very common and severely undertreated. In the United States, for instance, approximately 25% of the population experiences a mental health disorder during a given year and 50% in their lifetimes [7]; however, fewer than 30% of individuals with mental health disorders receive any treatment [8,9]. Thus, solutions for reducing prevalence in populations will require approaches that can be applied on a large scale [10].

To meet the need for outcome optimization at a population level, new approaches are required. Digital interventions have been developed to increase access and reduce costs, but engagement with these interventions is suboptimal, with many individuals engaging in only 1 or 2 sessions and fewer than half completing more than half of the treatment [11]. A recent study, for instance, of 93 mental health apps found that the medians of app 15-day and 30-day retention rates were 3.9% (IQR 10.3%) and 3.3% (IQR 6.2%), respectively [12]. To improve population-level outcomes, we have recently argued for the need for outcome optimization at the population level—an approach that simultaneously attempts to improve reach, uptake, engagement, and outcomes [1,13,14]. In this paper, we discuss how prevalence reduction in whole populations might be achieved through outcome optimization applied to both preventive and clinical interventions and suggest ways to address potential issues raised by this approach. Of note, outcome optimization applies to the entire population rather than the individual, and as we are focusing on prevalence, a reduction in caseness. However, as discussed later, the approach applies as well to any mental health or behavioral outcomes. We will argue that outcome optimization needs to be directed by a group of individuals with a diverse set of skills, which we refer to as an outcome optimization team, tasked with both reducing prevalence and incidence of a disorder in a population. Such teams would integrate data from various sources to increase reach, uptake, use/engagement, and outcomes for both preventive and clinical interventions and consider interactions/trade-offs among the variables. In a population model, focusing on incremental improvements in effect size may have much less impact on prevalence than focusing on increasing reach. For instance, Moessner and Bauer [15] noted that an increase of 10% from 10% in treatment utilization would decrease the number of cases by an additional 5%, whereas an improvement in treatment efficacy of 10% (from 10%) would only reduce the number of cases by approximately 2%. However, if the intervention used is not effective whatsoever or of very limited effectiveness, there would not be any population health benefits, irrespective of reach, uptake, and use, and there may also be harm associated with the intervention. Therefore, when considering how to optimize an intervention effect on a population, the aim should be to focus on reach while also considering how to improve efficacy.

The Overall Approach

Prevalence Reduction

The prevalence of a disorder in a population is measured by two primary factors: the number of individuals who are identified with the disorder and the number of individuals who develop the disorder during a defined period. To reduce prevalence, there are two necessary components: preventive interventions that reduce disorder onset (the incidence of new cases) and clinical interventions that are effective enough so that the individual already affected with the disorder no longer meets the clinical criteria for that condition.

A core premise of the approach we are suggesting is the recognition that interventions evaluated under carefully controlled circumstances in selected, often convenience populations may not be as feasible or effective when applied more broadly [5,16,17]. Therefore, to reduce the burden of mental health disorders in populations by delivering interventions at scale, interventions need to be carefully and systematically adapted to the needs, requirements, and interests of different defined populations in a dynamic way. In the following sections, we discuss the necessary components to achieve prevalence reduction through an outcome optimization model and use the example of eating disorders in young people to illustrate what we are proposing. The general model we follow is described in detail in a paper by Wilfley et al [18]. In this model, individuals in various populations are identified via an evidence-based screening [19] that sorts respondents into the following categories: possible anorexia, possible other eating disorder, risk of an eating disorder, and low risk of an eating disorder. The screening tool then provides appropriate prevention or treatment recommendations.

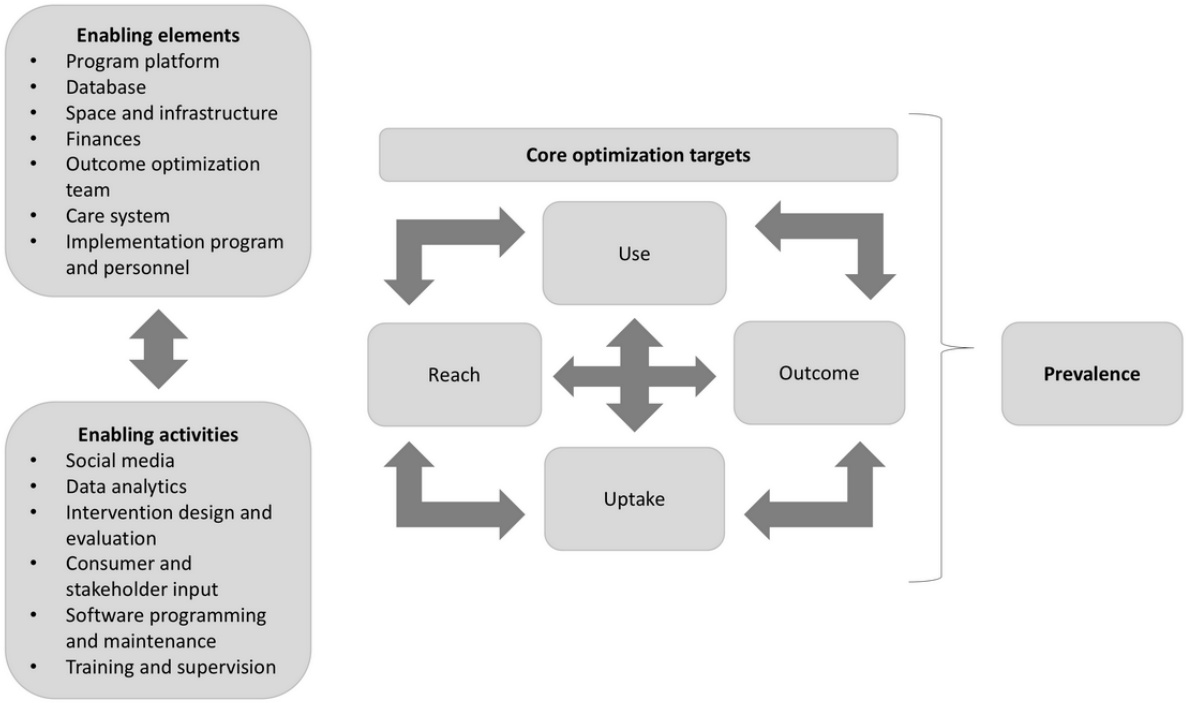
Outcome Optimization Teams

Affecting the key variables for optimizing outcomes is both a management and scientific process. As already mentioned, we have previously described the need for organizations to create

XSL•FO
RenderX

outcome optimization teams or similar groups, tasked with monitoring the key variables mentioned in Figure 1 and developing and implementing strategies to improve the various targets and key outcomes. These teams must possess a set of specific skills to enable effective use of the data to improve the key outcomes, use novel trial designs to conduct rapid evaluations, and iteratively adapt available interventions to improve the effects of the interventions on all critical variables (ie, from reach to outcome). The outcome optimization team needs to include individuals who are able to address administrative issues, manage complex problems, design social media and intervention strategies, and analyze and interpret data, as well technology partners and oversight members (including consumers), brought together with adequate resources for the task. There are now many examples of applications of technology (eg, e-commerce) in areas not related to health and companies that now employ the equivalent of outcome optimization teams to one extent or another, with diverse skills and competencies, focusing on data science to help optimize the outcomes of interest. These approaches are equally relevant to population-level mental health interventions.

Figure 1. Outcome optimization model enabling elements, activities, and targets.



Outcome Optimization Targets

In the optimization outcome approach, the primary components subject to intervention and evaluation (any of which could be targeted separately) are (1) *prevalence* of the disorder, the number of individuals in the population who are appropriate for a clinical intervention; (2) *risk* for the disorder, the number of individuals in the population who are appropriate for a preventive intervention (thus reducing incidence); (3) *reach*, the number of individuals in a population who are offered a relevant prevention or clinical intervention; and (4) *outcome*, the number of individuals who begin a program and achieve the desired outcome within a set time frame. The individual-level outcome would be categorical (case/no case), and the population-level outcome would be prevalence (defined as new cases and existing cases). In addition, two other variables are important for optimizing individual- and population-level outcomes: (5) *uptake*, how many of those individuals offered an intervention actually begin it (eg, create a user account for a web-based, guided self-help program and see a therapist); and (6) *engagement/use* (eg, how much of the intervention is used). To improve overall outcomes, we will argue that each of these parameters can be targeted individually and, yet, should be targeted and measured simultaneously, given that changes in one variable can have a downstream impact on another variable. For example, increased reach can reduce uptake rates, as less motivated individuals may be screened/included, or delivery of more effective treatments characterized by greater demands on client participation might be associated with reduced engagement.

We illustrate the different elements of our proposed approach by drawing from our research with eating disorders. Eating disorders are common and disabling problems with the highest mortality rates of any psychiatric illness. The risk factors for eating disorders are well known and have been shown to reduce incidence when addressed [20], and a variety of effective and cost-effective clinical interventions are available [21-23]. Furthermore, we have already developed and implemented an evidence-based screening on the web to categorize college students as being at risk for or with an eating disorder. Following completion of the web-based screening completion, we have provided appropriate and effective digital interventions or referrals to in-person intervention when appropriate at scale [24].

XSL•FO
RenderX

Reach

In this paper, reach represents the percentage of the population who are at risk for or who have a clinical disorder who complete a web-based screening and who are offered a preventive or clinical intervention appropriate to their needs and interests. We focus on reach via a web-based screening, but individuals at risk for or with clinical disorders could also be identified in other ways, including clinical interviews, algorithms, and, of course, through self-identification.

In the example of eating disorders, the goal of a prevalence reduction program should be to reach most individuals at risk for or with eating disorders. The prevalence of eating disorders in female college students has been estimated to be approximately 13.5% [25]; the rates for those at risk of eating disorders are even higher. For instance, Lipson et al [26] found that approximately 17.29% (1163/6723) of women in their survey had high weight and shape concerns, and many others had binge eating and/or weight and shape concerns without meeting clinical criteria. Conservatively, at least 25% of college-aged women are at risk for but do not have eating disorders. Unfortunately, reach in most populations is often very low. For instance, in our recent study, which used screening in college and universities to recruit women to use an eating disorder intervention, 1.42% (4284/300,613) of undergraduate women completed the web-based screening [27]. Thus, increasing reach is a high priority. An outcome optimization team might use three strategies to increase reach in the population. Hypothetically, the initial focus might be to provide screening through customary channels (eg, student listservs and flyers). In this model, approximately 5.00% (50/1000) of those with eating disorders completed the screening. To increase reach, the next optimization strategy might be to add the use of targeted Facebook advertisements, followed by using both targeted Facebook and Google advertisements. In the example, we assumed that adding Facebook advertisements increased the reach to 7.50% (75/1000), and by adding Google advertisements (strategy 3), the reach increased to 10.00% (100/1000). We realize that these examples are hypothetical, but there is very sparse literature on the use of different digital techniques (eg, social marketing) for increasing reach in targeted populations.

Ten percent is still far below the goal of reaching most of the population, and targeted Facebook and Google advertisements can be expensive. Other, perhaps more effective and less expensive approaches could be used. For instance, web-based screening for an entire student population for a variety of problems, including eating disorders, could be routinely required, or a web-based screening sent directly to individual student emails could be offered. The Healthy Minds Network routinely reaches approximately 16% of students using the latter approach [28]. The outcome optimization team should continuously experiment with new approaches for increasing overall reach and reach for subpopulations of interest.

Uptake

Once individuals have completed screening and have been offered feedback that directs them to a tailored intervention, the next critical variable is the uptake of the intervention, defined as having at least *begun* an intervention, such as engaging with at least some of a digital intervention or attending at least one individual or group therapy session. The uptake of interventions can be quite low. However, in our recent study of college women, 75.5% (690/914) of female students who were eligible for the clinical trial agreed to be randomized; of those randomized to the intervention, 79.2% (305/385) began it [27]. Data from another study that provides prevention and clinical interventions to students in public universities in the state of Missouri found that uptake rates of the digital programs offered for both the high-risk population and clinical population were about 44.2% (420/955) and 49.3% (187/379), respectively [24]. In another study, 26% (16/61) of students with possible anorexia nervosa who were recommended to seek treatment reported doing so at follow-up [29], and Lipson et al [26] found that 18.0% (345/1916) of students at risk of an eating disorder referred to an indicated preventive program began it. Overall, available data on the uptake of interventions have varied quite widely and may be impacted by such features as the population screened and the accessibility and perceived helpfulness of the interventions offered.

In our example, we assumed that for the clinical population, 50% of individuals provided with a referral recommendation will at least go on the web once, see a provider, or in some other way, engage in an intervention. A strategy that lends itself to digital technology is to provide a set of interventions that can be adapted to different needs and interests, with different resources to subpopulations within the larger targeted population. In our hypothetical example, the outcome optimization team found that the baseline uptake rate (number of individuals who were offered an intervention who began it) was 50%. The outcome optimization team explored and enacted a number of options to increase uptake. First, 25% of the sample was eligible but expressed little intention of acting on the recommendation and was provided a brief motivational intervention. Second, approximately 5% of the sample was found to favor Spanish as a primary language. A Spanish language version was authored and offered to interested individuals. A consumer survey found that approximately 25% of the reached population would choose in-person therapy but reported not being able to afford it. As a way to address this issue, the outcome optimization team created an option where therapists trained in evidence-based treatments for eating disorder (eg, cognitive behavioral therapy–guided self-help) would be reimbursed for treating individuals sponsored by a nonprofit organization, and this was offered, randomly, to half the individuals who might want to it. Another issue revealed by the outcome optimization team through consumer needs assessments was that many individuals with eating disorders live in areas with no available treatment expertise for eating disorders and/or preferred teletherapy to face-to-face therapy. The outcome optimization team, thus, developed a teletherapy eating disorder program based on previous work. To increase motivation in future users who met the characteristics identified through a moderator analysis, a postscreen feedback message was then designed to inform them that previous users with characteristics similar to theirs had achieved positive outcomes as a result of the program. Of course, the issues affecting uptake are not mutually exclusive, and variables such as cost, language, and expectation of success may have different impacts on

XSL•FO
RenderX

different individuals depending on their needs and interests. All subpopulations are part of the larger population, and the needs and relative importance of subpopulations will continue to change, reflecting, for instance, changing demographics in the targeted population. Uptake will also change depending on consumer needs and interests, available resources, cost, and other factors.

One of the challenges of population-based interventions is that little is actually known about what users may want. A general strategy is to first consider the interests/resources of the users/stakeholders, followed by using user-centered design [30] approaches to build out interventions designed with ultimate implementation in mind [5]. In building out the interventions, modern designs would be used (eg, the multiphase optimization strategy and others [1,2,31,32] to engineer optimized interventions, before evaluation in randomized controlled trials). Depending on the size of the targeted population, other general approaches could be examined in subsamples. For instance, futility studies of stepped-care approaches, which might appeal to organizations as a way of reducing costs, could be examined [18,33].

Use/Engagement

Program use, often defined as engagement, is the next key variable to consider. Use includes several factors, such as the amount of the prescribed intervention used (eg, sessions of therapy attended and web-based sessions opened); application of the skills outside of the program; and/or sufficient use to lead to reductions in clinical targets, clinical status, risk factor reduction, or other indicators of significant improvement [11]. For example, the results might indicate that individuals who complete a specific amount of intervention might show clinically significant improvements. This amount might refer to the number of sessions attended in face-to-face therapy or pages opened in web-based interventions. Important to determining sufficient engagement is to examine progression within an intervention both for symptom improvement and successful adaptation of behaviors, skills, and other targets considered necessary for a meaningful outcome. Digital technology and analytic methods offer the potential to passively assess a number of variables that might affect engagement and outcome (eg, activity, sleep, and search history). This offers advantages over other types of intervention in the availability of extensive data related to program use and, therefore, easier determination of use cutoffs.

Perhaps the most important use problem is that of early dropout. Early dropout has been defined in various ways, but it is defined, for the purpose of this discussion, as individuals who begin an intervention but complete an insufficient amount of a program to have a positive outcome (however, even this definition is problematic as many individuals seem to benefit from simply beginning an intervention or using it for only a short time) [1]. Recent meta-analyses of web-based studies report high rates of early dropout, partly depending on the amount of support offered [34]. Individuals may drop out of an intervention for a variety of reasons. A major goal of the outcome optimization team would be to understand why people are dropping out (particularly before they achieve a meaningful outcome) and to enact strategies to reduce dropout rates. Furthermore, the outcome optimization team would monitor use/dropout rates as they relate to clinical or preventive outcomes to help determine when a problem exists.

As with other components of the intervention, strategies to reduce early dropout and increase rates of sufficient engagement could be developed and tested. For instance, in a substudy of our recently completed trial [12], we examined the relationship between early and later session completion (n=47): individuals who completed at least three sessions were likely to complete at least 50% of the total program. In another analysis, we found that a very high score on a measure of thin body ideal (TBI) internalization predicted students likely to drop out of the intervention early. We then examined the components of the first three sessions to see what might have negatively impacted 21 individuals with a high TBI. One component tasked users to write a letter to their body. Of these 21 individuals, 10 rated it as not useful at all, 10 as somewhat useful, and only one as very useful. Therefore, we dropped this technique from the program and added a motivational interviewing piece. A subsequent sample of the first 59 users of the motivational interviewing piece was reviewed. The component was rated as not helpful by 10% (6/10) of individuals, a bit helpful by 44% (26/60), helpful by 42% (25/60), and very helpful by 5% (3/60). A number of other small changes were made over the course of a year, and we found that completion rates for the first two sessions increased from 69.5% (105/131) to 78% (70/89) following implementation of all the changes.

Within a dynamic, monitored population, there are innumerable approaches to improving engagement, and methods to do so have been well described [1,5]. However, it is worth noting that as reach is successful in subpopulations, each of these subpopulations is likely to generate new issues of uptake, engagement, and outcome.

Outcome

As with engagement, much has been written about how to improve outcomes using many of the strategies already discussed. The outcome optimization being discussed aligns with movements toward outcome-based care and measurement-based care [34], in which the delivery of interventions provided to clinical populations should focus on the use of outcomes to guide clinical decision making. We briefly mentioned what constitutes a positive outcome (eg, a clinically significant change) measured by loss of clinical status or even abstinence from eating disorder behavior. However, from a population-based prevalence reduction standpoint, there are some important considerations, as illustrated by the eating disorder program we are discussing. First, most of the studies on new models to improve outcome focus on digital interventions. Given some individuals' preference for face-to-face interventions, other treatment modalities and methods should be considered—we mentioned teletherapy above, for instance, as one obvious example. Second, as mentioned previously, focusing on incremental improvements in effect size may have less impact on prevalence than focusing on increasing reach. Third, interventions should be selected and designed with implementation in mind [5]. Mohr et al [33] noted

XSL•FO
RenderX

that when digital interventions that are shown to be efficacious in laboratory-based studies move to real-world settings, individuals do not engage with the tools and implementation often fails. Fourth, the complex issue of the ordering of interventions needs to be considered in population-based models. We, and others, have proposed stepped-care models as a *cost-effective* approach and also as a way to identify subpopulations that may need alternative approaches [14]. Thus, randomized controlled trials remain an important component of digital interventions. At the same time, our central thesis is that researchers must consider issues related to dissemination and implementation from the beginning [5]. This will lead to interventions that maximize public health impact.

Refining the Public Health Model

The model overlaps with other public health models, such as the highly influential Reach, Effectiveness, Adoption, Implementation and Maintenance model [35]. Similar to other public health models [36], we stressed the importance of offering both prevention and intervention approaches to reduce the prevalence. The model proposed here differs from existing models in terms of focusing on the use of digital databases and digital health interventions and innovations in reducing the prevalence of targeted populations. These data and analytics need to be monitored and interpreted by a multidisciplinary team involved in program codesign and delivery so that the digital health solution and its delivery can be quickly refined and improved, an approach consistent with Mohr et al's [5,17] recommendation that program evaluation should be more rapid and agile with iterative improvements. In the model we propose, the key innovation is that the key outcomes (reach, uptake, engagement, and outcome) are being simultaneously and continuously monitored, with ongoing testing and refining of strategies to improve these various indices. Fortunately, some emerging studies deal with at least several components of our proposed model. For instance, the Quit Sequential Multiple Assignment Randomization Trial Utah trial uses a cluster-randomized, multilevel sequential multiple assignment randomized model to examine strategies to increase reach and efficacy in community health centers [37].

System-Level Issues

Our proposed approach has focused on identifying the key elements for reducing symptoms in individuals, to reduce the prevalence of eating disorders in a population. However, program delivery and access and uptake of interventions are affected by a variety of system-level issues, including regulations/policies (eg, reimbursement), service delivery (eg, availability of trained providers), social environment (eg, support of family/friends), infrastructure (eg, availability of exercise equipment for those who may want to use it), accessibility (eg, availability of reliable transportation), and, of course, funding and reimbursement. For example, building out a teletherapy option may be limited by professional practice guidelines and issues of practicing across state lines, reimbursement, access to privacy-protected systems, and/or professional attitudes and training. The outcome optimization team needs to take these factors into consideration so that possible strategies are considered in the context of what can be achieved within the system responsible for prevalence reduction. Another issue is that there may be divergent agendas across different levels of the organization. Providers, supervisors, division leaders, and organizational leadership may have different ideas about what a service should be doing or offering.

Another major challenge for the outcome optimization model is for the outcome optimization team to have access to platforms in which interventions can be easily offered and adapted for use in various formats (eg, mobile apps and web browser access). Furthermore, the delivery system needs to be integrated with a broader health care delivery system. An example of a system designed to integrate digital practice into a common health care record is now being implemented in Australia [38]. Such a system allows feedback between outcome optimization teams and health care practitioners.

Putting It All Together: Prevalence Reduction Through Outcome Optimization

Using this approach at the population level has the potential not only to reduce population prevalence but also to provide important information about intervention and prevention mechanisms, subpopulation needs, and even how to provide personalized and customized interventions to individuals in a population. The model is based on data and data analyses, but most importantly, on having an expert and technically proficient team that is tasked with using data and novel trial designs to conduct ongoing redesign, implementation, and rapid evaluation of evolving interventions that adapt/provide for individuals' needs and address the effects of interventions on all critical variables (ie, from reach to outcome) that can potentially result in greater overall reductions in prevalence. *The creation of outcome optimization teams within organizations tasked with reducing prevalence would be the single most important step toward optimizing outcomes.*

Our approach also assumes that there is agreed-upon access to users' data and their willingness to provide information about their progress. The use of such data will need to be transparent, and the user will need to agree. Thorough and repeated efforts will need to be made to ensure that such information follows Health Insurance Portability and Accountability Act guidelines and remains confidential. The concerns of individuals about sharing data need to be taken seriously, and it is likely that many will opt out of data sharing, given the many examples of how personal data are being misused. However, there is also an increasing movement—sometimes referred to as the *quantified self* movement—toward health consumers having access to their own data and making this information available to their health providers and others to access.

Limitations of Dichotomous Measures as Outcomes

Before discussing how an outcome optimization team model might be enacted, it is important to note that prevalence rates determined by caseness have a number of limitations.

XSL•FO
RenderX

Psychopathology exists on a continuum, and individuals, over time, may fall in and out of risk and/or caseness. Continuous measures are much more likely to be informative than categorical measures. However, for the foreseeable future, prevalence reduction will be measured in terms of reduction/prevention of caseness; thus, the reduction of prevalence remains the most important goal of a population-based approach.

More broadly, digital technologies will need to and are able to address a number of other important issues besides any specific psychopathology or risk that is being addressed. First, dissemination and implementation research have tended to focus on single disorders. However, nearly all mental health problems presenting in practice have substantial comorbidities. In theory, one could reduce the prevalence of one disorder in an individual while having little effect on several other important problems. A number of studies are examining transdiagnostic approaches and considering outcomes across multiple categories and dimensions [12]. Second, being a case may be less important to the individual identified as such than other issues, such as quality of life and well-being. The inclusion of variables relevant to the individual should be part of an intervention system. Third, health behaviors and risk factors (eg, obesity) are intrinsically related to many disorders and should also be addressed. Unfortunately, few models of multicomponent, multidimensional interventions assessed over time have been reported. Fourth, individual-level data obtained through ecological momentary assessment and other techniques can be used as part of a population reduction strategy, and an outcome optimization team will need to consider both general and personalized interventions as part of an array of opportunities provided to individuals at risk for or with clinical disorders. The model we describe is applicable to multidimensional personalized interventions once models for doing so have been developed.

A Note on Prevention

A major strength of digital technology and the use of an outcome optimization team is that screenings and other methods can be used to identify both individuals at risk for and/or with clinical disorders [39] and to address issues of reach, uptake, use, and outcome in at-risk populations using the methods discussed earlier. In most settings, we know of where screenings are used that could identify both risk and caseness; only the latter group is addressed. Not providing preventive interventions is a lost opportunity. We realize that there are many issues with doing so: the burden of adding prevention onto the responsibilities of clinical services, the challenge of providing prevention at a low cost given the large number of individuals who might need to be provided with interventions, and issues of reimbursement, to name a few. However, scalable interventions shown to reduce onset for common problems such as depression [39] and eating disorders [20] are available. Implementing, altering, and adapting them to optimize reach, uptake, use, and outcome (reduction in risk) using the methods we have described may be one of the most important challenges we face.

How Reduction of Population-Wide Prevalence Could Happen

Efforts to reduce the population-wide prevalence of disorders will require considerable resources, and any effort to achieve such reductions will need to be aligned with government/institutional/provider/community perspectives and/or other values and goals. There are many health care systems (eg, Kaiser Permanente, the United States Department of Veteran Affairs) that consider prevalence reduction of disorders (eg, suicide, posttraumatic stress disorder [PTSD], and depression) within their populations to be of importance and routinely screen their populations for problems such as depression and PTSD. Reducing prevalence across populations is similarly critical to many governments and other institutions with missions to serve specific groups. For example, colleges and universities have a strong interest in reducing the prevalence of mental health disorders and ultimately reducing the number of individuals who drop out from college with the result that many could create outcome optimization teams focused on reducing these outcomes in their populations. Another option in health care systems is that groups of collaborative care teams could take responsibility for the outcome optimization of preventive and clinical interventions in their panels, sharing findings among the teams. However, deploying digital technologies focused on reach and uptake may create a major problem for the provider systems: it is likely to create a major increase in demand for services. As we go forward, considerations of the cost-effectiveness and creative design of service delivery models, including stepped-care approaches, will need to be layered into the models.

A population-wide prevalence reduction program also has implications for the health care system, where cost/benefits and trade-offs of combined prevention and intervention need to be considered and may complete competing interests, where, for instance, might be more interested in the short-term benefits of treating cases rather than preventing new cases and considering needs to be given to who directs the outcome optimization team goals.

An alternative approach to a total population-based approach is to focus first on outcome optimization within defined segments of the total population, for instance, individuals identified through screening. This approach does not focus on reductions in overall prevalence but only the reduction of risk and caseness within a subset of the entire population of interest. Working with the National Eating Disorders Association, we created a virtual outcome optimization team (comprising information scientists, statisticians, data managers, psychologists, psychiatrists, program designers, software providers, and others) to focus on both prevention and intervention for eating disorders for individuals identified through the National Eating Disorders Association screening tool. As an example of the need to enact new models of care delivery, over a 6-month period, 71,362 individuals completed the screening and most individuals (61,585/71,362, 86.30%) screened positive for an eating disorder. In addition, 10.20% (6602/71,362) were screened as being at high risk for the

XSL•FO
RenderX

development of an eating disorder. Of those who screened positive for an eating disorder, 85.90% (52,902/61,585) had never received treatment and only 3.00% (1847/61,585) were currently in treatment [40].

As another step forward, funding agencies should actively support innovative population-based interventions that use *newer* designs, including just-in-time adaptive interventions and other ways of thinking about and using digital technology and data to improve outcomes. As noted previously, we remain advocates of traditional randomized controlled trials, when implemented following intervention optimization, and we see them as being most meaningful when instituted within populations where they might eventually be deployed. Funding agencies might initially focus not on a comprehensive population prevalence reduction model but initially examine issues that would be relevant to the model. The investigators would be required to focus on defined, large populations, reach, uptake, engagement, and outcomes using modern data analytics and methods—and many such studies are now underway—with a demonstration as to how, where, and when they would be implemented in real-world situations.

Implementing prevalence reduction in populations is a challenge, but given the large number of people in any population at risk for or with a clinical disorder, scalable, innovative models of service delivery are urgently required [16]. Digital technologies can enable scalability, but new systems and models need to be developed to take advantage of this capability. We have argued that the implementation of outcome optimization teams represents an important possible approach to enable the delivery of technology-facilitated mental health interventions in a way that can optimize outcomes for the entire population.

Acknowledgments

This research was supported by R01 MH100455, T32 HL007456, T32 HL130357, K01 DK116925, and K08 MH120341 from the National Institutes of Health, and Partnership Grant APP1057411, and Centre of Research Excellence APP1170937 from Australia's National Health and Medical Research Council. The authors would like to acknowledge the help from Arielle Smith and Michelle Ferris in preparing the illustrations and manuscript.

Conflicts of Interest

CT is an unpaid scientific board member of SilverCloud and HappyBetter. SS is the Chief Clinical Officer for Eleos Health. DM has accepted consulting or honoraria from Apple Inc., Otuska Pharmaceuticals, and the One Mind Foundation and has an ownership interest in Adaptive Health, Inc. AG has received consulting fees from Actualize Therapy.

References

1. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
2. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for ehealth and mhealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
3. Krause J, van Lieshout J, Klomp R, Huntink E, Aakhus E, Flottorp S, et al. Identifying determinants of care for tailoring implementation in chronic diseases: an evaluation of different methods. Implement Sci 2014 Aug 12;9:102 [FREE Full text] [doi: 10.1186/s13012-014-0102-3] [Medline: 25112492]
4. Brown CH, Curran G, Palinkas LA, Aarons GA, Wells KB, Jones L, et al. An overview of research and evaluation designs for dissemination and implementation. Annu Rev Public Health 2017 Mar 20;38:1-22 [FREE Full text] [doi: 10.1146/annurev-publhealth-031816-044215] [Medline: 28384085]
5. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
6. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of emental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
7. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
8. Patel V, Maj M, Flisher A, de Silva MJ, Koschorke M, Prince M, WPA Zonal and Member Society Representatives. Reducing the treatment gap for mental disorders: a WPA survey. World Psychiatry 2010 Oct;9(3):169-176 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00305.x] [Medline: 20975864]
9. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]

XSL•FO
RenderX

10. Graham AK, Lattie EG, Mohr DC. Experimental therapeutics for digital mental health. JAMA Psychiatry 2019 Aug 21:- epub ahead of print. [doi: 10.1001/jamapsychiatry.2019.2075] [Medline: 31433448]
11. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
12. ClinicalTrials. 2019. Harnessing Mobile Technology to Reduce Mental Health Disorders in College Populations (iAIM EDU) URL: https://clinicaltrials.gov/ct2/show/NCT04162847?term=wilfley&draw=2&rank=1 [accessed 2020-04-04]
13. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
14. Taylor CB, Graham AK, Fitzsimmons-Craft EE, Sadeh-Sharvit S, Balantekin KN, Flatt RE, et al. Optimizing eating disorder treatment outcomes for individuals identified via screening: an idea worth researching. Int J Eat Disord 2019 Nov;52(11):1224-1228. [doi: 10.1002/eat.23169] [Medline: 31502312]
15. Moessner M, Bauer S. Maximizing the public health impact of eating disorder services: a simulation study. Int J Eat Disord 2017 Dec;50(12):1378-1384. [doi: 10.1002/eat.22792] [Medline: 29076172]
16. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
17. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 1;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
18. Wilfley DE, Agras WS, Taylor CB. Reducing the burden of eating disorders: a model for population-based prevention and treatment for university and college campuses. Int J Eat Disord 2013 Jul;46(5):529-532 [FREE Full text] [doi: 10.1002/eat.22117] [Medline: 23658106]
19. Graham AK, Trockel M, Weisman H, Fitzsimmons-Craft EE, Balantekin KN, Wilfley DE, et al. A screening tool for detecting eating disorder risk and diagnostic symptoms among college-age women. J Am Coll Health 2019;67(4):357-366. [doi: 10.1080/07448481.2018.1483936] [Medline: 29979922]
20. Taylor C, Fitzsimmons-Craft, Goel N. Prevention: current status and underlying theory. In: Agras WE, editor. The Oxford Handbook of Eating Disorders. Oxford, UK: Oxford University Press; 2017:246-266.
21. Jacobi C, Völker U, Trockel MT, Taylor CB. Effects of an internet-based intervention for subthreshold eating disorders: a randomized controlled trial. Behav Res Ther 2012 Feb;50(2):93-99. [doi: 10.1016/j.brat.2011.09.013] [Medline: 22137366]
22. Linardon J, Wade TD, de la Piedad GX, Brennan L. The efficacy of cognitive-behavioral therapy for eating disorders: a systematic review and meta-analysis. J Consult Clin Psychol 2017 Nov;85(11):1080-1094. [doi: 10.1037/ccp0000245] [Medline: 29083223]
23. Taylor C, Graham A, Flatt R, Fitzsimmons-Craft E. Current state of scientific evidence on internet-based interventions for the treatment of depression, anxiety, eating disorders, and substance abuse: an overview of systematic reviews and meta-analyses. Eur J Pub Health 2020:- (forthcoming)(forthcoming).
24. Fitzsimmons-Craft EE, Firebaugh M, Graham AK, Eichen DM, Monterubio GE, Balantekin KN, et al. State-wide university implementation of an online platform for eating disorders screening and intervention. Psychol Serv 2019 May;16(2):239-249. [doi: 10.1037/ser0000264] [Medline: 30407047]
25. Eisenberg D, Nicklett EJ, Roeder K, Kirz NE. Eating disorder symptoms among college students: prevalence, persistence, correlates, and treatment-seeking. J Am Coll Health 2011;59(8):700-707 [FREE Full text] [doi: 10.1080/07448481.2010.546461] [Medline: 21950250]
26. Lipson S, Sonneville K. Eating disorder symptoms among undergraduate and graduate students at 12 US colleges and universities. Eat Behav 2017 Jan;24:81-88. [doi: 10.1016/j.eatbeh.2016.12.003] [Medline: 28040637]
27. Fitzsimmons-Craft EE, Balantekin KN, Eichen DM, Graham AK, Monterubio GE, Sadeh-Sharvit S, et al. Screening and offering online programs for eating disorders: reach, pathology, and differences across eating disorder status groups at 28 US universities. Int J Eat Disord 2019 Oct;52(10):1125-1136. [doi: 10.1002/eat.23134] [Medline: 31268183]
28. Healthy Minds Network. 2019. The Healthy Minds Study: 2018-2019 Data Report URL: https://healthymindsnetwork.org/wp-content/uploads/2019/09/HMS_national-2018-19.pdf [accessed 2020-05-02]
29. Fitzsimmons-Craft EE, Eichen DM, Monterubio GE, Firebaugh M, Goel NJ, Taylor CB, et al. Longer-term follow-up of college students screening positive for anorexia nervosa: psychopathology, help seeking, and barriers to treatment. Eat Disord 2019 May 20:1-17 epub ahead of print. [doi: 10.1080/10640266.2019.1610628] [Medline: 31109255]
30. Graham AK, Wildes JE, Reddy M, Munson SA, Taylor BC, Mohr DC. User-centered design for technology-enabled services for eating disorders. Int J Eat Disord 2019 Oct;52(10):1095-1107. [doi: 10.1002/eat.23130] [Medline: 31313370]
31. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. Am J Prev Med 2014 Oct;47(4):498-504 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.021] [Medline: 25092122]
32. Collins LM, Kugler KC. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). New York, USA: Springer International Publishing; 2018.

XSL•FO
RenderX

33. Mohr DC, Lattie EG, Tomasino KN, Kwasny MJ, Kaiser SM, Gray EL, et al. A randomized noninferiority trial evaluating remotely-delivered stepped care for depression using internet cognitive behavioral therapy (CBT) and telephone CBT. Behav Res Ther 2019 Dec;123:103485. [doi: 10.1016/j.brat.2019.103485] [Medline: 31634738]
34. Harding KJK, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. J Clin Psychiatry 2011 Aug;72(8):1136-1143. [doi: 10.4088/JCP.10r06282whi] [Medline: 21295000]
35. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
36. Raza SA, Salemi JL, Zoorob RJ. Historical perspectives on prevention paradox: when the population moves as a whole. J Family Med Prim Care 2018;7(6):1163-1165 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_275_18] [Medline: 30613490]
37. Fernandez ME, Schlechter CR, del Fiol G, Gibson B, Kawamoto K, Siaperas T, et al. QuitSMART Utah: an implementation study protocol for a cluster-randomized, multi-level sequential multiple assignment randomized trial to increase reach and impact of tobacco cessation treatment in community health centers. Implement Sci 2020 Jan 30;15(1):9 [FREE Full text] [doi: 10.1186/s13012-020-0967-2] [Medline: 32000812]
38. Emery J, Boyle D. Data linkage. Aust Fam Physician 2017;46(8):615-619. [Medline: 28787562]
39. van Zoonen K, Buntrock C, Ebert DD, Smit F, Reynolds CF, Beekman AT, et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. Int J Epidemiol 2014 Apr;43(2):318-329 [FREE Full text] [doi: 10.1093/ije/dyt175] [Medline: 24760873]
40. Fitzsimmons-Craft EE, Balantekin KN, Graham AK, Smolar L, Park D, Mysko C, et al. Results of disseminating an online screen for eating disorders across the US: reach, respondent characteristics, and unmet treatment need. Int J Eat Disord 2019 Jun;52(6):721-729. [doi: 10.1002/eat.23043] [Medline: 30761560]

Abbreviations

PTSD: posttraumatic stress disorder
TBI: thin body ideal

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by Y Zhou, J Wang, Y Zhang; comments to author 10.03.20; revised version received 04.04.20; accepted 08.04.20; published 24.07.20.

Please cite as:
Taylor CB, Ruzek JI, Fitzsimmons-Craft EE, Sadeh-Sharvit S, Topooco N, Weissman RS, Eisenberg D, Mohr D, Graham A, Jacobi C, Oldenburg B
Using Digital Technology to Reduce the Prevalence of Mental Health Disorders in Populations: Time for a New Approach
J Med Internet Res 2020;22(7):e17493
URL: https://www.jmir.org/2020/7/e17493
doi:10.2196/17493
PMID:32706665

©C Barr Taylor, Josef I Ruzek, Ellen E Fitzsimmons-Craft, Shiri Sadeh-Sharvit, Naira Topooco, Ruth Striegel Weissman, Daniel Eisenberg, David Mohr, Andrea Graham, Corinna Jacobi, Brian Oldenburg. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Social Medical Capital: How Patients and Caregivers Can Benefit From Online Social Interactions

Pietro Panzarasa[1], PhD; Christopher J Griffiths[2], PhD; Nishanth Sastry[3], PhD; Anna De Simoni[2], PhD

[1]School of Business and Management, Queen Mary University of London, London, United Kingdom

[2]Centre for Primary Care and Public Health, Barts and The London School of Medicine and Dentistry, London, United Kingdom

[3]Department of Computer Science, University of Surrey, Surrey, United Kingdom

Corresponding Author:
Pietro Panzarasa, PhD
School of Business and Management
Queen Mary University of London
Mile End Road
London, E1 4NS
United Kingdom
Phone: 44 (0)20 7882 6984
Email: p.panzarasa@qmul.ac.uk

Abstract

The rapid growth of online health communities and the increasing availability of relational data from social media provide invaluable opportunities for using network science and big data analytics to better understand how patients and caregivers can benefit from online conversations. Here, we outline a new network-based theory of social medical capital that will open up new avenues for conducting large-scale network studies of online health communities and devising effective policy interventions aimed at improving patients' self-care and health.

(J Med Internet Res 2020;22(7):e16337) doi:10.2196/16337

KEYWORDS

online health communities; self-care; social networks; social capital; open and closed structures; social cohesion; brokerage

Recent years have witnessed the rapid growth of online health communities targeted at patients with long-term conditions [1-3]. Patients and caregivers have increasingly used forums and social networks as alternative or complementary sources of support to more traditional forms of health care provision. Recent advances in network science and data-intensive analytics, combined with the growing advent of big data from social media, promise to yield new insights into how patients can tap the full potential of health communities to improve disease self-care and find the support and information they need [4,5].

Ongoing initiatives such as the Values in Action Champions [6], the Health Service Executive Quit Smoking Programme [7] and its Facebook community [8], the Public Health England Stoptober smoking cessation campaign [9], and NHS Digital [10] show how harnessing the power of social networks can help people to spread behaviors, affect cultural changes in the health service, and improve health and well-being outcomes. Similarly, our recent work on communities targeted at patients with respiratory conditions [11] and stroke survivors [3] has provided evidence regarding how certain patients can fare better than others simply by leveraging their communication patterns. What these studies suggest is the idea that some patients can be at an advantage simply because they are somehow better connected than others. In other words, a patient's distinctive pattern of social engagement and communication in a community is an asset in its own right. That asset is what we call social medical capital.

As suggested by previous work on social capital and health care access, there is still controversy over the definition and measurement of social capital and its association with various types of health-related outcomes [12]. In this paper, we endorse a network-based perspective and describe a new conceptual framework for theorizing about the emergence of social medical capital as a function of communication patterns within online health communities. Previous studies of social capital in the health literature have also identified a variety of health-related outcomes that might be affected by social capital [12-15]. These include individuals' health (eg, health status, mortality), health-related behavior (eg, health care–seeking behavior, illness self-management), access to local health services (eg, to community health clinics), and psychological well-being (eg, self-esteem, mutual respect). Here we focus only on two broad

XSL•FO
RenderX

categories of outcomes (informational benefits and social support) and associate each of them with distinct network mechanisms viewed as structural sources of social capital.

We broadly define social medical capital as the advantages that any user (patient or caregiver) can gain from participation in the social networks provided by online health communities, where communication takes place at virtually no cost and across spatial and temporal boundaries. Like other forms of capital (eg, human or financial), social medical capital enables the achievement of certain ends (eg, emotional support); however, unlike other forms, it is based on the idea that social structure serves as a wellspring of advantages to users. That is, social medical capital is contingent on resources socially embedded in connections between users and accessible through these connections. This means not only that users can benefit from one another, but also that whether they can extract value from one another depends on how they interact [16,17]. This idea can be further articulated into four networking principles: (1) defragmentation, (2) bonding, (3) bridging, and (4) multiple-group membership.

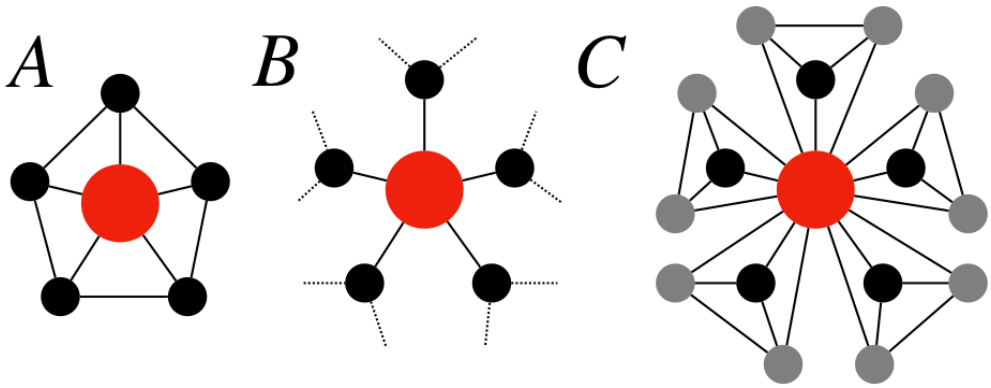
Defragmentation and Superusers

The value of participating in a community lies in the users' ability to gain prompt access to a range of people. As more users join a community, a catalyst of social medical capital is structural defragmentation; that is, social interaction becomes more valuable as more pairs of users become mutually reachable along some path. Recent work has suggested that communities undergo defragmentation by self-organizing into hub-dominated structures [11]. The emergence of a small number of "superusers" (ie, the hubs engaging in conversations with a disproportionally large number of other users) engenders normative control and safeguards the communities from splitting into disjoint components [18].

Bonding and Closed Structures

In a socially cohesive closed network, links are forged locally, and pairs of connected individuals tend to be tied to at least one third party that they have in common (Figure 1A). Previous studies have suggested that closed structures, rich in third-party relationships, engender a shared identity and a sense of belonging, foster trust and cooperation, and sustain emotional support [16,17,19]. Evidence from two communities where patients primarily seek social support has shed light on the role of social cohesion in eliciting social medical capital [11]. Two empirical regularities have been uncovered: (1) highly connected support-giving superusers preferentially communicate with poorly connected support-seeking users; and (2) pairs of support-seeking users who communicate with the same support-giving superuser tend to communicate with each other as well, thus creating closed connected triangles centered on superusers. Local redundancy is therefore the structural engine of a support-oriented community.

Figure 1. Three network structures associated with three distinct forms of social medical capital. The nodes represent the users, and each link refers to communication between users. The type of benefits that a focal user (the red node) can accrue from communication depends on whether the structure is: (A) closed, ie, the user's partners also communicate with each other; (B) open, ie, the user acts as the intermediary between other users; or (C) mixed, ie, the user belongs to multiple densely connected groups and acts as the only intermediary between members of different groups.



Bridging and Open Structures

An open network is rich in structural gaps and opportunities of intermediation between individuals (Figure 1B). Open structures, where connections tend to be weak and constitute "local bridges" between otherwise disconnected parts, have long been associated with informational benefits [16,17,20]. In an openly structured community, communication is likely to take users closer to complementary sources of information they do not already possess [21]. Brokerage opportunities between distinct social circles expose users to greater variance and novelty of resources, and eventually help them to satisfy their informational needs [20,21]. Paucity of local redundancy is therefore the structural engine of an information-oriented community.

Multiple-Group Membership and Mixed Structures

In cases where the community aims to provide both emotional and informational support, social medical capital lies at the interface between bonding and bridging [17,22]. In such cases, the community will be most beneficial to the users if they can combine local redundancy with brokerage opportunities (Figure 1C). In a limiting case, this can be achieved when the following conditions are met: (1) a given user is a member of multiple clusters of users; (2) these clusters do not have any other users in common except the focal one (who therefore is the only intermediary between otherwise disconnected clusters); and (3)

XSL•FO
RenderX

each cluster has an underlying closed cohesive structure (ie, all members are connected with one another).

Research and Policy Roadmap

Empirically testing our theoretical framework will require the construction of large-scale longitudinal network data sets with information on users and their time-stamped messages. To comparatively assess the efficacy of network mechanisms of social capital, content analysis of messages would be needed so as to uncover how users' phycological well-being and access to information vary as social connections also change over time.

To help policy makers to realize our vision of social medical capital as a crucial enabler of large-scale health care interventions in resource-constrained systems, we propose four targets: (1) promote patients' and caregivers' participation in online health communities to ensure continuous provision of socially embedded resources (eg, information, advice, support) that users can access through their (direct and indirect) social connections [14]; (2) support superusers' role in online health communities through appropriate training programs [1]; (3) enhance the quality of peer support and patients' self-care through robust evaluation systems; and (4) develop governance modes for maintaining privacy and confidentiality, cultivating trust and a participatory culture, and promptly detecting and preventing the spreading of misleading information and malicious behavior.

Acknowledgments

We thank Mr Yifan Qian for his help in the preparation of the figure. Nishanth Sastry was partially supported by grant number ES/M00354X/1 (A Shared Space and a Space for Sharing: A Transdisciplinary Exploration of Online Trust and Empathy).

Conflicts of Interest

Pietro Panzarasa and Anna De Simoni were invited by Values in Action, Health Service Executive, and the Chalfont Project to deliver a presentation at the "Network Sciences and Large-Scale Behavioural Change in Healthcare" workshop held in Dublin on February 20, 2019. All other authors declare no conflicts of interest.

References

1. De Simoni A, Taylor S, Griffiths CJ, Panzarasa P, Sheikh A. Online superusers as allies of the health care workforce. New England Journal of Medicine Catalyst 2018:1 [FREE Full text]
2. De Simoni A, Horne R, Fleming L, Bush A, Griffiths CJ. What do adolescents with asthma really think about adherence to inhalers? Insights from a qualitative analysis of a UK online forum. BMJ Open 2017;7(6):e015245 [FREE Full text] [doi: 10.1136/bmjopen-2016-015245] [Medline: 28615272]
3. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. BMJ Open 2016;6(4):e010501 [FREE Full text] [doi: 10.1136/bmjopen-2015-010501] [Medline: 27053271]
4. Garas G, Cingolani I, Panzarasa P, Darzi A, Athanasiou T. Beyond IDEAL: the importance of surgical innovation metrics. The Lancet 2019;393(10169):315. [doi: 10.1016/s0140-6736(18)32204-9]
5. Garas G, Cingolani I, Patel V, Panzarasa P, Alderson D, Darzi A, et al. Surgical innovation in the era of global surgery. Annals of Surgery 2020;271(5):868-874. [doi: 10.1097/sla.0000000000003164]
6. Values in Action Champions. Our Champions' Community. URL: https://www.hse.ie/eng/about/our-health-service/values-in-action/champions/ [accessed 2020-07-08]
7. Sign up and quit smoking. HSE. URL: https://www2.hse.ie/wellbeing/quit-smoking/sign-up-and-quit-smoking/sign-up-and-quit-smoking.html [accessed 2020-07-08]
8. You can quit. HSE. URL: https://www.facebook.com/pg/HSEquit/about/?ref=page_internal [accessed 2020-07-08]
9. Stoptober. URL: https://www.facebook.com/pg/stoptober/about/?ref=page_internal [accessed 2020-07-08]
10. New social media techniques used to boost cancer screening rates. NHS Digital. URL: https://digital.nhs.uk/news-and-events/latest-news/new-social-media-techniques-used-to-boost-cancer-screening-rates [accessed 2020-07-08]
11. Joglekar S, Sastry N, Coulson NS, Taylor SJ, Patel A, Duschinsky R, et al. How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the Asthma UK and British Lung Foundation online communities. J Med Internet Res 2018;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
12. Derose KP, Varda DM. Social capital and health care access: a systematic review. Med Care Res Rev 2009 Jun;66(3):272-306 [FREE Full text] [doi: 10.1177/1077558708330428] [Medline: 19174538]
13. Almedom AM. Social capital and mental health: an interdisciplinary review of primary evidence. Soc Sci Med 2005;61(5):943-964. [doi: 10.1016/j.socscimed.2004.12.025] [Medline: 15955397]
14. Dennis C. Peer support within a health care context: a concept analysis. International Journal of Nursing Studies 2003;40(3):321-332. [doi: 10.1016/S0020-7489(02)00092-5]

XSL•FO
RenderX

15. Kawachi I, Berkman LF. Social cohesion, social capital, and health. In: Berkman LF, Kawachi I, editors. Social Epidemiology. New York: Oxford University Press; 2000:174-190.
16. Lin N. Social Capital. A Theory of Social Structure and Action. New York, NY, USA: Cambridge University Press; 2001.
17. Latora V, Nicosia V, Panzarasa P. Social cohesion, structural holes, and a tale of two measures. J Stat Phys 2013;151(3):745-764. [doi: 10.1007/s10955-013-0722-z]
18. Barabási AL, Albert R. Emergence of scaling in random networks. Science 1999;286(5439):509-512 [FREE Full text] [doi: 10.1126/science.286.5439.509] [Medline: 10521342]
19. Coleman JS. Social capital in the creation of human capital. American Journal of Sociology 1988;94:S95-S120. [doi: 10.1086/228943]
20. Burt RS. Structural Holes. The Social Structure of Competition. Cambridge, MA, USA: Harvard University Press; 1992.
21. Granovetter MS. The strength of weak ties. American Journal of Sociology 1973;78(6):1360-1380. [doi: 10.1086/225469]
22. Burt RS. Brokerage and Closure. An Introduction to Social Capital. Oxford, UK: Oxford University Press; 2005.

Edited by G Eysenbach; submitted 19.09.19; peer-reviewed by P Han, M Stein; comments to author 19.01.20; revised version received 14.03.20; accepted 30.03.20; published 28.07.20.

Please cite as:
Panzarasa P, Griffiths CJ, Sastry N, De Simoni A
Social Medical Capital: How Patients and Caregivers Can Benefit From Online Social Interactions
J Med Internet Res 2020;22(7):e16337
URL: http://www.jmir.org/2020/7/e16337/
doi:10.2196/16337
PMID:32720910

©Pietro Panzarasa, Christopher J Griffiths, Nishanth Sastry, Anna De Simoni. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Bringing Code to Data: Do Not Forget Governance

Christine Suver[1*], PhD; Adrian Thorogood[2*], BCL, LLB; Megan Doerr[1], MSc, LGC; John Wilbanks[1], BA; Bartha Knoppers[2], PhD, ADE

[1]Sage Bionetworks, Seattle, WA, United States

[2]Centre of Genomics and Policy, McGill University, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Christine Suver, PhD
Sage Bionetworks
2901 Third Avenue, Suite 330
Seattle, WA, 98121
United States
Phone: 1 206 928 8242
Email: cfsuver@gmail.com

Abstract

Developing or independently evaluating algorithms in biomedical research is difficult because of restrictions on access to clinical data. Access is restricted because of privacy concerns, the proprietary treatment of data by institutions (fueled in part by the cost of data hosting, curation, and distribution), concerns over misuse, and the complexities of applicable regulatory frameworks. The use of cloud technology and services can address many of the barriers to data sharing. For example, researchers can access data in high performance, secure, and auditable cloud computing environments without the need for copying or downloading. An alternative path to accessing data sets requiring additional protection is the model-to-data approach. In model-to-data, researchers submit algorithms to run on secure data sets that remain hidden. Model-to-data is designed to enhance security and local control while enabling communities of researchers to generate new knowledge from sequestered data. Model-to-data has not yet been widely implemented, but pilots have demonstrated its utility when technical or legal constraints preclude other methods of sharing. We argue that model-to-data can make a valuable addition to our data sharing arsenal, with 2 caveats. First, model-to-data should only be adopted where necessary to supplement rather than replace existing data-sharing approaches given that it requires significant resource commitments from data stewards and limits scientific freedom, reproducibility, and scalability. Second, although model-to-data reduces concerns over data privacy and loss of local control when sharing clinical data, it is not an ethical panacea. Data stewards will remain hesitant to adopt model-to-data approaches without guidance on how to do so responsibly. To address this gap, we explored how commitments to open science, reproducibility, security, respect for data subjects, and research ethics oversight must be re-evaluated in a model-to-data context.

(J Med Internet Res 2020;22(7):e18087) doi:10.2196/18087

KEYWORDS

data management; privacy; ethics, research; data science; machine learning

Introduction

Sharing health data is essential to accelerate knowledge and discovery through opportunities for replication, validation, meta-analysis, and creative reuse [1]. Indeed, many research funding agencies, institutions, and scientific journals encourage or even require the disclosure of research data to the broader scientific community as a means to foster collaborations and to increase scientific accountability, transparency, and reproducibility [2]. Traditionally, data generated as part of research projects or routine clinical care are shared with the scientific community by means of direct download, with the data recipients analyzing data in their local computing environments. However, it is impractical to share some data sets in this manner because of their large size or because of legal restrictions on transfer between institutions or across sovereign borders. One common category of legal restriction is the confidentiality requirement applicable to health information, such as those required in the United States by the *Health Insurance Portability and Accountability Act* [3]. Another category of legal restriction is the limitations on cross-border transfers of certain data. For example, the European Union restricts the transfer of personal data to external countries under its *General Data Protection Regulation2016/679* (GDPR) [4].

XSL•FO
RenderX

Finland's *Secondary Use Act* requires health care data, including nonidentifiable data, to be processed in a secure data center based in Finland [5]. For more examples internationally, see International Compilation of Human Research Standards [6]. Furthermore, researchers and health care institutions who generate data may be concerned over relinquishing data assets given their perceived commercial or academic value, the costs of organizing and annotating data, the confidentiality and security of individual-level data, and loss of oversight of future data use.

The traditional data sharing governance role of health care institutions has been as custodians of data, whose primary responsibility was to keep data secure and confidential. This was achieved through silos, as illustrated by the patchwork of distinct health and accounting records one accumulates when navigating health care systems from hospitals to specialty clinics. The move to open data governance for broadening data sharing practices since 2000 represents a shift of that philosophy, one in which the community of data recipients is collectively responsible for protecting and maintaining the integrity of the data [7]. Open data is attractive in some ways, precisely because it externalizes the very real financial and logistical costs of data governance while providing opportunities for new research perspectives. However, open data governance may be insufficient for many holders of clinical data sets with individual-level privacy or intellectual property concerns.

In between data custodianship and open data governance, data stewardship is an institutional commitment to maximizing the organizational, scientific, and societal benefits of data sharing when also protecting data against privacy and security breaches and misuse [8]. Data stewards may be data generators themselves (eg, hospitals or research institutions) or an honest broker who acts as an independent mediator and trusted partner on behalf of one or more data generators and data users. An honest broker is generally mandated contractually or otherwise to protect and manage secure access to data under the ultimate legal control of another organization.

Data stewards (whether data generators or honest brokers) can now adopt different technical models for making data accessible to users. Typically, a data steward provides access by transferring copies of data to users. This copy-and-download model, however, raises concerns about unaccountable data management and use. Two alternative models promise greater security. In a researcher-to-data approach, a steward makes data available to users within a secure and auditable (cloud) computing environment. In a model-to-data approach, users can submit queries or algorithms to run on secure, hidden data. Each data access model involves different divisions of costs between stewards and users as well as tradeoffs between data use and data protection.

The first part of our paper compares these models primarily in the context of a single health data resource, controlled by a single entity (although this could be a pooled or centralized resource). We highlight the challenges that arise when attempting to scale each data access model beyond single health data resources to networks of multiple resources. The second part of our article focuses on the model-to-data approach. Admittedly, there are only a few existing implementations of model-to-data, and all these essentially involve single resources. Scaling model-to-data to connect multiple resources requires the establishment of a federated data system and, at this time, these systems remain to be largely theoretical. We argued that model-to-data can be a valuable addition to our data sharing arsenal, but it should only be adopted where truly necessary. This approach should supplement rather than replace existing data sharing approaches for the following reasons. First, model-to-data tends to limit scientific freedom and reproducibility. Second, although model-to-data reduces concerns over data privacy and loss of local control when sharing clinical data, it is not an ethical panacea. Model-to-data approaches require just as much, if not more, attention to matters of ethical and legal governance as other data sharing approaches. Third, furnishing the appropriate infrastructure and expertise for a model-to-data approach requires a significant and sustained investment of resources. Misaligned incentives remain to be a major barrier to data sharing, generally [1]. The challenges of aligning incentives are likely to be exacerbated for model-to-data, given the associated expenses of hosting both data and analyses. Finally, all the aforementioned scientific, legal-ethical, and resource challenges also threaten the scalability of model-to-data from single data resources to networks of multiple resources (federated data systems).

Technical Data Access Models

Data sharing has long been synonymous with a copy-and-download approach, where data stewards transfer copies of data sets to researchers. The resulting loss of control over data raises a range of privacy, credit, and proprietary concerns. Access governance mechanisms, such as due diligence review of access requests by a data access committee and data access agreements, can mitigate these concerns to some degree, but may limit access by researchers who do not have a recognized institutional affiliation. Data recipients must also have the expertise and means to provide data hosting, management, and analysis environment. Emerging technical data sharing models aim to further alleviate tradeoffs between data use and data protection (Figure 1). In a researcher-to-data approach, instead of transferring a copy of the data to data recipients, researchers programmatically access data sets that are maintained in a secure computing environment. This approach arose from the development of cloud computing. Cloud cyberinfrastructure has successfully expanded programmatic access to data and analysis tools to a global community of researchers, in many cases cheaply and efficiently, without the cost and logistical challenges of data transfer [9]. These secure computing environments are sometimes characterized as virtual data enclaves or data safe havens [10,11]. A data safe haven implies a collective resource kept in a secure computing environment and managed with appropriate ethical and legal governance for the mutual benefit of individuals, communities, and the society. Data safe havens can enable investigators to conduct exploratory research or can limit investigators to conduct a preapproved analysis, but only results, not individual data, are permitted to leave [11]. An example is the Centers for Medicare & Medicaid Services

XSL•FO
RenderX

Virtual Research Data Center, which allows research on Medicare and some Medicaid claims [12].

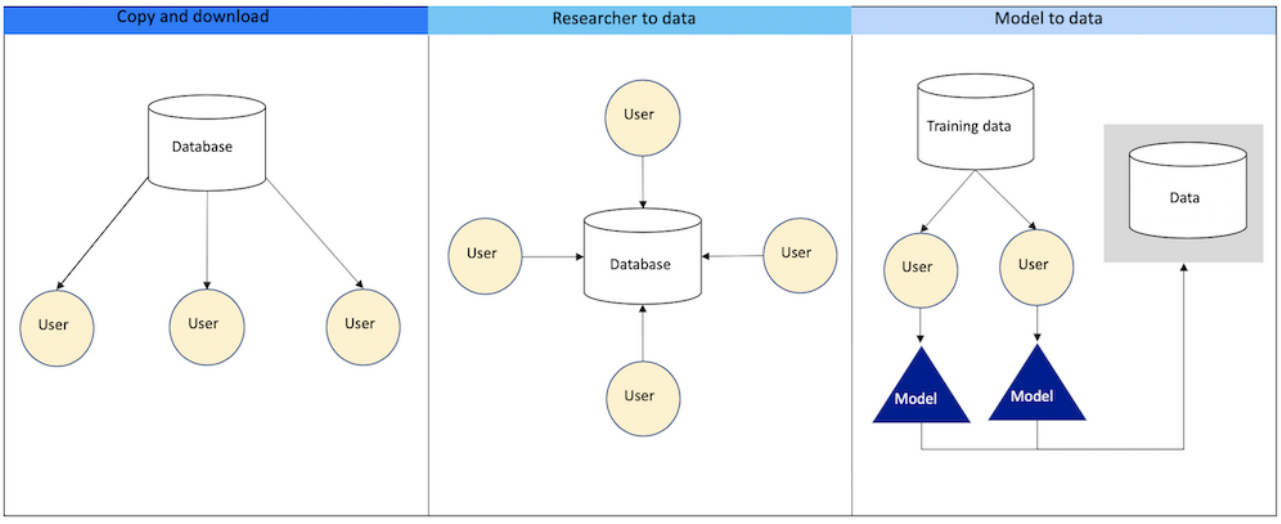
Data safe havens address the costs, risks, and logistical challenges of sharing large, sensitive data sets by obviating the need for copying and distributing the data. Additionally, the cloud platform may allow researchers without intensive local computing resources to run complex analytical models in rented cloud computing infrastructures. Nonetheless, data safe havens do not fully resolve concerns over privacy leaks and misuse because data users must still be trusted to some degree to keep the data within the secure environment and to use the data appropriately. However, these risks can be reduced through access oversight, data use agreements, and active monitoring. Finally, because users can directly interact with data in secure environments, data stewards can share at least some of the costs of data hosting and curation with them.

A comparison of the technical data access models has been provided in Table 1.

A more secure approach has been piloted to allow researchers to extract knowledge from sensitive data sets that remain sequestered. In a model-to-data approach, researchers submit their analytic code or model to a data steward, who maintains the restricted data set in a secure computing environment. The data steward runs the code and returns outputs (eg, summary or performance statistics) to researchers. Data are not moved or even made directly accessible to data users. Model-to-data is made possible by container technologies (eg, Docker) that simplify the bundling and transfer of software models and their dependencies across computing platforms [13]. The utility of model-to-data has been successfully demonstrated in crowdsourced research competitions involving confidential medical data [14,15]. For example, the Digital Mammography DREAM Challenge enabled the analysis of >640,000 deidentified digital mammography images from >86,000 individuals, without transferring or providing direct access to the images [13]. Lisa Austin and David Lie incorporate a version of model-to-data as part of a safe sharing sites proposal, where "computations may be performed on the data in a secure and privacy-protective manner without releasing the raw data, and all data sharing is transparent and auditable" [16]. It may be important depending on the context that the safe sharing site is an independent, honest broker. More recently, the Korean OpenData4Covid19 initiative used a model-to-data approach to share novel coronavirus disease–related health care data with researchers worldwide to fight the pandemic. Researchers must submit their analyses to be run by the Korean Data Centre. Only aggregate results are returned [17]. This example demonstrates how model-to-data can allow data to be rapidly shared with a broad community of users than would otherwise be possible.

We discussed the 4 categories of challenges confronting the model-to-data technical data access approach: (1) scalability to multiple resources, (2) scientific governance, (3) legal-ethical governance, and (4) incentives and sustainability. Data stewards must consider each of these challenges to determine if and how a model-to-data approach should be adopted in a particular context.

Figure 1. Common Data Sharing Models.



XSL•FO
RenderX

Table 1. Comparison of technical data access models.

Characteristics	Copy-and-download	Researcher-to-data	Model-to-data
Costs of data curation	Shared between data steward and data user	Primarily borne by data steward	Completely borne by data steward
Cost of computing infrastructure	Borne by data user	Borne by data steward (users may be charged for compute)	Borne by data steward (users may be charged for compute)
Researcher freedom (eg, methods and tools)	Greatest degree of freedom. Limited by terms of use	Limited freedom. Limited by computing infrastructure and terms of use	Least freedom. Limited by API[a] structure and computing infrastructure
Security and confidentiality of *individual-level data*	Weak. Many copies of data shared with many users, possibly in many countries, subject only to data access agreements. Difficult to audit. Difficult to withdraw data once shared	Strong. Data shared with third party users within a secure and auditable computing environment. Data can be easily withdrawn	Very strong. Data remains hidden from users. Data can be withdrawn at any time
Data privacy protection	Weak. Deidentification and data access agreement limiting reidentification	Strong. Individual-level data may be viewable but not downloadable. Only results are released. Outputs may need to be deidentified	Very strong. Individual-level data are not downloadable or viewable. Only results are released
Security and confidentiality of *Researchers' Ideas or Code*	Very strong. Researchers submit proposals but maintain control over methods	Medium. Researchers' activities are supervised/audited	Weak. Researchers must share query/workflow with data steward
Informed consent	Consent may be needed for research use, sharing, and cross-border transfer	Consent may be needed for research use only	Consent may be needed for research use only
Research ethics oversight	Data user may need research ethics approval	Data steward may need to provide research ethics approval	Data steward may need to provide research ethics approval
Scalability to multiple resources	Straightforward through a distributed data commons, with some shared infrastructure (eg, access portal)	Can only be done indirectly through an individual-level meta-analysis	Difficult but theoretically possible through a federated data system
Legal agreements	Data access/transfer agreement and data use agreement	Data use agreement and computing environment terms of use	Data use agreement only

[a]API: application programming interface.

Scalability of Model-to-Data to a Network of Multiple Resources

Thus far, we have only considered the case of providing access to a single data resource held by a single data steward. Much of the value of data sharing occurs, however, where multiple data resources are connected into a network. Researchers often seek to achieve greater statistical power or refine an algorithm by aggregating multiple data sets. This is particularly relevant to research requiring a comprehensive analysis of data sets at scale. For instance, efforts in achieving precision medicine rely on the ability to combine a vast amount of diverse data types to investigate health and wellness [18]. There is also a strong interest in combining data across sectors to improve knowledge. For example, AllIn partners are looking for solutions to intersect data from diverse community resources to improve community health outcomes [19]. However, combining data from diverse sources is challenging because of the lack of interoperable standards between various data models (eg, OMOP, PCORnet, and TrinetX) [20].

In addition, data sharing networks or consortia must agree on a shared technical solution, raising additional complexity [21]. Consortia of data stewards can agree to centralize or pool their data into a single repository governed by common ethical and legal governance rules and processes (eg, the *All of Us* Research Program [22]). Once the data are centralized, the data stewards must still decide on a technical data sharing model (ie, copy-and-download, researcher-to-data, or model-to-data). Where data distribution is problematic, a researcher-to-data approach may naturally align with a centralized consortium model. The aggregated data can be hosted in a single, secure computing environment. The stewards of the repository can then grant qualified researchers access to data and tools within the secure environment. Admittedly, centralization still involves relinquishing local control and trusting a third party to govern the data. However, these concerns can still be mitigated by adopting a researcher-to-data or a model-to-data approach as copies of data would not be further distributed to researchers.

Where data cannot be shared or centralized, model-to-data can still theoretically enable a virtual form of data aggregation through a federated data system (ie, distributed data system). A federated data system is composed of a network of autonomous data repositories or nodes that share a common data structure/schema and governance principles, but the data remains localized. Users can run an identical query or analysis at each data repository [23]. The results of the queries or analyses are aggregated and returned to the researcher who submitted the query. A federated data system enabling linkage and analysis of sensitive health care data in multiple repositories is proposed by the European Personal Health Train, where *analytical tasks visit data sources* [24]. Data sets are likened to

XSL•FO
RenderX

stations, algorithms are likened to the payload delivered by a *train*, and the network is likened to the train *track*. However, because of the variety of challenges explored below, federated data systems remain to be largely theoretical at this time.

Federated data systems potentially provide high levels of local security, control, and sovereignty as data remain hidden at each node in the network. They also potentially allow for the scalability of some analyses across multiple nodes into large, virtual cohorts. In addition, federated data systems can share ethical and legal governance standards (eg, access and use policies) and associated technical infrastructure like user authentication and authorization, such as the researcher passports infrastructure proposed by the Global Alliance for Genomics and Health [25]. Federated data systems can, in theory, also allow for virtual analysis of international cohorts without sacrificing compliance with data localization requirements, such as the cross-border transfer restrictions imposed by the European GDPR [26,27]. With model-to-data, only the results of algorithmic queries, and not the underlying individual-level data, need to leave a particular territory.

Challenges hampering the implementation of federated data systems include reaching and maintaining "agreement and commitment from federation members on the shared goals, operating principles, metrics for success and apportionment of the benefits and commercial value" [23]. Negotiation is also needed to determine what aspects of the data sharing architecture will be common versus what parts will be decentralized, and who will pay for building and maintaining the common aspects. Given that finding consensus solutions about these aspects in large groups can be difficult, large federated data systems may need to appoint a governing body to facilitate decisions among the federated members and designate data stewards at each node to implement the federation's resolutions. In a federated data system, each node is accountable to all other nodes. Thus, the effectiveness of governance may be inversely proportional to the number of nodes. In short, although complicated, large virtual cohorts can still theoretically be achieved through federated analyses, where algorithms run across multiple secure databases, with aggregation of results.

Many projects are piloting this federated data system approach. For example, the World Economic Forum aims to connect genomic databases from Canada, Australia, the United States, and the United Kingdom to enable queries of rare disease patients. Two European-funded Horizon 2020 projects implementing the federated data system approach include (1) the Research on European Children study and the Adults Born Preterm study, which connects multiple European cohort studies and aims to improve health, development, and quality of life for preterm and low-birth weight individuals [28,29] and (2) EUCAN-Connect, which aims to connect genomic databases to advance precision medicine [30].

Although an attractive alternative to the copy-and-download and researcher-to-data approaches, model-to-data, including federated model-to-data, highlights key scientific, ethical, and legal governance issues that must be addressed as we balance scientific innovation and discovery with data protection.

The current lack of guidance may discourage the real-world implementation of model-to-data solutions. We aim to address this gap by identifying key points to consider for data stewards, focusing on a single data resource model-to-data approach. We conclude that model-to-data should only be adopted in limited circumstances, as a complement to rather than as a substitute for existing data sharing models. This argument applies by extension to the more complicated case of federated data systems involving a network of multiple resources made available through model-to-data.

Scientific Governance of Model-to-Data

The effectiveness of any data sharing model is a function of the type, quality, and richness of the data; the number of problem solvers who can access the data; the diversity of questions and analysis methods that can be applied; and the ability to validate and compare outputs. Modern data sharing governance philosophy has been popularly articulated according to the *FAIR guiding principles* that data be findable, accessible, interoperable, and reusable [31]. The FAIRness of data is always heavily dependent on the actions and abilities of the data steward. FAIRness is all the more dependent on stewards in the case of model-to-data as the data user has no ability to curate or conduct quality assurance on the data. In this paper, we focus on the case of a single data resource. Many of these issues will be challenging in the case of a federated data system involving multiple resources, as participating resources will all have to agree on standards and cost-sharing arrangements.

Findable

To be findable, data resources must be associated with rich, accurate, and standardized metadata. Metadata must generally be openly accessible for researchers to discover relevant data sets through data search engines. New technologies such as genomic Beacons and clinical Beacons allow researchers to search for individual-level traits, outcomes, or genetic variants without revealing detailed individual-level data [32]. Beacon technology works through an application programming interface (API), allowing researchers to query a data set ("do you have any patients with genetic variant ABC?") and receive a simple *Yes* or *No* in response (or some slightly more detailed demographic or clinical information about the individual). Individual-level search may be equally possible across different data sharing models, with appropriate protections in place to ensure that searches do not leak excessive information about individuals. In model-to-data, enabling meaningful discovery may also call on data stewards to provide sufficient metadata or, better, an unbiased, representative training (testbed or sandbox) data set to help users understand the characteristics of the sequestered data set or to develop, train, and test their workflows and machine learning algorithms. The selection, composition, curation, and annotation of the representative training data sets is critical to the success of model-to-data as the training data set is used to *fit* models. Curation of the training (visible) and validation (sequestered) data sets must match perfectly for the data to be truly *findable*.

XSL•FO
RenderX

Accessible

Model-to-data enables analyses of sequestered data that cannot be directly accessed by researchers. At least superficially, the model-to-data approach is one way to resolve tensions between openness and confidentiality. In model-to-data, the data steward does not focus on securely transferring data to trustworthy and qualified users or on controlling access to a cloud computing environment, yet the data steward may still play an important governance and scientific oversight role. Determinations may still need to be made about who is authorized to analyze the data (ie, who can submit code to run on the data), and these determinations would need to be enforced through defined user authentication and authorization processes. In this manner, the model-to-data data steward will need to play a role akin to traditional data access committees in the review of data user access requests. Data stewards may also need to take on a more hands-on scientific advisory role, reviewing the appropriateness and scientific validity of the submitted analysis codes, providing feedback on model performance, and troubleshooting. Further complicating the responsibilities of the model-to-data data steward is the reality of finite computing and advisory resources. Because data are not directly accessible, model-to-data may enable data stewards to expand access to a wider group of data users than would otherwise be allowed because of reduced concerns over security and privacy. However, broadening access would need to be balanced with the need to be parsimonious against limited resources.

Interoperable

The responsibility for making data interoperable in model-to-data rests primarily with the data stewards, as data users cannot access the data directly. Furthermore, data stewards must ensure that their computing environment is interoperable to enable the third-party code to run accurately. This need for interoperable platforms is encouraging harmonization of not only data collection approaches but also technology architectures, data ontologies, formats, and governance structures. Indeed, the Global Alliance for Genomics and Health [33] has proposed a federated international ecosystem of genomic databases for which it is developing technical standards and policy frames to make the data available to authorized data users under relevant conditions [34]. The GA4GH Cloud Work Stream developed standardized computing routines, protocols, and tools (APIs) to enable the portability of algorithms, tools, and workflows across cloud environments in large-scale distributed projects [35]. These standards are being piloted in GA4GH Driver Projects, including the National Cancer Institute Genomic Data Commons [33,36] and the Canadian Distributed Infrastructure for Genomics [37]. These ongoing efforts to harmonize data ontologies, formats, and workflows support the interoperability of data sets both in the cloud as well as in the model-to-data research.

Reusable

The goal of data reuse is to enable the generation of new knowledge from data and to enable validation of previous findings to promote scientific consensus. In order for data to be reusable through model-to-data, data stewards must ensure that the data are high quality, rich, and fit-for-purpose. Preparing data for analysis is often the most taxing part of data science [38]. Regardless of the data sharing model, data generators and stewards bear a significant degree of responsibility for data curation. There are also efforts to harmonize clinical assays as evidenced by the proposed *Diagnostic Accuracy and Innovation Act* [39]. Where copies of data are distributed, researchers can participate to various degrees in curating data to render it usable for specific purposes. In model-to-data, however, the responsibility for curation and annotation of the sequestered data set is entirely the responsibility of data stewards.

Ensuring the Quality of Research Outputs

A concern for data sharing generally is not only ensuring the quality of data but also the quality of research outputs. These results are usually in the form of summary statistics, but increasingly data science outputs also include trained algorithms. Usually, the quality of research outputs is left to quality control by journals and peer review. Owing to their intensive role in data curation and algorithmic facilitation, in some cases, model-to-data data stewards may insist on coauthorship or may reserve the right to review abstracts or manuscripts before publication, although these approaches can raise concerns about scientific freedom.

Ethical and Legal Governance of Model-to-Data

Considerations for ethical data sharing go beyond scientific concerns. Data stewards adopting a model-to-data approach must also address issues of privacy, confidentiality, security, ownership, research ethics oversight, and sustainability.

Privacy, Confidentiality, and Security of Patients' Data

Concerns about risks to privacy are at the core of resistance to data sharing. Data sharing models involving the distribution of data to researchers rely heavily on deidentification (the process of removing direct identifiers). However, the residual reidentification risk of even limited genetic or demographic data sets has been clearly demonstrated [40]. Data access governance mechanisms (eg, due diligence review of access requestors, data access agreements) may mitigate risks, but ultimately depend on trust [41]. There are few means available to monitor obligations to keep data confidential and secure and, even if a breach is detected, it is not clear if and how a breach could be sanctioned. In model-to-data, by contrast, no copies of data are distributed, reducing privacy risks. Data stewards do not have to predict all possible future reidentification scenarios or assess the trustworthiness of data requesters to keep data secure when they enable model-to-data research.

However, there remain some limited privacy concerns regarding model-to-data. First, to make model-to-data data sets usable, limited training sets of data may need to be distributed. It may be possible to distribute synthetic or *noisy* training data sets as a way to preserve the data subject's privacy and data confidentiality. Second, privacy risks may differ depending on whether model-to-data access is provided for algorithm training, validation, or application. Unintended memorization of data has been shown to occur during the training of algorithms [42]. Outputs from applying a model could contain identifiable

XSL•FO
RenderX

information, or such information may be reverse engineered or inferred. Privacy leaks can be reduced by adding safeguards such as auditing of outputs and limiting what kinds of outputs may be returned to researchers. Even when the release is limited to outputs or knowledge, there may still be a need for careful risk assessment and calibration between openness and privacy protection.

Informed Consent and Research Ethics Oversight

Whereas model-to-data may allow data stewards to meet data confidentiality and localization requirements, *ethical* concerns remain related to the principles of informed consent and research ethics oversight. Ethically, patients have an interest in knowing who is being provided access to their data and what their data are being used for, even if those parties never directly access their individual-level data. Data privacy laws, particularly the European GDPR [4], require a legal basis, such as consent or public interest, to be able to analyze personal data for certain research purposes. The US National Institutes of Health's dbGAP genomic data repository imposes ethical data use limitations on data based on the scope of the participant's consent [43]. With model-to-data, *legal* constraints on use may appear to be diminished or inapplicable as individual-level data are not disclosed to third parties. Nevertheless, data use may raise ethical and social license concerns if data were used by controversial parties or for controversial purposes without patients' knowledge. Data stewards are responsible for ensuring appropriate oversight of data use. This responsibility overlaps with the scientific oversight mechanisms discussed above.

A traditional mechanism to ensure the ethical conduct of health research is to subject project proposals to approval and oversight by an independent research ethics oversight body (eg, Institutional Review Board, Research Ethics Committee). In a model-to-data context, however, the locus of research ethics oversight may need to shift from the data user's institution to the data steward, where the actual analysis of data is performed. This shift may in fact allow access to a wider range of data users from outside academic or health care institutions—who would not necessarily be subject to human subject research regulations—to analyze data (eg, commercial researchers and citizen scientists). The data steward would essentially provide a research ethics review as a service to these users. In the case of a federated data system involving multiple data stewards, some level of coordinated research ethics oversight between stewards would need to be developed. For example, a data user could receive approval from one data steward in a network and the other stewards could mutually recognize that approval. Fortunately, models for coordinated oversight of data-intensive research have already been developed and could be adapted to the model-to-data context [41,44].

Confidentiality and Security of Researchers' Ideas and Software

A central strength of model-to-data is that it improves data security. However, the trust patterns required to enable security in model-to-data research are novel. In traditional data sharing models, data subjects and data generators must trust data users to secure individual-level data. In model-to-data, by contrast, data users must trust that data stewards will keep their queries and code secure from unauthorized access and that data stewards will not appropriate the ideas or internet protocols of data users for themselves. This issue becomes more problematic when scaling model-to-data to multiple resources. From this perspective, allowing an honest broker to act as the data steward in model-to-data may be desirable (although this would require data generators to transfer control over their sensitive data to the honest broker). Involving an independent trusted third party can reduce real or perceived conflicts of interest. Another potential solution analogous to privacy-preserving record linkage is to enable secure computation without the data steward being able to directly inspect the query or code [45,46]. The query or code is encrypted and sent to a trusted third party to unencrypt at runtime. However, this may be technically complex and data stewards may insist on inspecting the nature of the submitted algorithm. Model-to-data stewards must also consider the security risks associated with allowing outside code to be run inside their protected computing environment (eg, tampering). Providing read-only access to data directories limits the risks that input data may be altered. Preventing network access while executing the analysis code reduces the risk that users may accidentally download data.

Data Sharing Incentives and Sustainability

The FAIRness of data in model-to-data contexts is heavily dependent on the role of the data steward. The FAIRness of analysis workflows is also increasingly key for the reproducibility of data-intensive health research. In the context of model-to-data, this FAIRness may additionally be essential for trust between data users and data stewards [24,47]. Data stewards need to possess not only *traditional* data governance expertise but also significant computational and analysis skills, a rare combination, and one without formally recognized qualifications at this time. These issues of oversight may hamper the adoption of model-to-data. Ultimately, by shifting more responsibility for data curation and infrastructure costs to data stewards, model-to-data represents a fundamental recalibration of the tripartite responsibilities of data generators, data users, and funders envisaged by the Toronto Statement on Pre-publication Data Sharing, a policy promoting rapid data sharing by large-scale community resource projects in biology [48]. Although model-to-data may seem like a technical decision, it significantly exacerbates the incentive problem already plaguing data sharing [49]. To succeed, model-to-data requires dedicated support from funders as well as from the highest level of leadership within organizations seeking to share data.

Similarly, all databases face sustainability challenges [50]. Who pays for the development, management, and maintenance of data sets, especially over time? For model-to-data, similar questions must also be asked about the sustainability of maintaining computing environments and security infrastructure. Open data distribution models promise a cheap solution to the preservation of data sets in part through the existence of numerous copies of the data set. In model-to-data, a data steward is trusted with preserving the data set and providing access to

XSL•FO
RenderX

it over time. Different sustainability models may need to be explored, from fee-for-access to sustained public funding. The choice of model may be influenced by the source of investment in data generation, curation, and infrastructure. Private investors may want to commercialize data access, whereas public funders may insist on an open science approach. One solution could be a commitment by funders to pay for stewardship as a utility contract, one that simultaneously sustains infrastructure and explicitly requires strict platform neutrality. Good contracting practices that guarantee the portability of data sets from steward to steward in the event of bankruptcy or bad faith would also be essential. Furthermore, data stewards must be careful to avoid conflicts of interest toward either the data generators who entrust them with the data or data users who entrust their hypothesis and analysis methods to them. Ultimately, the question for model-to-data approaches and, by extension, federated data systems, is whether the benefits to science and society are worth the cost. The first-of-its-kind economic evaluation of federated data systems concluded that their return on investment was unclear [51].

The major concerns related to sustainability in model-to-data are the evaluation and reproducibility of model-to-data research, the key to research quality, and transparency. If a data steward were to refuse or discontinue algorithmic access services, this would imperil research reproducibility. This issue becomes greater at scale. In a federated data system, all it takes is for one data steward in a network to withdraw algorithmic access to undermine reproducibility. The reproducibility of model-to-data research may be vulnerable to the caprices of funders, scientific priorities, politics, or even individual data subjects. Indeed, because there is only one copy of a data subject's data in model-to-data, it may also be easier for subjects to withdraw their data. On the one hand, model-to-data supports individual autonomy, allowing individuals to control how their data are (or are not) shared and used over time, which may bolster public trust and willingness to share. On the other hand, this may make reproducibility of research more susceptible to the caprices of individuals.

Conclusions

Model-to-data promises to unlock health data currently stuck behind institutional (fire)walls and country borders, providing access to researchers without sacrificing security, confidentiality, or local control. Model-to-data can improve the quality of research outputs by providing access to more data.

For algorithms, model-to-data access can serve as a means of rapid, continuous quality control. The accuracy of algorithms can be determined or compared through benchmarking, where the algorithm is run on a standard data set. For traditional research findings, model-to-data may allow peers to reach consensus about insights if running different models generates similar findings. In practice, however, this technical access model involves significant costs and drawbacks. It falls short of its utopian promise of technology, overcoming legal and human constraints on data sharing. We have argued that model-to-data should only be adopted where necessary to supplement rather than to replace existing data sharing approaches for scientific, legal-ethical, and resource reasons. By segregating researchers from data, model-to-data tends to limit scientific freedom, integrity, and reproducibility. Although model-to-data scores high on security and confidentiality, ethical concerns persist about individual control over who uses their data and for what purposes. Model-to-data also presents a coordination puzzle for research ethics oversight systems, with the research analysis occurring in an institution different from that of the researcher.

Model-to-data intensifies the incentive issues already plaguing open science [49]. Data stewards must provide expertise and resources to curate data, maintain technical infrastructure, and conduct scientific activities on behalf of users. Meanwhile, direct benefits tend to accrue to data users. Scalability is another concern for model-to-data. Scalability from a single resource to multiple resources is possible with a federated data system. However, as these systems scale, the challenges of maintaining scientific data harmonization, technical standardization, and agreements about cost and benefit sharing scale as well. The scale of federated data systems may therefore be naturally limited to a smaller number of trusted parties.

We predict that the model-to-data approach will become an increasingly attractive option for particular use cases as the scale, diversity, and regulatory complexity of international data sharing continue to rapidly increase. This assumes that the scientific, legal-ethical, incentives, and network issues described above are addressed. Model-to-data will be most useful for large-scale stand-alone resources or small networks of resources when scientific data standards are well established, where the inquiry is clearly defined in advance (eg, in a hackathon challenge), and when providing access to a specific community of users who would not otherwise be authorized. Ultimately, the success of model-to-data may require the migration of expertise in data curation, data science, and technical interoperability from institutions hosting researchers to those hosting data. It may further depend on the willingness of funders to support the creation and sustainability of data centers and their networks. Although expensive for single data stewards, model-to-data may be more efficient overall by reducing the need for redundant, secure computing environments.

We are cautiously optimistic that model-to-data can unlock at least some data sets for broader sharing, for at least some uses, but we remain skeptical that it can fully square the circle. To successfully adopt an model-to-data approach, the concept of data stewardship needs to be reconfigured, with implications that reach the organization's core business model. At least for now, responsible scientific, ethical, and legal data governance remains a reliably human endeavor.

XSL•FO
RenderX

Acknowledgments

AT and BK acknowledge funding support from Genome Canada, Genome Quebec, and the Canadian Institutes of Health Research.

Conflicts of Interest

None declared.

References

1. National Academy of Science, Engineering, and Medicine. Open Science by Design: Realizing a Vision for 21st Century Research. Washington (DC): National Academies Press (US); 2018.
2. Research Funders' Open Access Policies. Sherpa Services. URL: https://v2.sherpa.ac.uk/ [accessed 2020-01-09]
3. The HIPAA Privacy Rule. The US Department of Health and Human Services (HHS). 2015. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html [accessed 2020-05-07]
4. Data Protection in the EU. European Commission. 2017. URL: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en [accessed 2020-05-07]
5. Secondary Use of Health and Social Data. Finnish Ministry of Social Affairs And Health. URL: https://stm.fi/en/secondary-use-of-health-and-social-data [accessed 2020-05-07]
6. International Compilation of Human Research Standards. The US Department of Health and Human Services (HHS). URL: https://www.hhs.gov/ohrp/international/compilation-human-research-standards/index.html [accessed 2020-05-07]
7. Kim J. Overview of disciplinary data sharing practices and promotion of open data in science. Sci Educ 2019 Feb 20;6(1):3-9. [doi: 10.6087/kcse.149]
8. Council of Canadian Academies. URL: https://cca-reports.ca [accessed 2020-01-09]
9. Stein LD, Knoppers BM, Campbell P, Getz G, Korbel JO. Data analysis: create a cloud commons. Nature 2015 Jul 9;523(7559):149-151. [doi: 10.1038/523149a] [Medline: 26156357]
10. Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, et al. Data safe havens in health research and healthcare. Bioinformatics 2015 Oct 15;31(20):3241-3248 [FREE Full text] [doi: 10.1093/bioinformatics/btv279] [Medline: 26112289]
11. Platt R, Lieu T. Data enclaves for sharing information derived from clinical and administrative data. J Am Med Assoc 2018 Aug 28;320(8):753-754. [doi: 10.1001/jama.2018.9342] [Medline: 30083726]
12. CMS Virtual Research Data Center (VRDC). ResDAC. URL: https://www.resdac.org/cms-virtual-research-data-center-vrdc [accessed 2020-05-07]
13. Guinney J, Saez-Rodriguez J. Alternative models for sharing confidential biomedical data. Nat Biotechnol 2018 May 9;36(5):391-392. [doi: 10.1038/nbt.4128] [Medline: 29734317]
14. Ellrott K, Buchanan A, Creason A, Mason M, Schaffter T, Hoff B, et al. Reproducible biomedical benchmarking in the cloud: lessons from crowd-sourced data challenges. Genome Biol 2019 Sep 10;20(1):195 [FREE Full text] [doi: 10.1186/s13059-019-1794-0] [Medline: 31506093]
15. Mason M, Amatangelo M, Auclair D, Bassett D, Dai H, Dervan A. Abstract 4725: multiple myeloma dream challenge: a crowd-sourced challenge to improve identification of high-risk patients. Clin Res 2017. [doi: 10.1158/1538-7445.am2017-4725]
16. Austin LM, Lie D. Safe Sharing Sites. NYU Law Review. 2019. URL: https://www.nyulawreview.org/issues/volume-94-number-4/safe-sharing-sites/ [accessed 2020-01-09]
17. #opendata4covid19. Health Insurance Review & Assessment Service. URL: https://t.co/bbxDpRGnmJ [accessed 2020-05-08]
18. Schadt EE. The changing privacy landscape in the era of big data. Mol Syst Biol 2012;8:612 [FREE Full text] [doi: 10.1038/msb.2012.47] [Medline: 22968446]
19. All In: Data for Community Health. URL: https://www.allindata.org/about-us/ [accessed 2020-01-10]
20. Garza M, del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform 2016 Dec;64:333-341 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.016] [Medline: 27989817]
21. Contreras JL, Reichman JH. Sharing by design: data and decentralized commons. Science 2015 Dec 11;350(6266):1312-1314 [FREE Full text] [doi: 10.1126/science.aaa7485] [Medline: 26659039]
22. All of Us. National Institutes of Health. URL: https://allofus.nih.gov/ [accessed 2020-01-10]
23. Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data. The World Economic Forum. URL: https://www.weforum.org/whitepapers/federated-data-systems-balancing-innovation-and-trust-in-the-use-of-sensitive-data/ [accessed 2020-01-09]
24. Beyan O, Choudhury A, van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H, et al. Distributed analytics on sensitive medical data: the personal health train. Data Intell 2020 Jan;2(1-2):96-107 [FREE Full text] [doi: 10.1162/dint_a_00032]
25. GA4GH Passports and the Authorization and Authentication Infrastructure. GA4GH. 2019. URL: https://www.ga4gh.org/news/ga4gh-passports-and-the-authorization-and-authentication-infrastructure/ [accessed 2020-01-10]

XSL•FO
RenderX

26. Aarestrup AK, Væver MS, Petersen J, Røhder K, Schiøtz M. An early intervention to promote maternal sensitivity in the perinatal period for women with psychosocial vulnerabilities: study protocol of a randomized controlled trial. BMC Psychol 2020 Apr 28;8(1):41 [FREE Full text] [doi: 10.1186/s40359-020-00407-3] [Medline: 32345375]
27. European Translational Information and Knowledge Management Services: eTRIKS: Deliverable Report. eTRIKS. URL: https://www.etriks.org/wp-content/uploads/115446-eTRIKS-D7.9-Data-sharing-federation-model.pdf [accessed 2020-05-14]
28. Kunz. Empowering Distributed Analysis Across Federated Cohort Data Repositories Adhering to FAIR Principles. ERCIM News 121. 2019. URL: https://ercim-news.ercim.eu/en118/special/empowering-distributed-analysis-across-federated-cohort-data-repositories-adhering-to-fair-principles [accessed 2020-01-10]
29. Research on European Children and Adults Born Preterm. URL: https://recap-preterm.eu/ [accessed 2020-01-10]
30. EUCAN-Connect. URL: https://www.eucanconnect.eu/ [accessed 2020-01-10]
31. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
32. Fiume M, Cupak M, Keenan S, Rambla J, de la Torre S, Dyke SO, et al. Federated discovery and sharing of genomic data using beacons. Nat Biotechnol 2019 Mar;37(3):220-224 [FREE Full text] [doi: 10.1038/s41587-019-0046-x] [Medline: 30833764]
33. The Global Alliance for Genomics and Health (GA4GH). URL: https://www.ga4gh.org/ [accessed 2020-01-09]
34. Global Alliance for Genomics and Health. GENOMICS. A federated ecosystem for sharing genomic, clinical data. Science 2016 Jun 10;352(6291):1278-1280. [doi: 10.1126/science.aaf6162] [Medline: 27284183]
35. GA4GH/Wiki. GitHub. URL: https://github.com/ga4gh/wiki [accessed 2020-01-09]
36. The Next Generation Cancer Knowledge Network. NCI Genomic Data Commons - National Cancer Institute. URL: https://gdc.cancer.gov/ [accessed 2020-01-09]
37. CanDIG. URL: https://github.com/CanDIG [accessed 2020-08-07]
38. Patil D, Mason H. Data Driven. Newton, Massachusetts, United States: O'Reilly Media, Inc; 2015.
39. Schreier J, Feeney R, Keeling P. Diagnostics reform and harmonization of clinical laboratory testing. J Mol Diagn 2019 Sep;21(5):737-745 [FREE Full text] [doi: 10.1016/j.jmoldx.2019.04.002] [Medline: 31075512]
40. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul 23;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]
41. Regulatory & Ethics Toolkit. Global Alliance for Genomics and Health. URL: https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/ [accessed 2020-01-09]
42. Carlini N, Liu C, Erlingsson U, Kos J, Song D. The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks: arXiv; 2018:E.
43. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, Rutter JL, National Institutes of Health Genomic Data Sharing Governance Committees. Data use under the NIH GWAS data sharing policy and future directions. Nat Genet 2014 Sep;46(9):934-938 [FREE Full text] [doi: 10.1038/ng.3062] [Medline: 25162809]
44. Townend D, Dove ES, Nicol D, Bovenberg J, Knoppers BM. Streamlining ethical review of data intensive research. Br Med J 2016 Aug 2;354:i4181. [doi: 10.1136/bmj.i4181] [Medline: 27486077]
45. Baker DB, Knoppers BM, Phillips M, van Enckevort D, Kaufmann P, Lochmuller H, et al. Privacy-preserving linkage of genomic and clinical data sets. IEEE/ACM Trans Comput Biol Bioinform 2019;16(4):1342-1348. [doi: 10.1109/TCBB.2018.2855125] [Medline: 30059313]
46. Oprisanu B, de Cristofaro E. AnoniMME: bringing anonymity to the matchmaker exchange platform for rare disease gene discovery. Bioinformatics 2018 Jul 1;34(13):i160-i168 [FREE Full text] [doi: 10.1093/bioinformatics/bty269] [Medline: 29949992]
47. Lamprecht A, Garcia L, Kuzak M, Martinez C, Arcila R, del Pico E, et al. Towards FAIR principles for research software. DS 2020 Jun 12;3(1):37-59 [FREE Full text] [doi: 10.3233/DS-190026]
48. Toronto International Data Release Workshop Authors, Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, et al. Prepublication data sharing. Nature 2009 Sep 10;461(7261):168-170 [FREE Full text] [doi: 10.1038/461168a] [Medline: 19741685]
49. Longo DL, Drazen JM. Data sharing. N Engl J Med 2016 Jan 21;374(3):276-277. [doi: 10.1056/NEJMe1516564] [Medline: 26789876]
50. Anderson W, Apweiler R, Bateman A, Bauer G, Berman H, Blake J. Towards Coordinated International Support of Core Data Resources for the Life Sciences. bioRxiv. 2017. URL: https://www.biorxiv.org/content/10.1101/110825v3 [accessed 2020-06-29]
51. Dana G, Arnaud B. Global Data Access for Solving Rare Disease: A Health Economics Value Framework. The World Economic Forum. URL: https://www.weforum.org/reports/global-access-for-solving-rare-disease-a-health-economics-value-framework/ [accessed 2020-05-14]

XSL•FO
RenderX

Abbreviations

API: application programming interface
FAIR: findable, accessible, interoperable, reusable
GDPR: General Data Protection Regulation

Edited by G Eysenbach; submitted 02.02.20; peer-reviewed by G Dana, M Abdelhamid, M Lang; comments to author 31.03.20; revised version received 21.05.20; accepted 11.06.20; published 28.07.20.

Please cite as:
Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B
Bringing Code to Data: Do Not Forget Governance
J Med Internet Res 2020;22(7):e18087
URL: http://www.jmir.org/2020/7/e18087/
doi:10.2196/18087
PMID:32540846

©Christine Suver, Adrian Thorogood, Megan Doerr, John Wilbanks, Bartha Knoppers. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Seeking Meaningful Innovation: Lessons Learned Developing, Evaluating, and Implementing the Electronic Patient-Reported Outcome Tool

Carolyn Steele Gray[1,2], MA, PhD

[1]Bridgepoint Collaboratory for Research and Innovation, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Carolyn Steele Gray, MA, PhD
Bridgepoint Collaboratory for Research and Innovation
Lunenfeld-Tanenbaum Research Institute
Sinai Health System
1 Bridgepoint Drive
Toronto, ON, M4M 2B5
Canada
Phone: 1 4164618252 ext 2908
Email: Carolyn.SteeleGray@sinaihealth.ca

Abstract

Digital health solutions, in particular information communication technologies, often experience implementation failures leading to slower adoption than expected. This implementation challenge has spurred the development of frameworks to help navigate this uncertain and messy process. These frameworks point to environmental, organizational, individual, and technological factors that can drive or hinder implementation, with some in the field suggesting that perceived value may play a pivotal role. However, the concept of value can have varying meanings and be challenging to operationalize as a means to support implementation. Attending to philosophical and psychological meaningfulness for users and organizations in which technologies are adopted may offer a useful lens, by linking perceived value to individual behavior changes often required as part of implementing digital health technologies. Lessons learned from developing, evaluating, and implementing the electronic Patient-Reported Outcome (ePRO) tool demonstrate how qualitative methods can be used to uncover meaningfulness. By drawing from this example and other similar studies, this viewpoint offers suggestions on how future inquiry could deepen an understanding of meaningful innovation to help drive the implementation of digital health technologies.

(J Med Internet Res 2020;22(7):e17987) doi:10.2196/17987

KEYWORDS

digital health; implementation; meaningfulness

The Implementation Challenge Facing Digital Health

While digital technologies have transformed many industries worldwide, health systems remain laggards in adopting innovative and disruptive technology. In a recent reflection, Jadad and Jadad Garcia lament the glacial crawl of innovation adoption in health systems, stating that "sadly, the promises of information and communication technologies to transform healthcare services remain unfulfilled" [1]. Global health leaders have recognized this gap and are establishing strategies to support wider adoption of technologies that are viewed as foundational for robust health systems [2,3]. Included in these strategies is attention to overcoming implementation challenges that have slowed adoption. Implementation factors such as adaptability of the innovation, the implementation climate, the policy and system environment, and user characteristics are key hindrances [4,5]. In her viewpoint paper in the 20th anniversary issue of the *Journal of Medical Internet Research*, Buis argues that to achieve the vision of digital health, the next decade of inquiry requires greater attention to these implementation challenges [6].

Several theoretical frameworks and models of implementation have emerged and been applied to digital health. Two promising models include the Value-Proposition Design (VPD) [7], and the Non-adoption, Abandonment, Scale-up, Spread, and

XSL•FO
RenderX

Sustainability (NASSS) framework [8]. These frameworks represent an important step forward in implementation science, offering applied and practical approaches that build on heuristic models like the Consolidated Framework for Implementation Research (CFIR) [9]. VPD and NASSS include attention to multiple implementation factors, and both suggest that the value proposition of a given technology is a critical driver of its implementation, an idea also reflected within the World Health Organization guidelines on adopting digital health solutions [10]. The importance of the perceived and evidenced value of these technologies has spurred calls for better evaluation standards of digital health products to drive adoption [11].

Although perceived value is one of several factors to consider, the above frameworks suggest it to be a central concern as it can *drive behavior change* in the individuals and organizations needed to implement a technology. However, the concept of value can be understood in many ways and through multiple disciplinary lenses. This viewpoint suggests *meaningfulness* as one potentially useful approach to uncover how perceived value can drive implementation.

What is the Meaning of Meaning?

The notion of "meaning" in thinking of technology adoption in healthcare is not a new idea. The concept of "meaningful use" of electronic health records (EHRs) emerged in 2010 in an attempt to establish rules for how EHRs are to be used in order to improve quality and efficiency [12]. While this functionally focused notion of meaning has an important place, it may be too narrow to encompass different interpretations of meaning held by users. A broader understanding is offered through philosophical approaches exploring how individuals perceive that they are living meaningful lives. Some philosophers have argued that we see our lives and actions as being meaningful when we feel we are promoting good in the world, experiencing subjective satisfaction, and/or achieving our aims and life goals [13]. While the intention of this commentary is not to dive into the deep and rich philosophical literature on meaning, this approach to meaning can offer a lens to help explore and perhaps overcome some of the challenges experienced when implementing technology in health care.

Perceived meaningfulness is related to the implementation of digital health solutions in two ways. First, meaningfulness drives behavior, and behavior change is needed when adopting new technology. Social and cognitive psychology theories have shown how group interactional processes and individual cognitive and emotional processes motivate behavior change needed when implementing new practices [14], like when adopting a new electronic medical record or including telemonitoring as part of clinical care. Social determination theory, which suggests individuals' goals and aspirations (one of the components of meaningfulness mentioned earlier), are key drivers of behavior [15]. Social determination theory has also been used to understand patient adherence to digital health interventions [16].

For providers, psychological meaningfulness associated with work has been shown to incent better engagement with that work. As Kahn [17] describes in his foundational paper on the subject, psychological meaningfulness in work is "*the harnessing of organization members' selves to their work roles; in engagement, people employ and express themselves physically, cognitively, and emotionally during role performances*."

Here Kahn connects meaning in work to a broader sense of self, which can, in turn, act as a powerful motivator to drive behavior. Classic motivation theorist Maslow viewed meaningfulness as one of several higher-order needs which humans strive for through the ongoing process of self-actualization [18]. Meaningfulness in work is argued to help individuals reach a level of self-actualization where individuals are driven by the work itself, rather than completing a task. Building on Maslow's theory, Chalofsky and Krishna suggest meaning *in* work occurs when work speaks to an individual's whole self and is perceived to be *good work*. When perceived meaning *in* work is paired with meaning *at* work, where individuals feel emotionally committed to their organizational environment, they will be subsequently more engaged and willing to go "above and beyond" [19]. This feeling of *meaningfulness* in work represents a "deeper level of intrinsic motivation" [19]. Tapping into that level of motivation to change can be crucial in technology implementation, where resistance to change is a significant barrier to implementing digital health solutions [4].

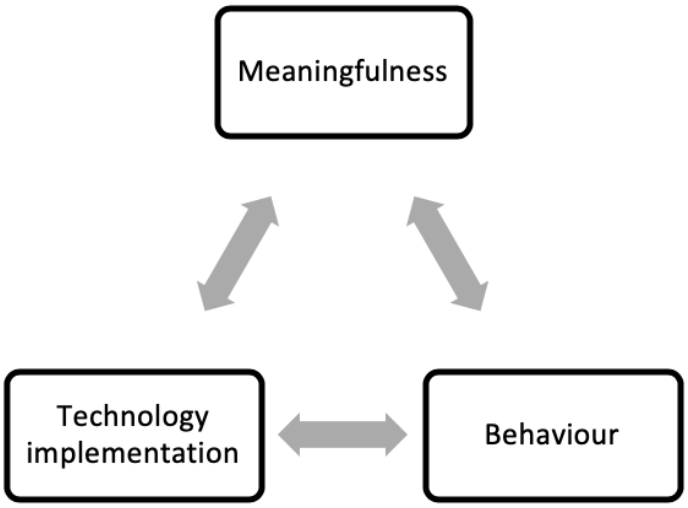
Second, while individuals' sense of meaning in their lives can drive their behavior, it can also be influenced by the technologies they use. Technology itself represents an artifact in the social and work environment that can influence meaningfulness to individuals. There may be a risk in assuming the technology is simply a tool to manipulate, having little to do with what is important to people and their identities. Another perspective, grounded in a sociological approach, suggests technology plays an important role in how we understand ourselves, our work, and our environment. As Lupton so eloquently argues, health technologies are "*sociocultural products located within pre-established circuits of discourse and meaning. [Technologies] are active participants that shape human bodies and selves as part of heterogeneous networks, creating new practices and knowledge*" [20].

As such, when seeking to drive implementation of technology, we not only need to determine what value this brings to the individuals interacting with the technology as suggested by VPD and NASSS, but we also need to unpack how using the technology may change perceptions of value and meaning over time.

Figure 1 offers a simple visualization of this proposed interrelationship between meaningfulness, behavior, and technology implementation.

XSL•FO
RenderX

Figure 1. Meaningfulness in digital health.



Seeking Meaningful Innovation

If meaningfulness is a key driver to implementation, then it is essential to uncover what is meaningful to users of systems within their organizations and system contexts. VPD and NASSS offer a set of questions to guide the implementation process iteratively. These questions are not only useful in uncovering objective opportunities and challenges in implementation (eg, whether there are necessary resources and regulations needed to adopt a particular technology) but can be used to uncover meaningfulness as well (eg, how individuals perceive a shared problem to be solved).

In adopting this approach, however, two challenges can arise. First, those asking the questions may struggle to pick up on meaningfulness because users will have different abilities and willingness to express what is meaningful. Simply put, you may ask the right question, but the answers may not be articulated in a way that meaningfulness can be easily understood. Second, it can be challenging to attend to what is important and meaningful to the many different users of a system which can include providers, administrators, patients, and families involved in health service delivery; all of whom have different professional and personal backgrounds, aims and goals, and personal perceptions of meaning.

Drawing on multiple *qualitative methods* can help not only ask the right questions at the right times but unpack and interpret the answers with an ear for meaningfulness. To demonstrate how this approach could be put into practice, the remaining sections draw on experiences implementing the electronic Patient-Reported Outcome (ePRO) tool, a mobile app and web-based platform designed to support goal-oriented care in primary care settings. Qualitative methods used to develop, evaluate, and implement ePRO across multiple settings offer lessons on how what was meaningful to patients, providers, and their organizations influenced the adoption of technology. Experiences with this technology are compared to other literature emphasizing meaningfulness to drive the implementation of digital health technologies.

Developing for Meaningfulness

A focus on meaningfulness suggests a shift from merely seeing technology as a tool to support a process toward a deeper understanding of a user's experience of receiving or delivering care in a particular context. User-centered co-design is a popular approach to employ the experiences and views of users to guide development [21]. It is essential to recognize that the process requires analytic work and iterative dialogue to link articulated desires for functionality, to why that function is important to users.

In the first rounds of developing the ePRO tool [22-24], interpretive description was used to understand user experiences. This method allows for situated and contextualized analysis with a view toward practical application of findings [25], which supported an iterative building of the interpretation of meaningfulness and drove the development of the technology. After undergoing multiple rounds of co-design, usability analysis found that although the technology had many shortcomings, it was highly used by patients, because they perceived it as meaningful [24].

The ARCHIE framework similarly suggests that the development of telehealth and telecare products must be "anchored in what matters to users" through iterative co-creation driven by a phenomenological lens [26]. The framework aligns with philosophical and sociological traditions that emphasize meaningfulness to users, in this case, patients, suggesting the use of phenomenological and ethnographic approaches to uncover meaningfulness and guide development.

Evaluating for Meaningfulness

As the development of the ePRO tool moved from design into the iterative evaluation of usability testing and pilot trial, the focus on meaningfulness to both patient and provider users became even more important. This phase required the next step of linking meaningfulness to other components of the intervention beyond just the technology, looking at contexts, processes, and mechanisms that drive any change in patient

XSL•FO
RenderX

outcomes. A narrative analysis of patient and provider interview data was used to identify dominant themes in the stories they told of using the technology [27].

Critical to both patient and provider stories was how the technology linked to what was important in their lives and work. For patients, this meant it supported their care goals (achieving personal aims) and reinforced a sense of shared accountability to those aims with their provider. Providers recounted that they were willing to put in more time and energy to learn and use the new technology, adapting workflows and processes, when the technology was perceived as important to their patients (linked to a sense of doing *good work*). When they viewed the technology as meaningful, patients and providers changed their behavior to support the regular use of the tool.

Other evaluation methods that have been applied to similar complex technology interventions have viewed meaningfulness as a central component. For example, Gomersall and colleagues suggest adopting network-based evaluation approaches, like social networking or realist approaches, to pick up on social and individual values and meaningfulness in evaluations of ambient assistive living technologies [28].

Using Meaningfulness to Drive Implementation

The ePRO example and other studies indicate that meaningfulness may play an important role in the implementation of technologies. In these examples, meaningfulness is linked to personally held values and beliefs of individuals interacting with technologies, demonstrating that when technology aligns with those beliefs, it may be more likely to be used. However, meaningfulness is only one factor among other implementation drivers occurring at organizational and system levels. Uncovering meaningfulness for organizations and systems may also be required to support implementation. One approach is to consider how a given technology aligns to the *why*, or the vision, of an organization or care model. The critical point is to not just attend to that vision but also think about the guiding principles and values that drive organizations toward their goal. The disconnect between technology and what providers perceive as being foundational to their work and organization is a significant barrier to implementation [29].

Ideally, technologies would fit the aims and activities of the model of care as well as the guiding principles and values of the organizations and individuals delivering and seeking care. The ePRO tool was co-designed with those engaging in goal-oriented, person-centered care models. When spreading the tool to new environments as part of the pragmatic trial (see protocol [30]), it was most successful in primary care clinics that also had a strong vision of person-centered care delivery along with processes they could adapt to meet that aim. Although we are still analyzing findings from this study, observationally we found that where there was alignment, the ePRO tool was more likely to be perceived as valuable to providers and patients leading to changes in how patients engaged in their care, and how providers managed patients and communicated with each other around supporting patient goals. Preliminary analysis of trial data suggests alignment between technology and the organization's vision and guiding principles can play an important role in implementation. A deep dive into the ethnographic data planned for this year will provide additional insight into this relationship.

Future Work

The experience of developing, evaluating, and implementing the ePRO tool, and the emerging literature in this space suggests that meaningfulness matters. In particular, philosophical and psychological meaningfulness may be playing an important role in the process of implementation. The ideas presented are based on just a few examples and require more empirical testing. While meaningfulness to users was explored as part of the ePRO study, the importance of meaning as related to implementation emerged throughout the study; as such we did not start explicitly looking for connections between meaningfulness and other implementation constructs until we got to later stages and had an opportunity to reflect on what was learned.

This viewpoint is intended to start a conversation about the role of meaningfulness in implementation. Many questions are yet to be explored, including: is there a unique understanding of meaningfulness that pertains to technology? How does meaningfulness interact with other components of implementation? Can strong empirical evidence of the value of a technology overrule intrinsic perceptions of meaning held by users? How does meaningfulness relate to other socially grounded concepts like ethics and equity when implementing technologies?

Perhaps the more important takeaway from this viewpoint is that philosophical and psychological meaningfulness for those who engage with technologies in their work and lives plays a role that deserves attention. Adopting qualitative methods can serve to uncover meaningfulness for diverse users and organizations, and could help to drive implementation decisions. Future work can help to uncover how we can adapt and build research methods to place meaningfulness at the center of our implementation efforts so that perhaps our reflections on the state of digital health a decade from now are much less grim than those we see today.

Acknowledgments

I would like to thank the dozens of wonderful colleagues, volunteer participants, trainees, and mentors with whom I have worked over the past 7 years as part of the ePRO project. It is through this work that the ideas shared in the commentary originated. A particular acknowledgment and deeply felt thank you to Drs Laura Desveaux and James Shaw, who invited me to speak on this topic at the Women's College Hospital Institute for Health System Solutions and Virtual Care (WIHV) Innovations Rounds in

XSL•FO
RenderX

Toronto in October 2019, and later provided thoughtful feedback on this viewpoint. It was in preparing for the talk that the ideas presented in this commentary were first brought together, presented, and then discussed with a terrific and engaged audience. I would also like to acknowledge Drs Ross Upshur and Kerry Kuluski, who offered support as I put thoughts together for this viewpoint. While this specific commentary was not explicitly funded, the ePRO project has been funded through the Canadian Institute for Health Research (FN-143559; FN-137200) and the Ontario Ministry of Health and Long-term Care Health Research Fund (#06034).

Conflicts of Interest

None declared.

References

1. Jadad AR, Jadad Garcia TM. From a Digital Bottle: A Message to Ourselves in 2039. J Med Internet Res 2019 Nov 01;21(11):e16274 [FREE Full text] [doi: 10.2196/16274] [Medline: 31682578]
2. World Health Organization. Global diffusion of eHealth: Making universal health coverage achievable: Report of the third global survey on eHealth. WHO Global Observatory for eHealth.: World Health Organization; 2016 Dec. URL: https://www.who.int/goe/publications/global_diffusion/en/ [accessed 2020-01-10]
3. eHealth: Digital health and care. European Commission. 2019. URL: https://ec.europa.eu/health/ehealth/cooperation_en [accessed 2019-12-19]
4. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
5. Van Velthoven MH, Cordon C. Sustainable Adoption of Digital Health Innovations: Perspectives From a Stakeholder Workshop. J Med Internet Res 2019 Mar 25;21(3):e11922 [FREE Full text] [doi: 10.2196/11922] [Medline: 30907734]
6. Buis L. Implementation: The Next Giant Hurdle to Clinical Transformation With Digital Health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
7. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018 Sep 20;1(1):48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
10. WHO Guideline: recommendations on digital interventions for health system strengthening. World Health Organization. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2019-12-10]
11. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2(1):38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
12. Jha AK. Meaningful use of electronic health records: the road ahead. JAMA 2010 Oct 20;304(15):1709-1710. [doi: 10.1001/jama.2010.1497] [Medline: 20959581]
13. Campbell SM, Nyholm S. Anti-Meaning and Why It Matters. J. of the Am. Philos. Assoc 2015 Dec 29;1(4):694-711. [doi: 10.1017/apa.2015.9]
14. Michie S, West R, Spring B. Moving from theory to practice and back in social and health psychology. Health Psychol 2013 May;32(5):581-585. [doi: 10.1037/a0030205] [Medline: 23646841]
15. Deci EL, Ryan RM. Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology/Psychologie canadienne 2008;49(3):182-185. [doi: 10.1037/a0012801]
16. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
17. Kahn WA. Psychological Conditions of Personal Engagement and Disengagement at Work. AMJ 1990 Dec;33(4):694. [doi: 10.5465/256287]
18. Maslow AH. The farther reaches of human nature. New York: Viking Press; 1971.
19. Chalofsky N, Krishna V. Meaningfulness, Commitment, and Engagement:The Intersection of a Deeper Level of Intrinsic Motivation. Advances in Developing Human Resources 2009 Mar 24;11(2):189-203. [doi: 10.1177/1523422309333147]
20. Lupton D. Apps as Artefacts: Towards a Critical Perspective on Mobile Health and Medical Apps. Societies 2014 Oct 29;4(4):610. [doi: 10.3390/soc4040606]

XSL•FO
RenderX

21. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
22. Steele Gray C, Miller D, Kuluski K, Cott C. Tying eHealth Tools to Patient Needs: Exploring the Use of eHealth for Community-Dwelling Patients With Complex Chronic Disease and Disability. JMIR Res Protoc 2014 Nov 26;3(4):e67. [doi: 10.2196/resprot.3500]
23. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving Patient Experience and Primary Care Quality for Patients With Complex Chronic Disease Using the Electronic Patient-Reported Outcomes Tool: Adopting Qualitative Methods Into a User-Centered Design Approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
24. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The Electronic Patient Reported Outcome Tool: Testing Usability and Feasibility of a Mobile App and Portal to Support Care for Patients With Complex Chronic Disease and Disability in Primary Care Settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
25. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: A noncategorical qualitative alternative for developing nursing knowledge. Res. Nurs. Health 1997 Apr;20(2):169-177. [doi: 10.1002/(sici)1098-240x(199704)20:2<169::aid-nur9>3.0.co;2-i]
26. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015 Apr 23;13(1):91. [doi: 10.1186/s12916-015-0279-6]
27. Steele Gray C, Gravesande J, Hans PK, Nie JX, Sharpe S, Loganathan M, et al. Using Exploratory Trials to Identify Relevant Contexts and Mechanisms in Complex Electronic Health Interventions: Evaluating the Electronic Patient-Reported Outcome Tool. JMIR Form Res 2019 Feb 27;3(1):e11950 [FREE Full text] [doi: 10.2196/11950] [Medline: 30810532]
28. Gomersall T, Nygård L, Mihailidis A, Sixsmith A, Hwang AS, Hedman A, et al. Network-based approaches for evaluating ambient assisted living (AAL) technologies. Evaluation 2017 Apr 18;23(2):192-208. [doi: 10.1177/1356389017697615]
29. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
30. Steele Gray C, Wodchis WP, Upshur R, Cott C, McKinstry B, Mercer S, Project Collaborators And Technology Partner, QoC Health Inc. Supporting Goal-Oriented Primary Health Care for Seniors with Complex Care Needs Using Mobile Technology: Evaluation and Implementation of the Health System Performance Research Network, Bridgepoint Electronic Patient Reported Outcome Tool. JMIR Res Protoc 2016 Jun 24;5(2):e126 [FREE Full text] [doi: 10.2196/resprot.5756] [Medline: 27341765]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
ePRO: electronic patient-reported outcome
NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability
VPD: Value-Proposition Design

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by P Anderberg, T Muto, K Goniewicz; comments to author 10.03.20; revised version received 05.05.20; accepted 03.06.20; published 29.07.20.

Please cite as:
Steele Gray C
Seeking Meaningful Innovation: Lessons Learned Developing, Evaluating, and Implementing the Electronic Patient-Reported Outcome Tool
J Med Internet Res 2020;22(7):e17987
URL: http://www.jmir.org/2020/7/e17987/
doi:10.2196/17987
PMID:32723725

©Carolyn Steele Gray. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predictors of Self-Determined Module Choice in a Web-Based Computer-Tailored Diet and Physical Activity Intervention: Secondary Analysis of Data From a Randomized Controlled Trial

Juul M J Coumans[1], MSc; Catherine A W Bolman[1], PhD; Anke Oenema[2], PhD; Lilian Lechner[1], PhD

[1]Department of Health Psychology, Faculty of Psychology, Open University of the Netherlands, Heerlen, Netherlands

[2]Department of Health Promotion, Caphri, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Juul M J Coumans, MSc
Department of Health Psychology
Faculty of Psychology
Open University of the Netherlands
Valkenburgerweg 177
Heerlen
Netherlands
Phone: 31 45 576 2635
Email: juul.coumans@ou.nl

Abstract

Background: Tailoring an online intervention to participant preferences (eg, by giving participants a choice which modules to follow) may increase engagement in the intervention, motivation for behavioral change, and possibly intervention effects. So far, little is known about what characteristics predict these module choices. Filling this knowledge gap is useful for optimizing program engagement.

Objective: We investigated participant choice for a dietary and/or physical activity (PA) promotion module in our web-based computer-tailored intervention based on self-determination theory (SDT) and motivational interviewing (MI). Furthermore, we investigated which demographic characteristics, current behavior, psychosocial constructs and constructs from SDT and MI, and program-related variables such as advice on which module to follow were associated with these choices.

Methods: Observational data were used from the randomized controlled trial MyLifestyleCoach of participants who were randomized into the intervention condition, completed the baseline questionnaire, and made a module choice in the opening session of the intervention. Here, they received advice on their own dietary and PA behavior. At the session's end, they chose which lifestyle modules they would like to follow (both, diet, PA, or no module). Measurements included demographic information; self-reported diet and PA; and several psychosocial, SDT, and MI constructs. In total, data from 619 Dutch adults (59.6% women; mean age was 51.9 [SD 13.5] years) were analyzed. A stepwise multinomial logistic regression analysis was conducted to investigate which characteristics are related to module choice; the diet module served as reference category as almost everyone was advised to follow this module.

Results: Of this sample, 54.8% (339/619) chose to do both the diet and PA module, 25.4% (157/619) chose to follow the diet module, 17.8% (110/619) preferred to follow no module, and 2.1% (13/619) chose to do the PA module only. Furthermore, it was found that older people, those who consumed more fruit, and those who scored lower on importance to change their current diet were more likely to choose no module compared to the diet module. People who had more motivation to change their current PA and those who received strong advice compared with slight advice to follow the diet module were more likely to choose both modules compared with the diet module only.

Conclusions: The results show that more than half of the sample was interested in following both the diet and PA module in this online lifestyle intervention. Several characteristics were found to be related to module choice. A future challenge is to examine how this knowledge can be used to improve future interventions, such as tailoring (messages or content) on specific groups or examining where and how MI could be used to motivate people to make a certain module choice.

Trial Registration: Netherlands Trial Register NL7333; https://www.trialregister.nl/trial/7333

(J Med Internet Res 2020;22(7):e15024) doi:10.2196/15024

XSL•FO
RenderX

KEYWORDS

eHealth; computer-tailoring; self-determination theory; multiple health behaviors; lifestyle promotion; module choice

Introduction

Health risk behaviors such as unhealthy dietary intake and insufficient physical activity (PA) are widespread among adults, making the prevention of these behaviors a public health priority [1-3]. The small effect sizes, limited sustainability of effects, and high dropout rates usually found with existing online diet and PA interventions suggest that there is room for improvement in this field [4,5]. In this regard, there is more and more evidence showing the importance of autonomous motivation in sustained behavioral change [6,7], and this might therefore be a potential strategy for improving diet and PA interventions. Self-determination theory (SDT) postulates that providing conditions that support autonomy and choice are important basic psychological needs to improve more autonomous forms of motivation [8-11]. This autonomy support may be relevant during participation in intervention programs but also necessary at the start of an intervention, allowing participants the option to choose for themselves what parts of the intervention program they want to participate in. For interventions targeting multiple behaviors, this means people should have the option to choose which behavior they want to work on, preferably from separate modules so they can focus on a behavior in which they are really interested. Until now, research about the choices people make in such interventions (ie, which modules people choose) has been limited. But research that focuses on which factors and determinants are related to these choices is even more scarce.

To date, research has mainly investigated predictors of participation, use, and revisits of eHealth interventions instead of why people choose certain topics in these interventions [12-19]. In general, these studies have shown that women, older people, people with more education, people with a higher income, and people with a healthier BMI and a healthier lifestyle are more likely to visit, start, use, and revisit web-based interventions. In addition, insufficient time, no personalized content, dissatisfaction with the content, and computer difficulties were linked to a decrease in adherence [20]. Giving participants a choice of which modules to follow may improve user engagement and favorably affect compliance and long-term success [21]. One study found that a majority of patients showed improvements in anxiety and depressive symptoms when they could decide what modules to use in a tailored web-based treatment, and only one participant (of out 27) dropped out [22].

Yet only a few studies have described which behaviors (modules) individuals choose to start with in multiple health behaviors interventions [21]. Two studies in the general adult population found that people prefer to select the PA module over dietary modules [19,23]. In a study with cancer survivors, a reversed pattern was present: the diet module was preferred over the PA module [24]. Another study found no favorable effects of messages tailored to a participant-selected topic over an expert-determined topic on PA levels [25]. Based on these mixed findings, the role of module choice should be further explored in tailored web-based interventions.

So far, two studies have identified factors that relate to module choice in online interventions [23,24]. These studies have demonstrated that age, marital status, advice (both content and number of referrals), and healthier lifestyle have been related to module choice. Research to date has not yet determined whether psychosocial factors also relate to module choice. According to the SDT, individuals experience more autonomous forms of motivation when the needs for relatedness, autonomy, and competence are met [10,26]. Thus, feeling a higher degree of autonomy and perceived competence in changing or maintaining improved dietary intake or PA may lead to more intrinsic motivation to change behavior, which could subsequently lead to engagement in actions to achieve the intended behavior change [8-11]. The client-centered counseling style and techniques from motivational interviewing (MI) can create a facilitating environment in which these needs are promoted [27]. However, (intrinsic) motivation alone is often not sufficient for initiating and maintaining improved nutrition or PA levels. Other psychosocial characteristics closely linked to motivation may also play a role in this process and possibly in module choice. It may be that people who have a higher commitment to change certain behaviors are more likely to choose a certain module. However, to our knowledge, no study has examined to what extent these factors relate to module choice [28].

With eHealth interventions becoming more and more popular in recent years, there is now a general idea concerning elements that make an intervention effective and increase its use. However, a detailed understanding of what actually occurs within an intervention is still lacking. Our main aim is to elucidate factors related to module choice. Identifying and understanding which characteristics relate to the choice of a specific module or the choice not to follow any module can inform intervention improvement. This knowledge can be used to optimally tailor the intervention to the specific characteristics and needs of the participants and prevent nonuse. This may result in increasing intervention exposure and relevance that could result in increased effectiveness of the intervention.

In this study, participant choice of a dietary and/or PA promotion module is investigated in our web-based computer-tailored intervention based on SDT/MI. The main aim of this study is to examine which factors are related to these module choices (diet, PA, both, or no module) in the intervention. For these factors, we will look at demographic characteristics, psychosocial constructs related to SDT and MI, current behavior, and program-related variables (advice on which module to follow and the importance of changing behavior after feedback). Several hypotheses have been formulated based on previously mentioned literature. Regarding demographic characteristics, it is expected that women, older people, and people with more education are more inclined to choose a module (both, diet, PA) compared with no module. Furthermore, it is expected that a higher intrinsic motivation to perform a certain healthy behavior is related to choosing that particular module (ie, a higher motivation to engage in sufficient PA is related to choosing the

XSL•FO
RenderX

PA or PA and diet module compared with the diet or no module). In addition, a healthier lifestyle is likely to affect module choice as well (eg, people with a healthier diet should be more inclined to choose the diet module compared with no module). Last, advice to follow a particular module and importance of the targeted change behavior is also expected to relate to choosing that particular module (both, diet, PA) compared with no module.

Methods

Design

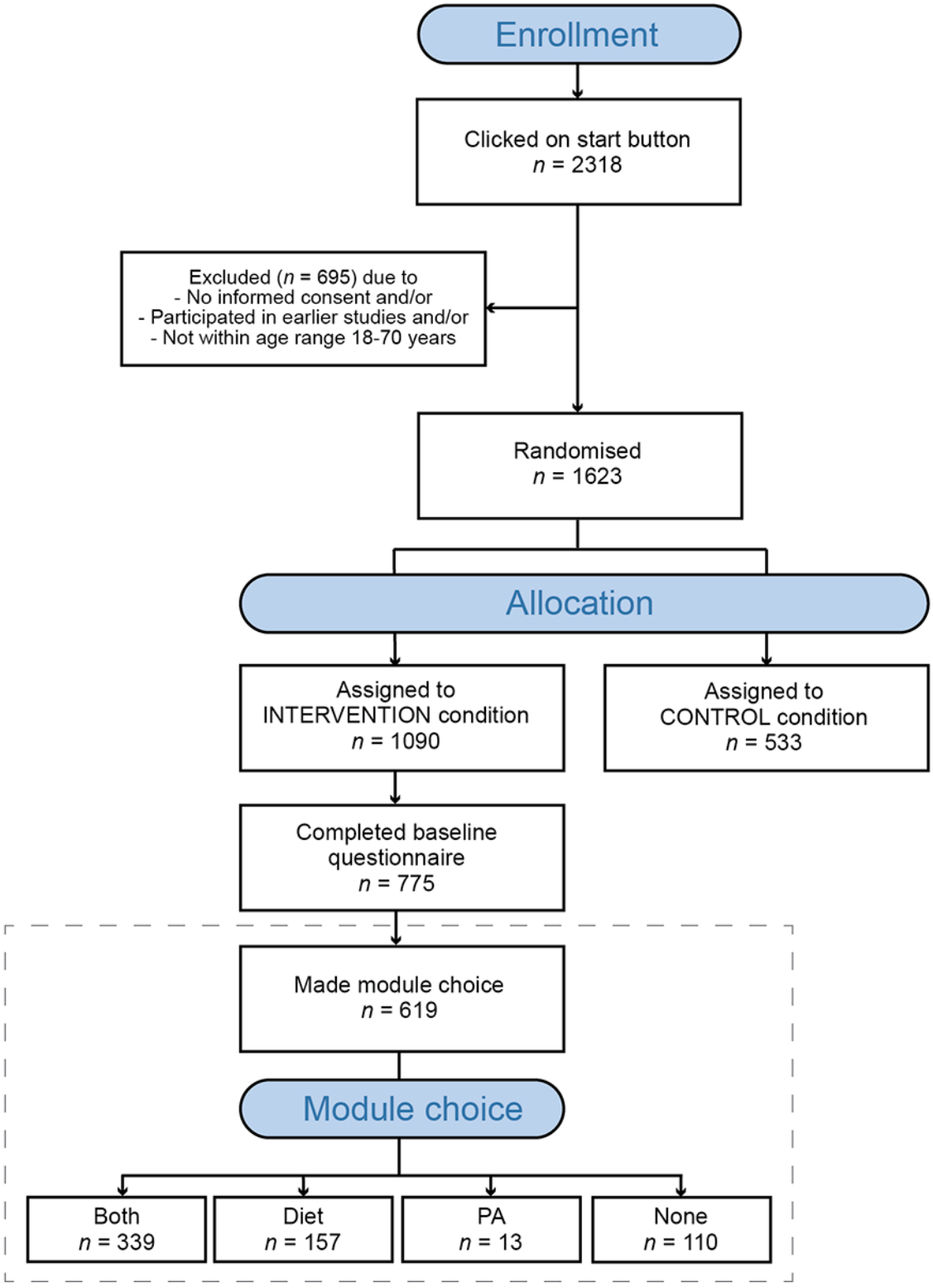
This observational study was conducted as part of a 2-armed randomized controlled trial, MyLifestyleCoach. Data were collected in the period from mid-October 2018 until mid-May 2019. A comprehensive description of the intervention and the underlying theoretical frameworks, details of the trial design, and sample size calculations have been published elsewhere [29,30]. This web-based computer-tailored dietary and PA promotion intervention was developed using the intervention mapping protocol, and it is based on principles from SDT and MI. MyLifestyleCoach consists of two modules: I Move, an existing effective computer-tailored intervention to promote PA (PA module), and I Eat, which is aimed at promoting healthy eating (diet module), using similar behavior strategies as the I Move module [31]. The I Eat module is new, however, and its effectiveness has not been evaluated yet [29]. This study has been reviewed and approved by the Committee for Ethics and Consent in Research of the Open University of the Netherlands (reference number: U2018/07266/SVW) and was registered with the Netherlands Trial Register [NL7333].

Participants

Participants for this study were recruited via an internet research panel of Dutch inhabitants who occasionally volunteer in web-based research. They were eligible to participate if they were aged 18 to 70 years. A total of 1090 participants were randomized into the intervention condition of this study, of which 71.10% (775/1090) of participants completed the baseline questionnaire and 79.9% (619/775) of participants made a module choice in the opening session of MyLifestyleCoach. See Figure 1 for an overview of the study design and flow of participants. The dashed line represents the focus of this study.

XSL•FO
RenderX

Figure 1. Overview of the study design.



Measurements

All the following constructs were measured in the online self-reported baseline questionnaire except for perceived importance; this was measured at the end part of the opening session, after feedback regarding current behavior had been provided. Researchers had no influence on how participants reported the outcomes. Participants had to complete each question to be able to proceed to the next section of the online questionnaire.

Demographics

The baseline questionnaire assessed gender, age, educational level, marital status (with or without a partner), health status (on a scale from 0 to 100), and height and body weight to calculate BMI. Educational level (ie, the highest level of education completed) was recoded into three categories: low

XSL•FO
RenderX

(primary or basic vocational school), medium (secondary vocational school or high school), and high (higher vocational school or university). Work was recoded into two categories: employed (full-time or part-time work) and unemployed (volunteering, having no job, incapacitated to work, retired, household chores, or studying). Furthermore, we asked whether participants had an impairment that would prevent them from being physically active (yes/no).

Psychosocial Constructs Including Self-Determination Theory/Motivational Interviewing–Related Constructs

Competence

Two Perceived Competence Scales were used to assess perceived competence (basic psychological need, SDT) for a healthy diet and PA [32]. This is a 4-item questionnaire on which participants indicated the extent to which each statement is true for them on a 7-point Likert scale (1 = not at all, 7 = very true). An example item is: "I feel confident in my ability to maintain a healthy diet/exercise regularly." The perceived competence scores were determined by calculating a mean score for the 4 items, separately for diet and PA. The internal consistency of these scales in this study was excellent (Cronbach α=.95 for diet and .96 for PA).

Self-Regulation and Motivation

The Treatment Self-Regulation Questionnaire concerns why people would engage in health-relevant behavior [33,34]. It was used to assess the degree to which a person's motivation for a healthy diet and PA is relatively driven by autonomous or self-determined reasons (SDT). Participants indicated the extent to which each of the 15 items was true for them on a 7-point Likert scale (1 = not at all, 7 = very true) for healthy eating and PA. Six items measure an autonomous regulatory style (Cronbach α=.92 for diet and .93 for PA), 6 items measure a controlled regulatory style (Cronbach α=.86 for diet and .89 for PA), and 3 items assess amotivation (Cronbach α=.70 for diet and .76 for PA). Responses to the respective items for each regulatory style were averaged to obtain a score for each of the three self-regulatory styles.

Intrinsic Motivation

According to the SDT, autonomous motivation consists of identified, integrated, and intrinsic regulation. However, the Treatment Self-Regulation Questionnaire does not differentiate between these autonomous forms of motivation. Therefore, the subscale intrinsic regulation from the Dutch Behavioral Regulation in Exercise Questionnaire was administered to determine intrinsic motivation, as this is the only fully self-determined form of motivation [35]. Participants rated on a 5-point Likert scale to what extent each of the 4 items was true for them with values of 1 = not true for me to 5 = very true to me. An example item was: "I exercise because it's fun" (translated to engage in PA in Dutch). To compare intrinsic motivation for exercise with a healthy diet as closely as possible, we replaced exercise in all items by eating healthily. The scores of the 4 items were averaged into one intrinsic motivation score, both for a healthy diet and PA. The internal consistency of these scales in this study was excellent (Cronbach α=.93 for diet and .98 for PA).

Importance

In the opening session, participants were asked to rate their perceived importance of eating (more) healthily and becoming more physically active using an importance ruler after they received feedback on their current behavior. This is a 1 to 10 scale, a tool to support the use of MI [36]. Including this factor next to the baseline importance scale was necessary because importance could have been changed as a result of the content (eg, feedback on current behavior) of the opening session.

Intention

Three items were used to measure intention for eating healthier and being physically active [37]. Participants rated on a 10-point scale to what extent, how strong, and how probably they intended to eat (more) healthily and become physically more active (or stay sufficiently active). A mean score was calculated for diet and PA based on these 3 items. The internal consistency of these scales in this study was excellent (Cronbach α=.91 for diet and .95 for PA).

Commitment

Three items were used to measure commitment to eating healthier and being physically active [38]. Participants rated on a 5-point Likert scale how important eating (more) healthily/PA is for them, how involved they were in eating (more) healthily/PA, and how committed they are toward eating (more) healthily/PA. A mean score was calculated for diet and PA based on these items. The internal consistency of these scales in this study was good (Cronbach α=.89 for diet and .92 for PA).

Diet

Diet was assessed with a validated Food Frequency Questionnaire [39,40]. It assesses the frequency and quantity of a variety of food items eaten in a typical week in the last month. Participants reported on how many days they typically consumed fruit, vegetables, and fish (ranging from 0 to 7 days per week). We added questions assessing the size of vegetables and fruit portions based on Willems et al [41]. The intake of pieces of fruit per day was calculated by multiplying the frequency by the number of pieces with the reported number of consumption days, divided by 7 (days a week). When participants reported eating fruit or vegetables on at least one day but did not fill in the portion size, the portion size was replaced by the median. Furthermore, participants reported the consumption frequency of a particular snack food in a typical week in the last month on a 7-point Likert scale: 1 = never/less than once a week, 2 = 1 to 3 times a week, 3 = 4 to 6 times a week), 4 = 1 time per day, 5 = 2 times per day, 6 = 3 times per day, or 7 = 4 or more times per day. Eight types of snacks were assessed, namely unsalted nuts, dried fruits, chocolate, candy, cookies, chips, ice cream, and savory pastries. The consumption frequency of unhealthy snacks per day was determined by summing the recoded frequencies for last 6 snacks (chocolate to savory pastries) and dividing by the numbers of days in a week.

Physical Activity Level

PA behavior was assessed using the validated self-administered Dutch Short Questionnaire to Assess Health [42]. This

XSL•FO
RenderX

questionnaire has been adapted to an online format. Participants were asked on how many days of in a typical week (0-7) during the past month they engaged in (1) walking to work/school, (2) cycling to work/school, (3) work, (4) household activities, (5) walking, (6) cycling, (7) gardening, and (8) odd jobs. If participants engaged at least once a week in the before mentioned activities, they reported how many hours and how many minutes they engaged in this activity. Work and household activities were separated for light/moderate and vigorous activity (examples were provided for light/moderate and vigorous activities). For sports, participants were asked which sports they engaged in the last month (they could choose up to 4 sports). Then they provided the number of days per week and duration for these sports. For walking, cycling, gardening, and sports, participants rated whether they considered it to be light, moderate, or vigorous. PA behavior was operationalized as the total number of minutes of moderate to vigorous physical activity per week (MVPA). This was calculated by multiplying the frequency (how many days per week), and duration (how many hours and minutes per day) of leisure and transport walking, leisure and transport cycling, gardening, household activities, odd jobs, and sports performed with moderate or vigorous intensity. Values were inspected using frequency tables, and extreme values (eg, 16 hours of heavy household activities) were replaced by the median of the sample. This self-report was used as this was the most feasible method (ie, convenience, low costs, and proven as a reliable and valid tool) in assessing PA compared with objective observations [43].

Procedure

Members of the research panel received an email advertising the study about lifestyle containing information regarding the study (background, providing personal advice regarding dietary and PA behavior, that the program could help to improve their dietary and PA behavior, and information about prizes they could win for fully completing the intervention) and a link to the study website. Participation was free. When they clicked on the link, they could read further extensive information regarding the project, target population, expectations, data protection regulations, benefits (such as insight in current dietary and PA behavior and having a chance to win prizes), and contact details. If they were interested, they could click on the start button.

All eligible participants who agreed to participate answered some questions relevant to the inclusion and exclusion criteria and signed an online consent form. They were then computer randomized into the intervention condition or waiting list control condition (2:1) using a computer-determined sequence and completed the baseline questionnaire. In this paper, only respondents of the intervention condition were included in the analysis. After completing the baseline questionnaire, participants in the intervention condition continued directly to the opening session. Here, they were introduced to the structure of the program by text and by video coach. The video coach did not go into detail about the content of the modules or give advice on which module to follow. Participants were also asked to rate their perceived importance of eating (more) healthily and becoming more physically active on a 10-point scale. They received advice on which module to follow using a traffic light system based on the responses to the baseline questionnaire on where there was room for improvement, separately for diet and PA (see more information about this in the advice section). Participants were given a feedback message consisting of a summary of the extent to which they adhered to the diet and PA guidelines, in which domains there was room for improvement, and how important a healthy diet and sufficient PA was to them. They also received a brief reminder about the structure of the program.

Subsequently, participants were invited to choose which module they wanted to follow: diet, PA, both, or no module. The ones who chose no module received an email 2 weeks later with a link to the program where they could again make a module choice. Within the program, participants who had chosen to follow either the diet or the PA module were given the option the follow the other module once they had completed a session. Participants who did not immediately enter the opening session after completing the baseline questionnaire got a friendly reminder to do so 1 week after they completed the baseline questionnaire.

Advice

Participants received traffic light–based advice presented on two separate pages, one each for diet and PA. The color of the traffic light displayed in the advice indicated how closely their current PA level and diet corresponded to health recommendations. The corresponding text advised which modules were most relevant for them to use. For diet, the targeted behaviors (based on a pilot study) were to consume at least 200 grams of fruit (2 portions) and 250 grams of vegetables daily, eat fish at least once a week, and consume no unhealthy (ie, energy-dense) snacks [44]. For PA, the advice recommended accumulating at least 2.5 hours of MVPA every week [45]. A green light was shown when participants met the guideline. We then praised the positive scores and advised that it was not of high priority to follow the corresponding module but participants were still free to have a look at the module. An orange light was shown when the participants were close to meeting the guidelines (ie, the recommendations for at least one of the targeted behaviors for diet but not all were unmet or participant had 120 to 150 minutes of MVPA per week). Here, we also praised the reasonable positive score but advised following the module because there was room for improvement. Last, a red light was displayed when the participant failed to meet the guideline (ie, less than 120 minutes of MVPA per week). We strongly advised the participant to follow the corresponding module because the module could help them improve their behavior and health. If participants did not meet a certain guideline, they got detailed results about their current activity toward that particular behavior (fruit, vegetables, fish, unhealthy snacks, or PA) in relation to the guidelines. See Figure 2 for an example of the traffic light–based advice. The traffic lights were meant to provide participants with insight into their behavior and what they could change; they were not necessarily intended to induce compliance with dietary recommendations and PA guidelines.

XSL•FO
RenderX

Figure 2. Screenshot of the page where people get advice in the opening session.



Statistical Analyses

Descriptive statistics were used to depict demographic and psychosocial characteristics of the participants included in this study. To investigate the predictors of module choice, we performed Spearman correlations between all predictors to determine the intercorrelation of predictors. In case of high intercorrelation, predictors could be removed beforehand. A stepwise multinomial logistic regression analysis was conducted to identify which demographic (step 1), psychosocial including SDT and MI constructs (step 2), dietary behavior (step 3), and program-related variables such as advice and changed importance (step 4) were related to module choice. Significance was set at $P<.05$. Multicollinearity was assessed using variance inflation factors (VIF). Furthermore, the fits of the models were compared with likelihood ratio tests. If $P<.05$, the more complex model was significantly better than the simpler model, and thus the more complex model was favored. All statistical analyses were performed with the statistical software R version 3.4.4 (R Foundation for Statistical Computing).

Results

Participants

The study population was slightly overrepresented by women (369/619, 59.6%; men: 250/619, 40.4%). The mean age was 51.9 years, and about two-thirds of the participants were in a relationship. Furthermore, the majority of the participants had a high level of education. Of this sample, more than half of the participants had a paid job (full-time or part-time). Of those who did not have a paid job, 8.9% (20/225) were volunteering, 7.1% (16/225) were unemployed, 16.4% (37/225) were incapacitated to work, 44.0% (99/225) were retired, 16.4% (37/225) were studying, and 7.1% (16/225) did household chores as their main activity. In total, 4.8% (30/619) of participants had a physical impairment that prevented them from being physically active. The mean BMI was 26.6 kg/m^2; 1.8% (11/619) were classified as underweight, 42.2% (261/619) had a healthy weight, 35.2% (218/619) were classified as having nonobese overweight, and 20.8% (129/619) were classified as obese. About one in three participants complied with the fruit guideline of 2 portions of fruit per day, 11.8% (73/619) adhered to the vegetable guideline of 250 grams of vegetables per day, 62.5% (387/619) of this sample consumed fish at least once a week, and 7.8% (48/619) never consumed unhealthy snacks. More than 90% of this sample engaged in 150 minutes or more MVPA per week (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the study population.

Characteristics	Module choice				
	Both (n=339)	Diet (n=157)	PA[a] (n=13)	None (n=110)	Total (n=619)
Demographics					
>Age in years, mean (SD)	52.1 (12.8)	49.3 (15.1)	52.8 (14.3)	54.7 (12.6)	51.9 (13.5)
Gender, n (%)					
Women	215 (63.4)	97 (61.8)	7 (53.8)	50 (45.5)	369 (59.6)
Men	124 (36.6)	60 (38.2)	6 (46.2)	60 (54.5)	250 (40.4)
Education, n (%)					
Low	13 (3.8)	6 (3.8)	1 (7.7)	4 (3.6)	24 (3.9)
Medium	94 (27.7)	38 (24.2)	4 (30.8)	23 (20.9)	159 (25.7)
High	232 (68.4)	113 (72.0)	8 (61.5)	83 (75.5)	436 (70.4)
Marital status, n (%)					
Partner	226 (66.7)	102 (65.0)	10 (76.9)	77 (70.0)	415 (67.0)
Single	113 (33.3)	55 (35.0)	3 (23.1)	33 (30.0)	204 (33.0)
Employment status, n (%)					
Employed	210 (61.9)	104 (66.2)	7 (53.8)	73 (66.4)	394 (63.7)
Unemployed	129 (38.1)	53 (33.8)	6 (46.2)	37 (33.6)	225 (36.3)
Physical impairment, n (%)					
No	322 (95.0)	147 (93.6)	13 (100)	107 (97.3)	589 (95.2)
Yes	17 (5.0)	10 (6.4)	0 (0.0)	3 (2.7)	30 (4.8)
BMI group, n (%)					
Underweight	4 (1.2)	4 (2.5)	1 (7.7)	2 (1.8)	11 (1.8)
Normal	137 (40.4)	63 (40.1)	4 (30.8)	57 (51.8)	261 (42.2)
Overweight	112 (33.0)	62 (39.5)	5 (38.5)	39 (35.5)	218 (35.2)
Obese	86 (25.4)	28 (17.8)	3 (23.1)	12 (10.9)	129 (20.8)
BMI (kg/m^2), mean (SD)	27.1 (5.7)	26.4 (5.0)	26.1 (4.7)	25.4 (3.6)	26.6 (5.2)
Health status (1-100), mean (SD)	68.3 (16.2)	69.2 (16.9)	73.7 (9.0)	74.1 (13.7)	69.7 (16.0)
Psychosocial and SDT[b]/MI[c], mean (SD)					
Competence (1-7)					
Diet	5.0 (1.5)	4.9 (1.3)	4.7 (1.4)	5.3 (1.4)	5.1 (1.4)
PA	5.2 (1.5)	5.3 (1.3)	4.8 (1.7)	5.3 (1.5)	5.2 (1.4)
Amotivation (1-7)					
Diet	2.3 (1.2)	2.4 (1.2)	2.4 (1.3)	2.4 (1.2)	2.3 (1.2)
PA	2.0 (1.2)	2.3 (1.3)	2.3 (1.5)	2.2 (1.2)	2.2 (1.2)
Controlled regulatory style (1-7)					
Diet	2.8 (1.3)	2.8 (1.1)	2.7 (1.3)	2.7 (1.1)	2.8 (1.2)
PA	2.7 (1.3)	2.8 (1.2)	2.7 (1.2)	2.7 (1.1)	2.7 (1.2)
Autonomous regulatory style (1-7)					
Diet	5.5 (1.1)	5.4 (1.2)	5.1 (1.1)	5.4 (1.1)	5.5 (1.1)
PA	5.7 (1.1)	5.5 (1.2)	5.5 (1.3)	5.6 (1.1)	5.6 (1.1)
Intrinsic motivation (1-5)					
Diet	3.4 (1.0)	3.4 (1.0)	3.1 (1.0)	3.5 (1.0)	3.4 (1.0)

XSL•FO
RenderX

Characteristics	Module choice				
	Both (n=339)	Diet (n=157)	PA[a] (n=13)	None (n=110)	Total (n=619)
PA	3.8 (1.1)	3.9 (1.1)	3.6 (1.4)	3.9 (1.1)	3.8 (1.1)
Intention (1-10)					
Diet	7.9 (1.5)	7.7 (1.4)	7.5 (2.1)	7.9 (1.6)	7.8 (1.5)
PA	7.8 (1.7)	7.9 (1.8)	7.0 (1.4)	7.7 (1.8)	7.8 (1.7)
Commitment (1-5)					
Diet	3.9 (0.6)	3.8 (0.7)	3.8 (0.7)	3.9 (0.6)	3.9 (0.7)
PA	3.8 (0.8)	4.0 (0.8)	3.8 (0.8)	3.8 (0.8)	3.9 (0.8)
Compliance to diet and PA guidelines, n (%)					
Diet					
Fruit	106 (31.3)	50 (31.8)	3 (23.1)	52 (47.3)	211 (34.1)
Vegetables[d]	36 (10.6)	16 (10.2)	1 (7.7)	12 (10.9)	65 (10.5)
Fish	206 (60.8)	98 (62.4)	9 (69.2)	74 (67.3)	387 (62.5)
Snacks	29 (8.6)	7 (4.5)	3 (23.1)	9 (8.2)	48 (7.8)
Diet	3 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.5)
PA					
MVPA[e,f]	314 (93.7)	152 (97.4)	9 (69.2)	104 (96.3)	579 (94.6)
Program-related variables					
Module advised, n (%)					
Both	21 (6.2)	4 (2.5)	5 (38.5)	4 (3.6)	34 (5.5)
Diet	315 (92.9)	153 (97.5)	8 (61.5)	106 (96.4)	582 (94.0)
PA	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
None	3 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.5)
Importance (1-10), mean (SD)					
Diet	8.2 (1.4)	8.0 (1.3)	7.8 (1.5)	7.9 (1.4)	8.1 (1.4)
PA	8.5 (1.3)	8.4 (1.3)	8.5 (1.2)	8.4 (1.4)	8.5 (1.3)

[a]PA: physical activity.

[b]SDT: self-determination theory.

[c]MI: motivational interviewing.

[d]Vegetables: corrected values were used. Portion sizes of 75 grams or larger were divided by 50, as one portion equals 50 grams.

[e]MVPA: moderate to vigorous physical activity.

[f]Corrected values were used according to the Short Questionnaire to Assess Health manual; extreme values were furthermore replaced by the median. As a consequence, 7 participants were excluded as they reported to be physically active over 6720 minutes per week.

Module Advice and Module Choice

Table 2 displays the provided module advice and subsequent module choice. In total, 24.6% (152/619) of participants were strongly advised to follow the diet module (red traffic light), 75.0% (459/619) of participants were slightly advised to follow the diet module (orange traffic light), and only 0.5% (3/619) of participants were not advised to follow the diet module (green traffic light). Regarding the PA advice, 3.6% (22/619) of participants were strongly advised to follow the PA module (red traffic light), 1.9% (12/619) of participants were slightly advised to follow the PA module (orange traffic light), and 94.5% (585/619) of participants were not advised to follow the PA module (green traffic light). Of the participants who made a module choice in the opening session of the intervention, 54.8% (339/619) of participants chose to follow both modules, 25.4% (157/619) of participants only chose the diet module, 2.1% (13/619) of participants chose only the PA module, and 17.8% (110/619) of participants chose no module. Thus, more than 80% of the participants were interested in following at least one module, and more than half of the participants were interested in following both modules.

XSL•FO
RenderX

Table 2. Overview of advice and module choice.

Traffic light[a]		Module choice				Total
Diet	PA[b]	Both, n (%)	Diet, n (%)	PA, n (%)	None, n (%)	N (%)
Red	Red	5 (1.5)	1 (0.6)	0 (0.0)	1 (0.9)	7 (1.1)
Red	Orange	2 (0.6)	0 (0.0)	1 (7.7)	0 (0.0)	3 (0.5)
Red	Green	86 (25.4)	39 (24.8)	1 (7.7)	16 (14.5)	142 (22.9)
Orange	Red	9 (2.7)	1 (0.6)	3 (23.1)	2 (1.8)	15 (2.4)
Orange	Orange	5 (1.5)	2 (1.3)	1 (7.7)	1 (0.9)	9 (1.5)
Orange	Green	229 (67.6)	114 (72.6)	7 (53.8)	90 (81.8)	440 (71.1)
Green	Red	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Green	Orange	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Green	Green	3 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.5)
Total	—[c]	339 (54.8)	157 (25.4)	13 (2.1)	110 (17.8)	619 (100.0)

[a]Traffic light advice was based on self-reported (uncorrected) values. A red traffic light indicates that the module was strongly advised, an orange traffic light indicates that the module was slightly advised, and a green traffic light indicates that the module was not advised but that participants could have a look at the module.

[b]PA: physical activity.

[c]Not applicable

Which Factors Are Related to Module Choice?

First, it was investigated using Spearman correlations whether there was collinearity between the continuous predictors. There were several predictors that were highly intercorrelated ($r \geq .60$) with the other predictors. See Multimedia Appendix 1 for the correlation table. Based on these results, it was decided to remove the psychosocial constructs of intention, as the correlation coefficients were approaching a correlation coefficient of .80 with perceived competence (intention diet with perceived competence of diet, $rs=.78$, $P<.001$, and intention PA with perceived competence of PA, $rs=.76$, $P<.001$) and because perceived competence was of more theoretical relevance in this study (basic psychological need in the SDT).

Since only 13 participants chose the PA module only, the frequency was too low to enable group comparisons and coefficients could not be reliably estimated. We, therefore, removed those cases from the analyses. Then the stepwise multinomial logistic regression analysis was conducted with module choice as the outcome (both modules, diet, or no module). The diet module served as the reference category, because everyone was advised to follow the diet module, except for 3 participants who were excluded here. Multimedia Appendix 2 presents the results of this analysis. The means each predictor presented in Table 1, stratified for module choice, may aid understanding.

Model 1 shows that people with a higher age were more likely to choose both modules as well as no module compared with the diet module; however, the effect that older adults were more likely to choose both modules became nonsignificant. When the psychosocial and SDT/MI predictors were added to the previous (model 2), it was found that people who had higher levels of amotivation to change their current PA levels were less likely to choose both modules compared with the diet module. Model 3 additionally shows that people who consumed more fruit per day were more likely to choose no module compared with the diet module. When the program-related variables (model 4) were added to model 3, it could be seen that people who were strongly advised to follow the diet module were more likely to choose both modules than those who were slightly advised compared with the diet module. Last, people who found it more important to change their current diet were less likely to choose no module compared with the diet module.

Model Comparisons

Adding the predictors to each subsequent model led to an increased Nagelkerke R^2. Models 2 and 3 did not have a significantly improved fit compared with model 1 (see Multimedia Appendix 3). However, model 4 had a significantly better fit than models 2 and 3, whereas compared with model 1 there was found a marginal statistically significant effect. Therefore, model 4 seems to explain our data best, as this model had a significantly better fit than the other models (except for model 1) and also was indicated by the highest Nagelkerke explained variance (see Multimedia Appendix 2).

Additional Information Regarding the Regression Analysis

A VIF larger than 10 is an indication of multicollinearity (see Multimedia Appendix 4). We removed autonomous motivation and commitment for diet and PA from the analysis. Importance was nevertheless included in the analysis, as this turned out to be an important factor in relation to module choice. The results of the full model, including those predictors that were excluded because of a too large VIF, can be seen in Multimedia Appendix 5. Although minor differences in the results were found, our main findings seem to hold.

XSL•FO
RenderX

Discussion

Principal Findings

This study investigated participant choice for a dietary and/or PA promotion module in a web-based computer-tailored intervention based on SDT/MI. The main aim was to examine which factors were associated with these module choices. The results showed that more than half of the sample (55%) chose to follow both the diet and the PA module. About a quarter of the sample chose to follow to diet module only, 18% of the participants did not want to follow any module, and 2% chose the PA module only.

Our findings are in line with a previous study in which more participants used the diet module (58%) than the PA module (22%) [24]. But our findings are not consistent with several other studies in which the PA module was preferred over dietary modules that focused on limiting fat intake and increasing fruit and vegetable consumption [19,23,25]. Several explanations could be offered for this discrepancy. First, improvements in diet seemed to be far more necessary in our sample than improvements in PA levels. In our study, almost no one completely adhered to the dietary guidelines, whereas most participants already adhered to the MVPA guideline, which may be an overestimation of findings in the general Dutch population [46]. Second, people were completely free to decide for themselves which behavior they wanted to work on (ie, improving or maintaining current behavior) rather than simply trying to achieve or maintain the recommended levels as in line with Kanera and colleagues [41]. This may also be evident for a large proportion of the participants who chose both modules even though they were not advised to do so, indicating that for them it is important to maintain a healthy lifestyle. In other studies, there were restrictions in making a choice, such as only choosing modules when participants received orange or red traffic light advice [23]. Third, in our study, as well as in the study by Kanera and colleagues [41], the diet module was rather broad targeting multiple dietary behaviors; in other studies, individuals could choose more specific topics (eg, fat) which might be more relevant for some participants. Still, there were 110 participants in our study who were not interested in following either the diet or PA module, even though all participants were advised to choose at least the diet module. This is interesting since most studies did not investigate the explicit option of following no module, even though it is an important choice that might be common in the intervention practice and research.

The main aim of this study was to examine why some people are likely to make a particular module choice. Our results showed that age was the only demographic factor related to module choice. Older people were more likely to choose no module compared with the diet module. This is somewhat in contrast to our expectations and previous findings in which older people generally participated more in eHealth interventions in terms of start and (re)visit [14,15,17,19].

Surprisingly, amotivation was the only motivational construct found to be related to module choice, while we expected that especially intrinsic motivation was related to module choice. People who lacked motivation to become more physically active were less likely to choose both modules compared with the diet module alone, possibly because many already adhered to the MVPA guidelines and only those who were more motivated (ie, having lower amotivation scores for PA) were interested in the PA module. Next to these findings, it could still be that there are other motivational constructs that are more important in explaining module use, adherence/dropout, and behavioral change [6]; these will be investigated in a future study.

Previous research found that an already healthier lifestyle was related to module choice [23,24]. In our study, we only found that people who consumed more fruit per day were more likely to choose no module compared with the diet module. Almost half of the people who preferred no module already ate enough fruit according to the fruit consumption guideline of two servings per day. Sufficient fruit intake might be a reason why they were less interested in following the diet module. Eating behavior is very complex, and when people estimate how healthy their current diet is, they might use various indicators and fruit consumption may be an important one.

As expected, dietary advice did affect the modules people in our study chose. Participants who were strongly advised compared with those who were slightly advised to choose the diet module were more likely to choose both modules. It could be that these participants felt more need to change their lifestyle at a broader level. PA advice, on the other hand, did not affect module choice. These findings somehow validate our nondirectional goal of advice; many participants already adhered to the PA guideline but still preferred to choose the PA module in addition to the diet module. With regard to participant perceptions of how important they find eating healthy and getting sufficient PA, we found that people who considered it more important to eat more healthily during the opening session were less likely to choose no module compared with the diet module. To note, importance was measured after advice including feedback about their current behavior was given. It is conceivable that in this basic opening session awareness about their own unhealthy behavior was increased. As a result, people might have changed their perceptions on the importance of their dietary behavior. So in terms of the transtheoretical model of behavioral change, people who were in the precontemplation stage (ie, having no intention to change behavior) could have moved toward the contemplation stage (ie, intending to start the healthy behavior) as a result of the opening session [47]. It should further be noted that this model (including the program-related variables) had the highest explained variance and the best fit, meaning that compared with all other models we tested, this model explained our observed data the best.

By giving individuals the opportunity to not follow any of the offered modules—in accordance with SDT and MI principles [48,49]—we definitely lost people before the more specific parts of the intervention started. It is possible these participants were mainly interested in completing the questionnaires (eg, for the possibility of winning rewards). This may have resulted in additional dropout. However, participants received reminders in a later stage that they could still begin using the modules if they wanted, which has not been accounted for in this study. It would be interesting to see if the option of choosing no module

XSL•FO
RenderX

would lead to less dropout during the actual modules, as it is expected that only motivated participants are actually participating. It could also be that for those who chose to follow no module, the opening session provided sufficient information to fulfill their needs (for example, by receiving feedback on their PA and dietary behavior) and possibly even change their health behavior.

The findings of this study are preliminary, making it difficult to provide concrete recommendations for intervention development. Several factors were identified that were linked to choosing no module instead of the advised diet module: an older age, eating more fruit, and finding eating healthily less important. More research is needed to find out why those individuals are more likely to choose no module compared with the diet module, thus examining what could be improved (ie, more relevant content, better tailoring to specific groups, using MI to improve importance as early as possible) to make them more likely to choose the diet module. Future research is also necessary to investigate how and when MI could be applied to target those individuals who are more likely to choose no module and for whom the opening session is not sufficient to get them engaged in intervention activities and behavioral change.

Strengths and Limitations

To our knowledge, this is the first study to investigate to what extent the combined demographic characteristics, SDT/MI constructs, psychosocial factors, diet-related behavioral outcomes, and program-related variables (eg, advice) are related to participant module choice in a web-based tailored eHealth lifestyle intervention that is directed to more than one behavior. The findings of this study should be interpreted in the light of several limitations. First, we only used self-report to gather data. For example, even though the Dutch Short Questionnaire to Assess Health was the most feasible method to measure PA levels in our study, it may be subject to recall bias. In our study, a larger proportion of the people adhered to the PA guidelines compared with the general Dutch population, which could also be an overestimation of their actual PA level [46]. Second, our findings may not be fully generalizable to the entire Dutch adult population. A large part of our sample was highly educated. Even though our results showed that education did not affect module choice, generalizability may be questioned as a large proportion of the sample also adhered to the PA guideline, which may also concern overrepresentation [50]. Furthermore, women and people who are classified as obese were overrepresented as well, but age seems to be in line with the Dutch population. Third, attrition bias could play a role in several phases of the intervention period, such as between completing the baseline questionnaire but not completing the opening session (ie, making a module choice); this may threaten external and internal validity. This study did not provide insight into this matter; however, the process evaluation and effectiveness study will. Fourth, it may be that a combination of predictors and/or other predictors that are not measured in this study are also relevant in predicting module choice. No interaction effects were tested in this study since our sample size was not sufficient to test these effects on top of the already numerous predictors in our current analysis. Interactions between advice and the different types of motivation would be an interesting avenue to explore—for example, one can expect that people who score higher on controlled motivation might be more likely to follow advice within the intervention. Future research is necessary to examine additional relevant combinations of factors that could provide further insight into what predicts module choice.

Conclusions

Most participants chose to follow both the diet and PA modules in our web-based tailored lifestyle intervention. Our study has provided new insights into which characteristics are related to module choice in a lifestyle intervention. Additional research is necessary to examine how to target those individuals who are likely to choose no module and thereby who are at risk of dropping out in studies where this option is not explicitly offered. Most of the SDT/MI–related concepts did not affect module choice, suggesting that for initial module choice constructs from the SDT were less relevant than expected. These factors may, nevertheless, be relevant at a later stage of the intervention, and they could be related to participation or behavioral change. This will be examined in a follow-up study. A stepped approach to develop and pilot an intervention may be of relevance as various factors may be related to different aspects of an intervention (ie, module choice, participation, and behavioral change). To conclude, the findings in this study have important implications for developing eHealth interventions containing multiple health behaviors, as they can provide input for intervention improvement by providing insights that could help to optimally tailoring to the needs and characteristics of participants, such as making it more interesting for older adults, increase its effectiveness, and prevent dropout in a later stage.

Acknowledgments

We would like to thank Mark Lardinois for providing support with Tailorbuilder. This project was funded by an internal research fund of the Open University of the Netherlands.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation matrix for predictors.
[DOCX File , 36 KB - jmir_v22i7e15024_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Results from the stepwise multinomial logistic regression analysis predicting module choice.
[DOCX File , 17 KB - jmir_v22i7e15024_app2.docx]

Multimedia Appendix 3
Results from the model comparisons using likelihood ratio tests.
[DOCX File , 16 KB - jmir_v22i7e15024_app3.docx]

Multimedia Appendix 4
Variance inflation factors for each predictor in the stepwise multinomial regression analysis.
[DOCX File , 23 KB - jmir_v22i7e15024_app4.docx]

Multimedia Appendix 5
Results from the stepwise multinomial regression analysis using the full model.
[DOCX File , 39 KB - jmir_v22i7e15024_app5.docx]

References

1. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. Lancet 2010 Nov 20;376(9754):1775-1784. [doi: 10.1016/S0140-6736(10)61514-0] [Medline: 21074255]
2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
3. Pomerleau J, McKee M, Lobstein T, Knai C. The burden of disease attributable to nutrition in Europe. Public Health Nutr 2003 Aug;6(5):453-461. [doi: 10.1079/PHN2002456] [Medline: 12943561]
4. Broekhuizen K, Kroeze W, van Poppel MNM, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
5. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Feb 15;12(3):211-230. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]
6. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
7. Pelletier L, Dion S, Slovinec-D'Angelo M, Reid R. Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. Motiv Emotion 2004 Sep;28(3):245-277 [FREE Full text] [doi: 10.1023/B:MOEM.0000040154.40922.14]
8. Ryan R, Patrick H, Deci E, Williams G. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Health Psychol 2008;10:2-5 [FREE Full text] [Medline: 18648109]
9. Fortier MS, Duda JL, Guerin E, Teixeira PJ. Promoting physical activity: development and testing of self-determination theory-based interventions. Int J Behav Nutr Phys Act 2012 Mar 02;9:20 [FREE Full text] [doi: 10.1186/1479-5868-9-20] [Medline: 22385751]
10. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
11. Deci E, Ryan R. Self-determination theory: a macrotheory of human motivation, development, and health. Canadian Psychology/Psychologie canadienne 2008;49(3):182-185 [FREE Full text] [doi: 10.1037/a0012801]
12. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health 2007;28:413-433. [doi: 10.1146/annurev.publhealth.28.021406.144145] [Medline: 17150029]
13. Robroek SJW, Brouwer W, Lindeboom D, Oenema A, Burdorf A. Demographic, behavioral, and psychosocial correlates of using the website component of a worksite physical activity and healthy nutrition promotion program: a longitudinal study. J Med Internet Res 2010 Sep 30;12(3):e44 [FREE Full text] [doi: 10.2196/jmir.1402] [Medline: 20921001]
14. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
15. Schulz DN, Schneider F, de Vries H, van Osch LADM, van Nierop PWM, Kremers SPJ. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]

XSL•FO
RenderX

16. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
17. Van 'T Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
18. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based ehealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
19. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]
20. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
21. Kremers SPJ, Visscher TLS, Brug J, Chin A Paw MJM, Schouten EG, Schuit AJ, et al. Netherlands research programme weight gain prevention (NHF-NRG): rationale, objectives and strategies. Eur J Clin Nutr 2005 Apr;59(4):498-507. [doi: 10.1038/sj.ejcn.1602100] [Medline: 15714217]
22. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]
23. Schulz DN, Kremers SPJ, De Vries H. Tailored eHealth lifestyle promotion: which behavioral modules do users prefer? J Health Commun 2015;20(6):663-672. [doi: 10.1080/10810730.2015.1012243] [Medline: 25928161]
24. Kanera IM, Willems RA, Bolman CAW, Mesters I, Zambon V, Gijsen BC, et al. Use and appreciation of a tailored self-management eHealth intervention for early cancer survivors: process evaluation of a randomized controlled trial. J Med Internet Res 2016;18(8):e229 [FREE Full text] [doi: 10.2196/jmir.5975] [Medline: 27554525]
25. Quintiliani LM, Campbell MK, Bowling JM, Steck S, Haines PS, DeVellis BM. Results of a randomized trial testing messages tailored to participant-selected topics among female college students: physical activity outcomes. J Phys Act Health 2010 Jul;7(4):517-526 [FREE Full text] [Medline: 20683094]
26. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. Philadelphia: Springer; 1985.
27. Markland D, Ryan RM, Tobin VJ, Rollnick S. Motivational interviewing and self–determination theory. J Soc Clin Psychol 2005 Sep;24(6):811-831. [doi: 10.1521/jscp.2005.24.6.811]
28. Kelly CW. Commitment to health: a predictor of dietary change. J Clin Nurs 2011 Oct;20(19-20):2830-2836. [doi: 10.1111/j.1365-2702.2010.03654.x] [Medline: 21592243]
29. Coumans JMJ, Bolman CAW, Friederichs SAH, Oenema A, Lechner L. Development and testing of a personalized web-based diet and physical activity intervention based on motivational interviewing and the self-determination theory: protocol for the mylifestylecoach randomized controlled trial. JMIR Res Protoc 2020 Feb 04;9(2):e14491 [FREE Full text] [doi: 10.2196/14491] [Medline: 32014841]
30. Trial NL7333 (NTR7549). URL: https://www.trialregister.nl/trial/7333 [accessed 2020-07-02]
31. Friederichs SAH, Oenema A, Bolman C, Lechner L. Long term effects of self-determination theory and motivational interviewing in a web-based physical activity intervention: randomized controlled trial. Int J Behav Nutr Phys Act 2015;12(1):101 [FREE Full text] [doi: 10.1186/s12966-015-0262-9] [Medline: 26283094]
32. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [Medline: 9773724]
33. Ryan RM, Connell JP. Perceived locus of causality and internalization: examining reasons for acting in two domains. J Pers Soc Psychol 1989 Nov;57(5):749-761. [doi: 10.1037//0022-3514.57.5.749] [Medline: 2810024]
34. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ. Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors. Health Educ Res 2007 Oct;22(5):691-702 [FREE Full text] [doi: 10.1093/her/cyl148] [Medline: 17138613]
35. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exercise Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
36. Rollnick S, Butler CC, Stott N. Helping smokers make decisions: the enhancement of brief intervention for general medical practice. Patient Educ Couns 1997 Jul;31(3):191-203. [Medline: 9277242]
37. Sheeran P, Orbell S. Implementation intentions and repeated behaviour: augmenting the predictive validity of the theory of planned behaviour. Eur J Soc Psychol 1999 Mar;29(2-3):349-369. [doi: 10.1002/(sici)1099-0992(199903/05)29:2/3<349::aid-ejsp931>3.0.co;2-y]
38. Webb TL, Sheeran P. Integrating concepts from goal theories to understand the achievement of personal goals. Eur J Soc Psychol 2005 Jan;35(1):69-96. [doi: 10.1002/ejsp.233]
39. Huybrechts I, Börnhorst C, Pala V, Moreno LA, Barba G, Lissner L, IDEFICS Consortium. Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption

XSL•FO
RenderX

frequencies among European children. Int J Obes (Lond) 2011 Apr;35 Suppl 1:S69-S78. [doi: 10.1038/ijo.2011.37] [Medline: 21483425]
40. van Assema P, Brug J, Ronda G, Steenhuis I. The relative validity of a short Dutch questionnaire as a means to categorize adults and adolescents to total and saturated fat intake. J Hum Nutr Diet 2001 Oct;14(5):377-390. [Medline: 11906579]
41. Willems RA, Bolman CAW, Mesters I, Kanera IM, Beaulen AAJM, Lechner L. The Kanker Nazorg Wijzer (Cancer Aftercare Guide) protocol: the systematic development of a web-based computer tailored intervention providing psychosocial and lifestyle support for cancer survivors. BMC Cancer 2015;15:580 [FREE Full text] [doi: 10.1186/s12885-015-1588-z] [Medline: 26260318]
42. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [Medline: 14680666]
43. Helmerhorst HJF, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012;9:103 [FREE Full text] [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]
44. Brink E, van Rossum C, Postma-Smeets A, Stafleu A, Wolvers D, van Dooren C, et al. Development of healthy and sustainable food-based dietary guidelines for the Netherlands. Public Health Nutr 2019 Sep;22(13):2419-2435 [FREE Full text] [doi: 10.1017/S1368980019001435] [Medline: 31262374]
45. Weggemans RM, Backx FJG, Borghouts L, Chinapaw M, Hopman MTE, Koster A, Committee Dutch Physical Activity Guidelines 2017. The 2017 Dutch Physical Activity Guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58 [FREE Full text] [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
46. Het aandeel van de Nederlandse bevolking van 4 jaar en ouder dat voldoet aan de beweegrichtlijnen. 2017. URL: https://www.sportenbewegenincijfers.nl/kernindicatoren/beweegrichtlijnen [accessed 2019-02-12]
47. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
48. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd Edition. New York: The Guilford Press; 2013.
49. Ryan R. Handbook of Self-Determination Research. Rochester: University of Rochester Press and Boydell & Brewer; 2004.
50. Nicolaou M, Gademan MGJ, Snijder MB, Engelbert RHH, Dijkshoorn H, Terwee CB, et al. Validation of the SQUASH physical activity questionnaire in a multi-ethnic population: the HELIUS study. PLoS One 2016;11(8):e0161066 [FREE Full text] [doi: 10.1371/journal.pone.0161066] [Medline: 27575490]

Abbreviations

MI: motivational interviewing
MVPA: moderate to vigorous physical activity
PA: physical activity
SDT: self-determination theory
VIF: variance inflation factor

Edited by N Bruining; submitted 19.06.19; peer-reviewed by A Enrique, C Short, R Rahman, P Hsiao, I Mesters; comments to author 18.09.19; revised version received 13.01.20; accepted 11.03.20; published 23.07.20.

Please cite as:
Coumans JMJ, Bolman CAW, Oenema A, Lechner L
Predictors of Self-Determined Module Choice in a Web-Based Computer-Tailored Diet and Physical Activity Intervention: Secondary Analysis of Data From a Randomized Controlled Trial
J Med Internet Res 2020;22(7):e15024
URL: http://www.jmir.org/2020/7/e15024/
doi:10.2196/15024
PMID:32706659

©Juul MJ Coumans, Catherine AW Bolman, Anke Oenema, Lilian Lechner. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of an Unguided, Web-Based Distress Self-Management Program After Breast Cancer Diagnosis: Sub-Analysis of CaringGuidance Pilot Study

Robin M Lally[1*], PhD, MS, RN, AOCN, FAAN; Kevin Kupzyk[2*], PhD; Steve Gallo[3*], MS; Donna Berry[4*], PhD, RN, AOCN, FAAN

[1]College of Nursing, Fred & Pamela Buffett Cancer Center, University of Nebraska Medical Center, Omaha, NE, United States

[2]College of Nursing, University of Nebraska Medical Center, Omaha, NE, United States

[3]Center for Computational Research, Roswell Park Cancer Center, Buffalo, NY, United States

[4]School of Nursing, University of Washington, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Robin M Lally, PhD, MS, RN, AOCN, FAAN
College of Nursing
Fred & Pamela Buffett Cancer Center
University of Nebraska Medical Center
UNMC College of Nursing Center For Nursing Science
985330 Nebraska Medical Center
Omaha, NE, 68198-0001
United States
Phone: 1 4025595464
Email: robin.lally@unmc.edu

Abstract

Background: Unguided, web-based psychoeducational interventions are gaining interest as a way to reach patients while reducing pressure on clinical resources. However, there has been little research on how patients with cancer use these interventions.

Objective: The objective of this analysis was to evaluate how women newly diagnosed with breast cancer used the unguided web-based, psychoeducational distress self-management program CaringGuidance After Breast Cancer Diagnosis while enrolled in a pilot feasibility study.

Methods: Women with stage 0 to II breast cancer diagnosed within the prior three months were recruited from clinics primarily in the Northeastern United States for participation in a 12-week pilot study of CaringGuidance plus usual care versus usual care alone. Usage prompts included sets of emails sent weekly for 12 weeks; standardized congratulatory emails after every two hours of program use, and informative emails for each cognitive-behavioral exercise. Individual user activity on the site was automatically tracked by an analytics system and recorded directly in the CaringGuidance database.

Results: Complete usage data were available for 54 subjects. Ninety-eight percent of the intervention group logged into CaringGuidance independently at least once. Thirty-eight (70%) logged in during all three months, 15 (28%) were intermittent users, and one (2%) was a non-user. Users (n=53) averaged 15.6 (SD 9.85) logins. Mean logins were greatest in month 1 (7.26, SD 4.02) and declined in months 2 (4.32, SD 3.66) and month 3 (4.02, SD 3.82). Eleven (21%) used CaringGuidance with both the frequency and activity level intended at study outset, 9 (17%) exceeded intended frequency and activity (high-high users), and 10 (19%) were below expected usage on both login frequency and activity (low-low users). Low-low users and high-high users differed significantly (P<.001) in the total number of views and unique views of all program components. Change in depressive symptoms and the number of sessions (r=.351) and logins (r=.348) between study months 1 and 2 were significantly correlated (P=.018, .019). Higher baseline distress was associated with more unique views of program resources (r=.281, P=.043). Change in intrusive/avoidant thoughts from baseline to month 3, and the number of users' unique exercise views were negatively correlated (r=–.319, P=.035) so that more unique exercise views, equated with greater decline in intrusive/avoidant thoughts from baseline to month 3.

XSL•FO
RenderX

Conclusions: These findings favor the hypothesis that the key ingredient is not the amount of program use, but each user's self-selected activity within the program. More research is needed on the ideal ways to maintain use, and capture and define engagement and enactment of behaviors by people with cancer accessing unguided, self-management web-based programs.

(J Med Internet Res 2020;22(7):e19734) doi:10.2196/19734

KEYWORDS

eHealth; psychoeducation; supportive oncology; distress; self-management; oncology; breast cancer

Introduction

Background

Interest in unguided, web-based psychoeducational and behavioral interventions is growing due to ease of delivery and need for few clinical resources. The outcomes of such interventions, however, rely on patients accessing and using these programs independently [1]. Identifying how populations of patients use web-based programs and the best way to support individuals in meaningful program use and engagement is an emerging science [2]. Perceived program quality and usability [3] and supporting potential users' intent to and persistence in program use, are crucial to reducing use-attrition and increasing the likelihood that patients receive health benefits [4]. Duration of use may also be insufficient to describe the dose of web-based intervention necessary to achieve the intended benefit due to the inability to monitor off-line processing of program content, difficulty capturing the depth of online engagement [5,6], and not all users need to use the full program to attain their personal goals [7,8].

Moment-to-moment use of a website is one part of program users' engagement and has been coined, "microlevel engagement" [6]. Together with measures of the depth of user interaction with the program, which results in the enactment of behavior change (ie, macrolevel engagement), micro- and macro-level engagement result in "effective engagement" [6]. At least some amount of program use, measurable through the program's analytic system, must occur for there to be engagement and intervention effect. Analyses presented are focused upon the moment-to-moment use of a new unguided, web-based program for women with breast cancer.

Published reports on completely unguided, web-based psychosocial distress self-management interventions for adults with cancer are limited [9-13]. To the best of our knowledge, CaringGuidance After Breast Cancer Diagnosis [13-16] is one of only two such interventions specifically designed to address the psychosocial needs of women with breast cancer. The primary differences between CaringGuidance and the other program (BREATHE) [11] are that CaringGuidance is presented in English; was developed with input from, and includes vignettes of, Black as well as White American breast cancer survivors [16]; and is intended for initial use by women as soon as possible after diagnosis as opposed to after treatment has ended. While it is best to initiate cancer-related distress reduction to alleviate anxiety and depressive symptoms early after diagnosis [17], the initial post-diagnosis period is also busy with pre-treatment examinations, physician appointments, surgery, and chemotherapy. Therefore, it was expected that program use statistics for CaringGuidance might differ from unguided programs in which users access the program when treatment has been ongoing or is complete.

Objective

The objective of this analysis was to evaluate how women newly diagnosed with breast cancer used the unguided web-based, psychoeducational distress self-management program CaringGuidance while enrolled in its first pilot study [13-15]. The overall goal is to further inform the science of unguided web-based interventions by describing the (1) frequency, duration, and activity of CaringGuidance use by women newly diagnosed with breast cancer in total and by the month of study participation; (2) baseline characteristics of women who demonstrated high, moderate, low, and no program use; and (3) to evaluate high, moderate, and low program users' study completion, program satisfaction, enactment of distress management skills, and distress outcomes.

Methods

Participants

Women with stage 0 to II breast cancer diagnosed within the prior three months were recruited from clinics primarily in the Northeastern United States for participation in a 12-week pilot feasibility study of CaringGuidance plus usual care versus usual care alone. Enrollment was limited to stage 0–II breast cancer to reduce variation in the treatment experience among subjects. Details of recruitment and eligibility for this study of 100 women have been previously reported [13,14]. Institutional Review Board approval was received, and written consent obtained from all participants before study assessments. This feasibility pilot study, not involving drugs or devices, was not deemed eligible in 2013 for clinical trial registration by the university research administration.

Intervention

Baseline demographic and psychosocial measures of distress (ie, Distress Thermometer, Center for Epidemiologic Studies Depression Scale, and Impact of Events Scale), as well as Social Constraints Scale on spouse/partner and family/friends, were completed following written consent. These measures were repeated at months 1-3. Participants randomized to the CaringGuidance user condition received individual usernames and passwords. Usernames and passwords cannot be changed by users, thus permitting tracking use by assigned usernames.

CaringGuidance After Breast Cancer Diagnosis (version 1) is an unguided, web-based, psychoeducational program based on theories of stress and coping [18], and adjustment to illness through cognitive processing of life-threatening events [19-22]. The program's topical outline was guided by findings of the

XSL•FO
RenderX

PI's grounded theory describing women's thought processes following diagnosis [23,24]. Interventional components of CaringGuidance are based on cognitive-behavioral, problem-solving, and supportive oncology techniques, which have demonstrated efficacy in both in-person and web-based delivery for the breast cancer population [25,26]. From 2011-2013, the program was developed in an iterative process of review and revision by a team of oncology professionals, including psychologists, breast cancer survivors, web developers/programmers, and software engineers [16].

CaringGuidance program components include 5 learning modules divided into 17 topical sections (Textbox 1). Resources include 90 video vignettes filmed with 6 breast cancer survivors (ages 30-70, stage 0-III breast cancer, equal representation of Black and White American women) and 20 self-management "cognitive-behavioral homework" exercises (eg, visualization, cognitive reframing). There are also 13 resource modules consisting of a glossary, links to the program exercises, a library of full-length videos from which the vignettes were derived, breast cancer risk factors, signs of depression and anxiety, links to cancer resources, and myths and facts about breast cancer. The five modules are listed in tabs at the top of each page. Pages include a list of topical sections within each module to provide quick access (Multimedia Appendix 1). Users can orient and navigate using breadcrumb navigation at the top of each page and the titles of the next and previous sections appearing at the bottom of each page. Additional descriptions of CaringGuidance may be found elsewhere [14,15].

At login, first-time users are directed by the program through three introductory pages, including a welcome video from the program creator/PI, instructions for use, and a 12-statement tailoring exercise that guides users to program modules based on their self-selected greatest concern. The purpose of tailoring in this program is to help match the content to each user's needs with the expectation that this will increase content relevance [27] and contribute to users achieving their health goals [28]. Upon subsequent logins, users are directed to a personalized homepage but may access the introduction and tailoring exercise at any time. The 140 program components may be accessed by users at will in a flexible manner to direct their distress self-management. There is no required order in which to use the components nor requirements for completion before moving on to a different component.

Study participants were instructed that the suggested usage frequency and duration were 20-30 minutes for 2 to 3 times per week (range 40-90 minutes/week) for 12 weeks (480-1080 total minutes, 8-18 hours), but that they were permitted to access the program as much or as little as desired to create their own experience. The frequency/duration suggested was based on the traditional 12-week program of 1-hour counseling sessions. It was not expected that all users would access all components as everything within the program was not relevant to every user. Participants were informed during the consent process that program use would be tracked by the program's analytic system in a manner invisible to them.

At enrollment, research staff showed participants the login page and the first page containing the welcome video but did not go further into the program. When enrollment was completed by email, an explanation of the first three program pages was provided in the email containing the user's login information. All participants received a hard copy pictorial guide on basic website use (eg, how to enlarge the font and adjust the volume).

During the study, usage prompts included standard emails sent weekly for 12 weeks supporting continued use or encouraging use; standard congratulatory emails after every two hours of program use, and pre-written informative emails for each cognitive-behavioral exercise sent following two or more minutes of a user accessing that component (these emails also directed users to relevant program components for self-reflection). All emails were automated or sent by staff using a CaringGuidance Gmail account to simulate an automated message, thus avoiding personal interaction. To assess safety and review of the daily symptom/support log maintained by all participants, one research assistant phoned monthly for a scripted conversation. These calls averaged 18.8 minutes (SD 7.73) in month 1, down to a mean 13.5 minutes (SD 4.64) in month 3. The script directed the research assistant to suggest portions of the program not yet accessed during the call. Calls were recorded and 10% reviewed by the principal investigator to assure script adherence. Emails and calls placed by participants to the research office for technical support (n=6) were tracked [14].

XSL•FO
RenderX

Textbox 1. CaringGuidance modules and topical sections

- Are my reactions normal?
  - Fears and anger
  - Exploring other emotions
  - Why might I think about cancer differently than other women?
- What does this diagnosis mean?
  - Why me? or Why not me?
  - Questions and misconceptions
- Who am I now?
  - Self-concept
  - Accepting support
  - Is a support group right for me?
  - The meaning of survivor
  - Body image and sexuality
  - How will people act toward me now?
- What are strategies to care for myself?
  - Coping with cancer
  - Talking with people around you
  - Personal control strategies
- Moving forward
  - Personal growth from this experience
  - How much will cancer be a part of my life?
  - Setting healthy goals

Usage Data Storage and Retrieval

The CaringGuidance web analytics system was developed by our team to collect user activity information that would facilitate the analysis of various usage modalities. The activity of individual users on the site was automatically tracked by the system and recorded directly in the CaringGuidance database. Each time a user visited an individual page on the site (referred to as a "page visit"), the analytics system collected the information specified in Textbox 2.

Textbox 2. Information collected by the analytics system.

- The name and internal identifier of the user,
- Whether or not the activity is a login event (eg, the user logged into the site using their username and password),
- A session identifier that facilitates the tracking of a user's activity during a particular visit,
- Whether or not the user viewed a resource, article, exercise, or video and which of these items were viewed
- A timestamp with second resolution

Usage Definitions and Measures

Page visits and login information stored in the database were used to construct a set of sessions that provided details as to how each user utilized the site during that time. We defined a session as a set of consecutive page visits, starting with a login event and ending with a logout or a period of inactivity. The amount of time that a user spent on any given page visit was calculated using the difference between the timestamps of the current page visit and the next page visit within the session. Because users may simply close their web browser, turn off their computer, or leave their computer for some time in addition to clicking the "logout" button, we did not have a reliable method for calculating the time spent on the final page visit in a session (Figure 1). To overcome this limitation, we used a configurable period of inactivity to infer a logout event and expired the session—currently set to 30 minutes. In the case where a user's session expired, and they later returned to the

XSL•FO
RenderX

site, this activity was treated as a new login event and initiated a new session for that user (Figure 2). The activity/inactivity rules provided a lower bound for the time that a user visited CaringGuidance during any given session.

Figure 1. User session comprised of 4-page visits where clocks represent page view timestamps. The user visited for at least 20 minutes; however total time is underestimated as it is not known how long the user spent on the site during page visit 4.

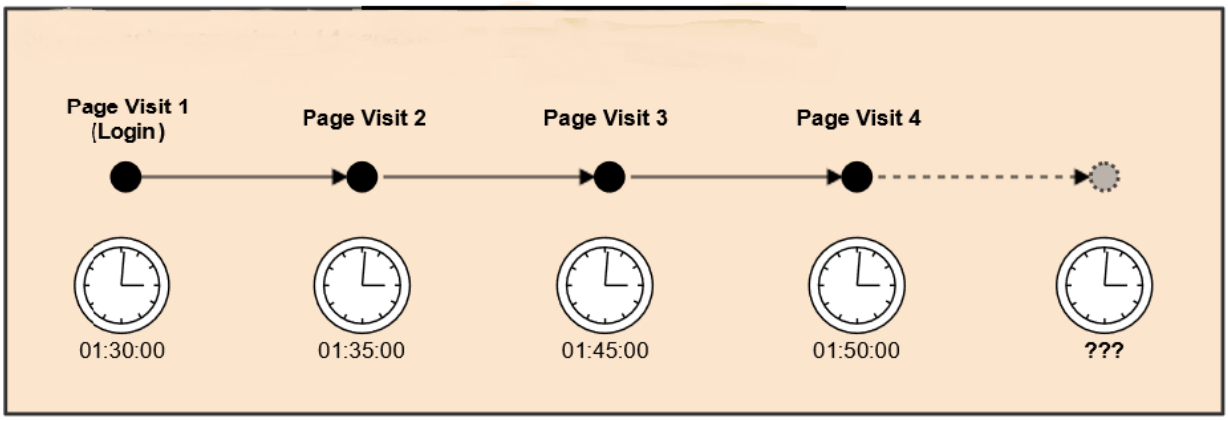
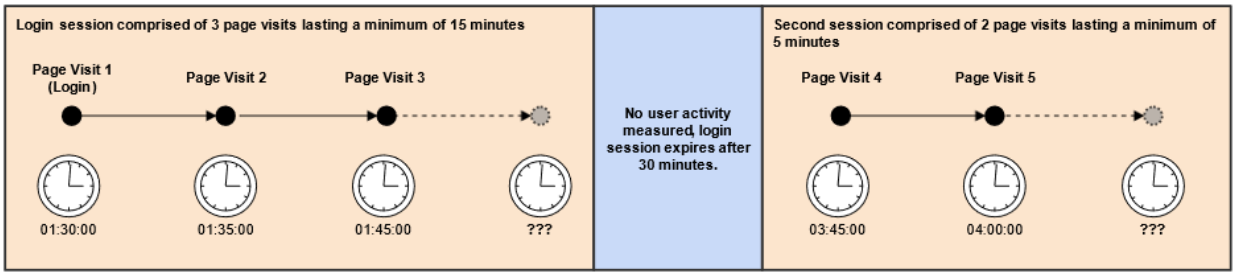


Figure 2. User session wherein a period of inactivity triggers use to be treated as two individual sessions.



Usage

In this current micro-level engagement analysis, usage was defined as frequency, duration, and user program activity. This definition is consistent with definitions of usage applied in prior research [29]. The *frequency* was defined as the number of logins to CaringGuidance per participant per study month over the 3-month study enrollment. Frequency also was characterized by the number of sessions in which users engaged per month and over the study enrollment period, since more than one session may have been engaged in during a single login (Figure 2). The *duration* was defined as the total number of minutes logged in each month and over the 3-month study period. The *activity* was defined as the number of total page views and the number of unique page views per participant per month and overall for each of the CaringGuidance program components (ie, module, exercise, videos, and resource pages). Finally, activity was examined for which components appeared to be preferred based on use and repeated access.

Users and Non-Users

It was expected that participants would access CaringGuidance with varying frequency and duration for multiple reasons, including the recency of their cancer diagnosis, levels of distress, and ongoing treatment. *Non-users* were defined as study participants who received a username and password but who registered ≤1 log-ins/sessions and for whom zero minutes were recorded. Consistent with definitions described by van den Berg et al [29], *continuous users* were defined as registering at least one login/session per month for each of the 3 study months. *Intermittent users* were defined as those who logged in initially and registered at least one login/session in study month 1 but then registered <1 session in study month 2 or 3, or both study months 2 and 3.

Macrolevel Engagement

For this pilot study, macro-level engagement [6] was measured with three questions rated from 1 (strongly disagree) to 5 (strongly agree) on the self-report satisfaction survey at the conclusion of month 3 [14]. These questions were, “I used things I learned from CaringGuidance to change (a) my thoughts about breast cancer; (b) how I talked or acted around people; and (c) my self-care behaviors.

Participant Usage Groups

In order to characterize CaringGuidance usage concerning baseline and monthly psychological outcomes, user patterns were divided into three usage groups designated as low, moderate, and high. In so doing, we were inspired by definitions used by van den Berg et al [29]. For our usage analysis, *low frequency* was defined as 1-12 logins total throughout the 12 weeks; in other words, 50% or less of the minimum number of logins suggested to participants (ie, 24 logins) at enrollment. *Moderate frequency* was defined as 13-24 logins throughout the study, and *high frequency* was defined as ≥25 logins over the 12 weeks. Activity was likewise defined as *low activity*

XSL•FO
RenderX

equal to opening 0-25% of the program's 140 components (ie, unique views), *moderate activity* equal to opening 26-50%, and *high activity* equal to opening 51-100% of the program components. High activity was defined as viewing >50% of the components because not all components applied to all study participants. Logins and activity were cross-tabulated to characterize use as *low-low*, *moderate-moderate*, or *high-high*.

Total duration logged into the program was not used in the usage group calculation because of varying speeds of Internet connections, differing participants' reading speed, differing times needed to use various program components, and the analytics systems' inability to measure the duration of the last session, thus underestimating total time on the program.

Results

Demographics

Study participants were women diagnosed with stage 0 to II breast cancer within the prior three months at baseline. Ability to read English and access a computer with Internet service and email were required since the program at that time was not mobile accessible. Fifty-seven participants were randomized to the CaringGuidance condition; however, one participant did not receive a password until week 5, and two participants withdrew after randomization due to feeling too busy to participate. Thus, CaringGuidance analytic data were available for 54 participants.

These 54 participants ranged in age from 36 to 78 years (mean 55.02, SD 9.4). They reported prior experience using the Internet, with a median of 1-hour Internet use (range 10 to 600 minutes) per day at baseline. Of these 54 participants, the majority were White (n=50, 93%) and married/partnered (n=32, 59.3%), while 22 (40.7%) were single/divorced/widowed. Forty-nine of the 54 (91%) had attended at least some college. Most were employed full-time (n=33, 61.1%) while 10 (18.5%) reported part-time employment at baseline, and of those reporting income (n=47), 53.2% had a household annual income of $75,000 or above. See Textbox 3 for clinical characteristics. Full demographic data on the 100 participants enrolled in the pilot study have been previously reported [13].

Textbox 3. Clinical characteristics and treatment during the study (N=54).

- Cancer stage at baseline
  - 0 (n=13)
  - I (n=24)
  - II (n=16)
  - "early" (n=1)
- Time since diagnosis at baseline
  - <4 weeks (n=20)
  - 1-2 months (n=28)
  - 2-3 months (n=6)
- Breast surgery procedure during the study (n=29)
- Chemotherapy received during the study (n=33)
- Radiation therapy received during the study (n=43)

Attrition

Eight participants assigned to the CaringGuidance condition withdrew or were lost to follow up, meaning that they did not complete all study psychosocial measures [14]. These include the 3 participants noted above who did not receive their login information or withdrew after randomization. The other 5 participants continued to use CaringGuidance despite not completing all psychosocial measures. Four of these were intermittent users (ie, logged in only in month 1) while one was a continuous user (ie, logged in in all three study months). These five participants did not differ demographically at baseline from others assigned to the CaringGuidance condition.

Of the 54 CaringGuidance participants for whom usage data are available, 38 (70%) were continuous users, 15 (28%) were intermittent users, and one (2%) was a non-user. The non-user was age 62, Black, unemployed and widowed, with Stage 0 breast cancer. She completed the psychosocial surveys and monthly calls with the research assistant, during which she indicated her intention to use a library computer to access CaringGuidance, but transportation barriers prevented her from doing so. The "non-user" was eliminated from this analysis and findings for the 53 "users" (continuous or intermittent) with mean age 54.9 years (SD 9.4), and 94% White are reported here.

Frequency

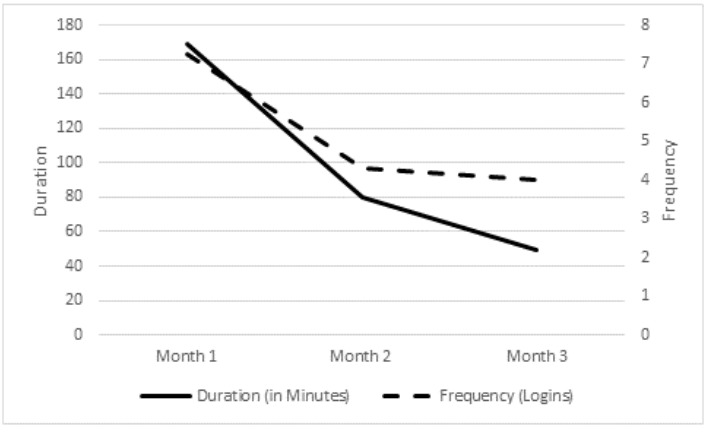
Users (n=53) logged in an average of 15.6 (SD 9.85) times during their 12-week access period. The mean number of logins was highest in month 1 (mean 7.26; SD 4.02) and declined after that such that the mean logins in month 2 were 4.32 (SD 3.66) and 4.02 (SD 3.82) in month 3. Overall, login attrition was significant across all three study months ($F_{2,104}$=28.9, P<.001) with the sharpest decline in logins occurring between study months 1 and 2 ($F_{1,52}$=38.8, P<.001) (Figure 3).

XSL•FO
RenderX

Users averaged 16.94 sessions (SD 10.42) over 12 weeks. Session attrition mimicked login attrition, as would be expected, with the most considerable decline in sessions between months 1 (mean 7.85, SD 4.34) and 2 (mean 4.83, SD 4.28; $F_{1,52}$=27.4, P<.001) and remaining stable from month 2 to month 3 (mean 4.26, SD 3.84; $F_{1,52}$=17.0, P=.217).

With one exception, no significant correlations were identified between use frequency (number of logins or sessions) and change in psychological outcomes; the one exception being a significant positive correlation between change in depressive-symptoms and the number of sessions (r= .351) and logins (r= .348) between study months 1 and 2 (P=.018 and .019 respectively), but not between any other study months.

Figure 3. Change in login frequency and duration of program use over 3 months.



Duration

Time spent on CaringGuidance by users during their 12 weeks of access ranged from 9.27 to 1265.3 minutes (21.1 hours; mean 4.98 hours, SD 3.61). The mean duration of use declined significantly over the 3 study months from 169.38 minutes (SD 120.08) in month 1 to 49.06 minutes (SD 45.06) in month 3 ($F_{2,104}$=40.2, P<.001). The decline in duration of use was also significant when examined by month; from month 1 to month 2 ($F_{1,52}$=33.7, P<.001) and month 2 to month 3 ($F_{1,52}$=7.9, P=.007).

The total duration of CaringGuidance use was found to have a significant positive relationship (r=.291, P=.036) with users' baseline intrusive/avoidant thinking such that users with more initial intrusive/avoidant thinking used CaringGuidance for more total minutes over the 3 study months [14,15]. This was also true for baseline spouse/partner derived social constraints and the total duration of CaringGuidance use (r=.370, P=.031) and minutes of use in study month 2 (r=.422, P=.013) and month 3 (r=.345, P=.045) [15]. No additional statistically significant correlations were found between duration and baseline or month 3 psychological outcomes.

Activity

Modules

The 17 written learning component pages (modules) were accessed 1 to 103 times per user (mean 41.11, SD 26.57). All 53 users accessed at least one module. There were between 1 and 17 *unique* module page views per user (mean 12.06, SD 4.65), meaning that all module pages were accessed at least once. Eleven users (21%) accessed all 17 module pages. The five most viewed modules were: "Fears and Anger" (211 views), "Why me? Why not me?" (218 views), "Self-concept" (242 views), "Personal growth from this experience" (243 views), and "Coping with cancer" (290 views). Least accessed modules were: "Is a support group right for me?" (51 views), and "How will people act toward me now?" (59 views).

Exercises

Each of the 20 cognitive-behavioral exercises was accessed from 33 to 199 times [14]. Fifty-two of the 53 users (98%) viewed exercises yielding between 0 and 162 exercise views per user (mean 38.91, SD 34.69). Of these, there were between 0 and 20 *unique* exercise views (mean 12.06, SD 5.61), meaning that some users viewed the same exercises multiple times. Repeat viewing was encouraged by several exercises that directed users to review prior exercises as a means to self-monitor changes in thinking over time. Six of 53 users (11%) viewed all 20 exercises.

Videos

Fifty users (94%) accessed videos resulting in between 0 and 119 video views per user (mean 29.7, SD 28.29). Unique video views ranged from 0-80 (mean 23.91, SD 21.17). Eighty-nine of the 90 videos (99%) were viewed. The 10 most viewed videos were viewed between 29 to 42 times and featured four of the six survivors (two Black and two White American survivors). An overall theme of these most viewed videos was self-concept as a newly diagnosed survivor.

Resources

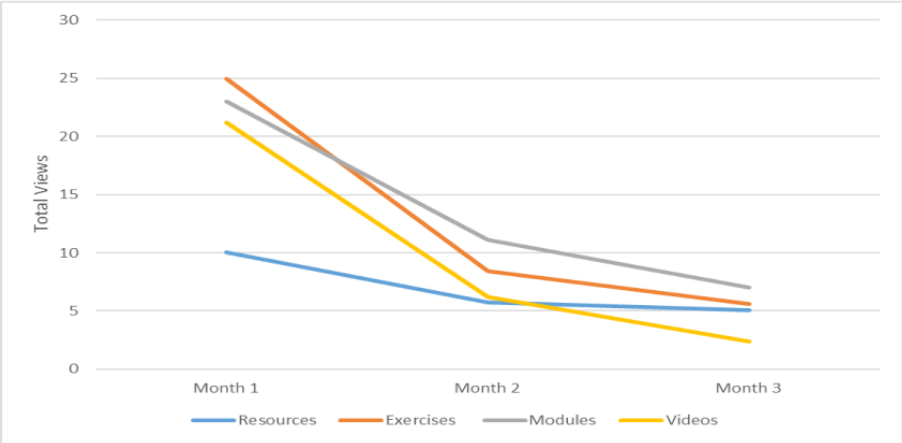
Each of the 13 resource components was accessed from 30 to 252 times by program users. Fifty-one of 53 users (96%) accessed at least one resource page. Three users (6%) accessed all resource pages. Resource pages were accessed between 0 and 67 times per user (mean 20.91, SD 14.93). On average, 7.62 (SD 3.62) of the resource page views were *unique*, indicating that some users returned to the same resource pages multiple times to review the content. The five most accessed pages were:

XSL•FO
RenderX

breast cancer risk factors (96 times), cancer information resources (98 times), common questions about breast cancer (99 times), mindfulness-based stress reduction (109 times) and all program exercises list (252 times). Least accessed were questions to ask your doctor (30 times) and first appointments—talking with your doctors (40 times).

Overall, the most significant program activity occurred in month 1 for all components and significantly decreased between months 1 and 2 (P<.001). Views of program modules and videos continued to decrease between study months 2 and 3 (P=.004 and .003) (Figure 4).

Figure 4. Program component views over 3 study months.



Psychosocial Associations

Baseline overall distress correlated positively with activity such that higher baseline distress was associated with more unique views of program resources by users (r=.281, P=.043). Additionally, at study month 3, the number of unique module views was negatively associated with spouse/partner derived social constraints (r=–.418, P=.021). In other words, more unique modules were viewed by subjects when perceived social constraints were lower.

When the *change* in overall distress, depressive-symptoms, and intrusive/avoidant thinking was examined, no correlations were found with user activity, with one exception. A significant negative correlation (r=–.319, P=.035) was identified between the change in intrusive/avoidant thoughts from baseline to month 3 and the number of users' unique exercise views. In other words, the higher number of unique exercise views, the more significant the decline in intrusive/avoidant thoughts from baseline to month 3.

Use Groups

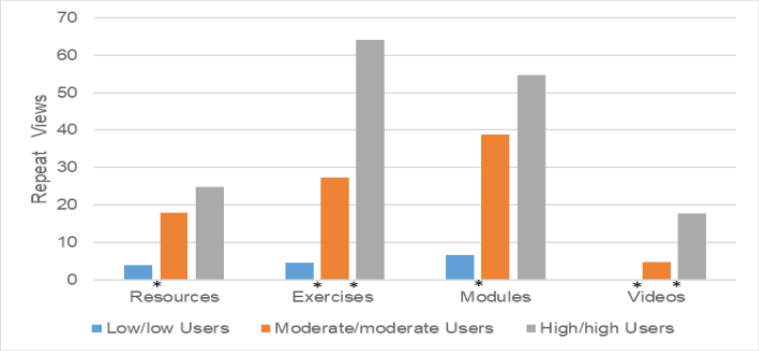
Eleven of 53 (21%) users were defined as moderate program users. In other words, they used CaringGuidance in the range that approached the intended usage told to them at study entry for *both* frequency and activity (*mod-mod*). Nine users (17%) exceeded intended usage for *both* login frequency and activity (*high-high*), while 10 users (19%) were below expected usage on *both* login frequency and activity (*low-low*). In total, 39 (74%) users were in the moderate to high activity range during the period they were logged in the program, and 30 (57%) logged-in with moderate to high frequency. Overall frequency and activity were highly associated (r=.565, P<.001). Low-low users and high-high users differed significantly (P<.001) on the total number of views and unique views of all program components.

Low Frequency/Low Activity (Low-Low)

Three of the 10 low-low users were also part of the group who did not complete any of the study's monthly mailed psychological assessments, and one of these subjects formally withdrew due to feeling too busy to participate. The low-low use group's number of logins ranged from 2 to 10 (mean 5.2, SD 3.26). This group consisted of the one subject of all 53 users who viewed no exercises, the two subjects who viewed no resources, and the three subjects who viewed no videos. Low/low user activity involved 0 to 13 unique exercise views per subject (mean 5.3, SD 3.62) of the possible 20, 1 to 14 unique module views per subject of a possible 17 (mean 5.9, SD 3.87), 0 to 10 unique views of resource pages per subject of a possible 13 (mean 3.7, SD 3.62), and 0-18 unique video views per subject of a possible 90 (mean 4.0, SD 5.44). Low-low users returned to view individual study components, on average, 3.8 (resources) to 6.6 (modules) times, exclusive of videos. No videos were viewed a second time by these users (Figure 5). Despite being low on frequency *and* activity, 30% of subjects in the low-low group were "continuous" users, logging-in 8 to 10 times each over the 12-week study.

XSL•FO
RenderX

Figure 5. Repeat views by user group. An asterisk indicates significant differences between adjacent groups (left to right): Resources *P*=.002; Exercises *P*<.001 and .012; Modules *P*<.001; Videos *P*=.001 and .016.



High Frequency/High Activity (High-High)

In contrast to the low-low user group, the 9 high-high users returned all monthly psychosocial assessments and completed the study. All 9 were "continuous" users logging-in from 25 to 30 times (range) throughout the study (mean 27.3, SD 2.0). Their activity involved 15 to 20 unique exercise views per subject of a possible 20 (mean 18.56, SD 1.67; 33% accessed all 20 exercises), 10 to 17 unique module views per subject of a possible 17 (mean 15.56, SD 2.35; 56% accessed all of the modules), 8 to 13 unique views of resource pages out of a possible 13 (mean 10.56, SD 1.74; 1 subject accessed all resource pages), and 34 to 74 unique video views per subject of a possible 90 unique video views (mean 51.0, SD 14.76). High-high users returned to view program components an average of 17.7 (videos) to 64.1 (exercises) times. The most return views were to exercises and modules (mean 54.8 views). This behavior resulted in significant differences in repeat viewing behavior between usage groups (Figure 5).

Demographic Characteristics of User Groups

There were no significant differences identified in the demographic characteristics (age, race, education, employment status, income, breast cancer stage, pre/post-op status, and self-rated knowledge about breast cancer at baseline) of low, moderate, and high login frequency users nor users with varying amounts of activity. Additionally, no significant difference was found between pre-study self-reported daily Internet minutes across low, moderate, or high CaringGuidance login frequency or activity. This remained true when low-low and high-high groups' demographic characteristics were compared.

Psychosocial Outcomes by User Group

Study outcomes on psychological distress indicated that access to CaringGuidance versus not was favorable for the intervention group as evidenced by significant differences in the reduction of distress and depressive symptoms between study months 2 and 3 [13]. When low, moderate, and high number of logins and low, moderate and high activity groups were examined individually, or combined low-low, mod-mod, or high-high, no significant differences in change for overall distress, depressive-symptoms, intrusive/avoidant thinking, or reports of social constraints from family/friends or spouse/partners were found from baseline to month 3 or between study months. In other words, no usage group was superior to another regarding the change in distress over the study period as represented by these variables.

User Satisfaction Survey Completion

The CaringGuidance user satisfaction survey was completed by 60% (n=6) of the low-low users and 100% of high-high users. Low-low and high-high users did not differ significantly in CaringGuidance satisfaction except that the high-high users perceived that using CaringGuidance increased their knowledge about their breast cancer (mean 4.56, SD 0.73) versus mean 3.50 (SD 0.84) of possible 5 points among low-low users, (*P*=.049). Forty-nine percent of survey respondents reported that CaringGuidance use changed their thoughts about breast cancer, 44% indicated that program content led them to change self-care behavior, and 40% reported that what they learned from CaringGuidance changed how they talked or behaved with people [14].

Discussion

Principal Findings

This analysis explored the unguided use of a web-based psychoeducational distress self-management program by women recently diagnosed with breast cancer. We described the characteristics of users and non-users, the frequency and duration of use, users' activity within the program, and outcomes associated with different use patterns.

Ninety-eight percent of the women who proceeded in the study logged into the CaringGuidance program independently at least once after randomization, a very positive result given reports of other studies in which rates of initial login were less than 50% [30]. Our data closely approximates that of similar unguided programs tested in a research environment where 90% of those assigned to the intervention logged in [29]. Those who volunteer for research are likely a motivated population and, of course, are aware that they have committed to study participation. On the other hand, the usage results from this study are notable because subjects not only voluntarily logged into a website on their own time, they did so after recently receiving likely the worst news of their lives—a cancer diagnosis, and while making hospital and clinics visits to undergo tests, surgery, chemotherapy and radiation treatments.

XSL•FO
RenderX

Of course, use attrition was significant after month 1, but this is also typical of web-based interventions [30-33]. We did not collect information on users' reasons for waning use, which could represent that some users achieved their goal for using the program sooner than others (e-attainers) [8]. Research continues to be needed to understand better how people recently diagnosed with cancer may be motivated to use and engage with unguided programs.

We also found that, consistent with prior research, total duration users spent on CaringGuidance did not correlate with psychological outcomes. Neither did the sheer number of logins nor the number of program components viewed correspond with distress as measured in this study. In other words, simply more use was not better. Researchers have argued that it is the depth of engagement with the program and the ability to glean what the user desires to support their needs that matters [5,6]. Evidence varies, however, among studies concerning the effect of higher amounts of intervention use on psychological outcomes, with some finding more logins and time spent reduces distress [34]. In contrast, other studies found the number of logins and duration not to affect depression or anxiety [35]. The variability in findings is likely related to the uniqueness of interventions' targeted populations, behaviors targeted by the intervention, and how program use is defined, supporting the need to evaluate micro and macro-engagement for specific interventions [6].

Based on our findings, we hypothesize that given at least a minimal amount of program use, it is the content that viewers access and whether that content engages and satisfies the user that holds the most significant import to psychological outcomes. While this hypothesis requires additional testing, support is provided by our finding that the higher number of unique exercise views made by users, the more significant their decline in intrusive/avoidant thoughts. This finding supports the mechanism of action of the cognitive-behavioral influenced program exercises, which are intended to assist users to reframe their thinking and process the cancer experience, in turn reducing intrusive/avoidant thoughts over time.

Additionally, our findings demonstrated that women who reported higher depressive symptom scores logged in with greater frequency over months 1 to 2; in turn, significant differences between months 2 and 3 in depressive symptoms favored the intervention over control [13]. Although it is acknowledged that correlation does not inform us whether depressive symptoms led to more frequent logins or visa-versa, it is logical to presume that if logging in increased women's feelings of depression, they would have stopped this voluntary activity.

Repeated views of program components favor the hypothesis that women found value in using CaringGuidance. If users found no value, they would not have returned dozens of times to multiple components, as we saw in this study. Similar conclusions have been drawn in prior studies [29]. Not only did we note repeated viewing of program components, but a greater propensity among high frequency/high activity users compared to other users to return to all components while showing repeat interest, particularly in program exercises and written content. This finding is promising because these components convey the program's cognitive-behavioral ingredients, and thus, further development of these types of components is supported. Also promising is that the most accessed program modules focused on coping, personal growth from the cancer experience, and supporting survivors' self-concept. The most viewed videos also dealt with self-concept, indicating that CaringGuidance users independently focused on program components that were meaningful for distress reduction through confronting and re-evaluating how cancer reshaped their identity and world view which are vital to cognitively processing the cancer experience [19-24].

Finally, we were surprised to find that there was no relationship between high distress and greater use of program videos. At the time of design, it was anticipated that women with more considerable distress at baseline would gravitate toward the program videos as an activity that took less focused attention. At the same time, those further in their ability to cognitively process the diagnosis would attend to program exercises. Breast cancer survivors involved in program development expressed concern over overwhelming distressed women, which in turn inspired our addition of more videos, the glossary of terms, and "Questions for your doctor" resource components. The hypothesis seemed to be somewhat supported in that women distressed at baseline gravitated to the program resources component, however women with intrusive/avoidant thoughts and experiencing spouse/partner social constraints used the overall program with greater duration and women with depressive symptoms used the program with greater frequency.

Limitations

A limitation of this usage analysis was the potential underestimation of the total time users spent on the program because of the tracking system's inability to define the time users spent on their last page visit of a session. This underestimation may be balanced somewhat by the fact that all participants in this pilot study received scripted monthly calls from the research assistant about which the primary purpose was to assess for adverse study events but included questions about users' ability to login and find the information they sought in the program. Although it was not found that users viewed program components that were discussed during the calls, the contact may have prompted more program use than if users had only received the email prompts [36,37]. Other researchers have also found phone calls to be less effective than email prompts [38]. Data were not collected on whether email prompts were opened, although the sender (CaringGuidance program) likely acted as a weekly reminder. Additional limitations include the small, educated, and racially homogenous sample of women whom all had prior Internet experience and computer access, thus limiting generalizability. Lastly, although data were collected on user satisfaction and enactment of behaviors gleaned from the program, data were not systematically collected that would allow for analysis regarding reasons for use attrition.

Conclusions

In conclusion, distressed women recently diagnosed with breast cancer self-selected the CaringGuidance program components that satisfied their needs and used these components with

XSL•FO
RenderX

considerable variability in frequency, duration, and activity. These findings favor the hypothesis that the key ingredient is not the amount of use, but rather the self-selected activity of each user within the program. Given the ease of accessibility and low resource utilization associated with CaringGuidance as well as the safety, satisfaction, and preliminary efficacy findings [13-15], CaringGuidance offers a potential clinically implementable option for distress management in this population. Future work should focus on implementation such as encouraging women's acceptance of mental health online support, helping users to find the components within the program that they desire, increase users' motivation to continue using to their maximum benefit and to explore the depth of engagement and cognitive processing in which users engage off-line. Overall, more information is needed on the ideal ways to capture and define engagement and enactment of behaviors by people with cancer accessing unguided, self-management web-based programs.

Acknowledgments

RL is the creator and led the development and testing of CaringGuidance. SG developed the analytical data system for CaringGuidance and was responsible for data retrieval. KK conducted all data analyses for this manuscript. All authors contributed to the writing, review, and approval of this manuscript for publication. Funding for the development and testing of CaringGuidance was provided by an American Cancer Society Mentored Research Scholar Grant MRSG-11-101-01-CPPB – Principal Investigator RL. RL receives research funding from the National Institute of General Medical Sciences, 1U54GM115458-01, and Pfizer in collaboration with the National Comprehensive Cancer Network #36030539. CaringGuidance After Breast Cancer Diagnosis Copyright 2016 The Research Foundation for The State University of New York, licensed for academic research at the University of Nebraska.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of pages from CaringGuidance™ After Breast Cancer Diagnosis Copyright © 2016 The Research Foundation for The State University of New York, licensed for academic research to the University of Nebraska.
[PNG File , 912 KB - jmir_v22i7e19734_app1.png]

References

1. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
2. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
3. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
4. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
5. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
6. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
7. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
8. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
9. Goldberg JI, Schulman-Green D, Hernandez M, Nelson JE, Capezuti E. Self-Management Interventions for Psychological Distress in Adult Cancer Patients: A Systematic Review. West J Nurs Res 2019 Oct;41(10):1407-1422. [doi: 10.1177/0193945919845104] [Medline: 31007160]

XSL•FO
RenderX

10. Beatty L, Kemp E, Coll JR, Turner J, Butow P, Milne D, et al. Finding My Way: results of a multicentre RCT evaluating a web-based self-guided psychosocial intervention for newly diagnosed cancer survivors. Support Care Cancer 2019 Jul;27(7):2533-2544. [doi: 10.1007/s00520-018-4526-1] [Medline: 30411239]
11. van den Berg SW, Gielissen MFM, Custers JAE, van der Graaf WTA, Ottevanger PB, Prins JB. BREATH: Web-Based Self-Management for Psychological Adjustment After Primary Breast Cancer--Results of a Multicenter Randomized Controlled Trial. J Clin Oncol 2015 Sep 01;33(25):2763-2771. [doi: 10.1200/JCO.2013.54.9386] [Medline: 26169621]
12. Chambers SK, Ritterband LM, Thorndike F, Nielsen L, Aitken JF, Clutton S, et al. Web-Delivered Cognitive Behavioral Therapy for Distressed Cancer Patients: Randomized Controlled Trial. J Med Internet Res 2018 Jan 31;20(1):e42 [FREE Full text] [doi: 10.2196/jmir.8850] [Medline: 29386173]
13. Lally RM, Kupzyk KA, Bellavia G, Hydeman J, Gallo S, Helgeson VS, et al. CaringGuidance™ after breast cancer diagnosis eHealth psychoeducational intervention to reduce early post-diagnosis distress. Support Care Cancer 2020 May;28(5):2163-2174 [FREE Full text] [doi: 10.1007/s00520-019-05028-0] [Medline: 31414245]
14. Lally RM, Bellavia G, Gallo S, Kupzyk K, Helgeson V, Brooks C, et al. Feasibility and acceptance of the CaringGuidance web-based, distress self-management, psychoeducational program initiated within 12 weeks of breast cancer diagnosis. Psychooncology 2019 Apr;28(4):888-895. [doi: 10.1002/pon.5038] [Medline: 30803084]
15. Lally RM, Kupzyk K, Mills A, Gallo S, Meneses K. Effects of social constraints and web-based psychoeducation on cancer-related psychological adjustment early-after breast cancer diagnosis. J Psychosoc Oncol 2019;37(6):677-698. [doi: 10.1080/07347332.2018.1546787] [Medline: 31631813]
16. Lally RM, McNees P, Meneses K. Application of a novel transdisciplinary communication technique to develop an Internet-based psychoeducational program: CaringGuidance™ After Breast Cancer Diagnosis. Appl Nurs Res 2015 Feb;28(1):e7-11. [doi: 10.1016/j.apnr.2014.10.006] [Medline: 25467063]
17. Alfano CM, Leach CR, Smith TG, Miller KD, Alcaraz KI, Cannady RS, et al. Equitably improving outcomes for cancer survivors and supporting caregivers: A blueprint for care delivery, research, education, and policy. CA Cancer J Clin 2019 Jan;69(1):35-49 [FREE Full text] [doi: 10.3322/caac.21548] [Medline: 30376182]
18. Folkman S, Greer S. Promoting psychological well-being in the face of serious illness: when theory, research and practice inform each other. Psychooncology 2000;9(1):11-19. [doi: 10.1002/(sici)1099-1611(200001/02)9:1<11::aid-pon424>3.0.co;2-z] [Medline: 10668055]
19. Creamer M, Burgess P, Pattison P. Reaction to trauma: A cognitive processing model. Journal of Abnormal Psychology 1992;101(3):452-459. [doi: 10.1037/0021-843x.101.3.452]
20. Lepore S. A social-cognitive processing model of emotional adjustment to cancer. In: Baum A ABL, editor. Psychosocial interventions for cancer. Washington DC: American Psychological Association; 2001:99-116.
21. Lepore S, Helgeson V. Social constraints, intrusive thoughts, and mental health after prostate cancer. J Soc Clin Psychol 1998;17(1):89-106 [FREE Full text] [doi: 10.1521/jscp.1998.17.1.89]
22. Redd W, DuHamel K, Johnson Vickberg SM. Long-term adjustment in cancer survivors: integration of classical-conditioning and cognitive-processing models. In: Baum, A ABL, editor. Psychosocial interventions for cancer. Washington DC: American Psychological Association; 2001:77-97.
23. Lally RM. Acclimating to breast cancer: a process of maintaining self-integrity in the pretreatment period. Cancer Nurs 2010;33(4):268-279. [doi: 10.1097/NCC.0b013e3181d8200b] [Medline: 20467302]
24. Lally RM, Hydeman JA, Schwert K, Henderson H, Edge SB. Exploring the first days of adjustment to cancer: a modification of acclimating to breast cancer theory. Cancer Nurs 2012;35(1):3-18. [doi: 10.1097/NCC.0b013e318227ca62] [Medline: 22173485]
25. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
26. Ye M, Du K, Zhou J, Zhou Q, Shou M, Hu B, et al. A meta-analysis of the efficacy of cognitive behavior therapy on quality of life and psychological health of breast cancer survivors and patients. Psychooncology 2018 Jul;27(7):1695-1703. [doi: 10.1002/pon.4687] [Medline: 29500842]
27. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
28. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
29. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MFM, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. J Med Internet Res 2013 Aug 19;15(8):e170 [FREE Full text] [doi: 10.2196/jmir.2566] [Medline: 23958584]
30. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
31. Morgan C, Mason E, Newby JM, Mahoney AEJ, Hobbs MJ, McAloon J, et al. The effectiveness of unguided internet cognitive behavioural therapy for mixed anxiety and depression. Internet Interv 2017 Dec;10:47-53 [FREE Full text] [doi: 10.1016/j.invent.2017.10.003] [Medline: 30135752]

XSL•FO
RenderX

32. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
33. Owen JE, Bantum EO, Gorlick A, Stanton AL. Engagement with a social networking intervention for cancer-related distress. Ann Behav Med 2015 Apr;49(2):154-164 [FREE Full text] [doi: 10.1007/s12160-014-9643-6] [Medline: 25209353]
34. Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J. Exposure to a patient-centered, Web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015 Jun 03;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]
35. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
36. Berry DL, Blonquist TM, Nayak MM, Grenon N, Momani TG, McCleary NJ. Self-Care Support for Patients with Gastrointestinal Cancer: iCancerHealth. Appl Clin Inform 2018 Oct;9(4):833-840 [FREE Full text] [doi: 10.1055/s-0038-1675810] [Medline: 30463094]
37. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
38. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]

Edited by G Eysenbach; submitted 29.04.20; peer-reviewed by P Chow, D Attai; comments to author 21.05.20; revised version received 01.06.20; accepted 04.06.20; published 06.07.20.

Please cite as:
Lally RM, Kupzyk K, Gallo S, Berry D
Use of an Unguided, Web-Based Distress Self-Management Program After Breast Cancer Diagnosis: Sub-Analysis of CaringGuidance Pilot Study
J Med Internet Res 2020;22(7):e19734
URL: https://www.jmir.org/2020/7/e19734
doi:10.2196/19734
PMID:32628117

©Robin M Lally, Kevin Kupzyk, Steve Gallo, Donna Berry. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adherence to Established Treatment Guidelines Among Unguided Digital Interventions for Depression: Quality Evaluation of 28 Web-Based Programs and Mobile Apps

Stefan Bubolz[1], BSc; Gwendolyn Mayer[1], Dipl-Psych; Nadine Gronewold[1], MSc; Thomas Hilbel[2], Prof Dr; Jobst-Hendrik Schultz[1], PD, Dr med

[1]Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg, Germany

[2]Westphalian University of Applied Sciences, Gelsenkirchen, Germany

Corresponding Author:
Nadine Gronewold, MSc
Department of General Internal Medicine and Psychosomatics
Heidelberg University Hospital
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 6221 56 38176
Email: nadine.gronewold@med.uni-heidelberg.de

Abstract

Background: Web-based interventions for depression have been widely tested for usability and functioning. However, the few studies that have addressed the therapeutic quality of these interventions have mainly focused on general aspects without consideration of specific quality factors related to particular treatment components. Clinicians and scientists are calling for standardized assessment criteria for web-based interventions to enable effective and trustworthy patient care. Therefore, an extensive evaluation of web-based interventions at the level of individual treatment components based on therapeutic guidelines and manuals is needed.

Objective: The objective of this study was to evaluate the quality of unguided web-based interventions for depression at the level of individual treatment components based on their adherence to current gold-standard treatment guidelines and manuals.

Methods: A comprehensive online search of popular app stores and search engines in January 2018 revealed 11 desktop programs and 17 smartphone apps that met the inclusion criteria. Programs and apps were included if they were available for German users, interactive, unguided, and targeted toward depression. All programs and apps were tested by three independent researchers following a standardized procedure with a predefined symptom trajectory. During the testing, all web-based interventions were rated with a standardized list of criteria based on treatment guidelines and manuals for depression.

Results: Overall interrater reliability for all raters was substantial with an intraclass correlation coefficient of 0.73 and Gwet AC1 value of 0.80. The main features of web-based interventions included mood tracking (24/28, 86%), psychoeducation (21/28, 75%), cognitive restructuring (21/28, 75%), crisis management (20/28, 71%), behavioral activation (19/29, 68%), and relaxation training (18/28, 64%). Overall, therapeutic meaningfulness was rated higher for desktop programs (mean 4.13, SD 1.17) than for smartphone apps (mean 2.92, SD 1.46).

Conclusions: Although many exercises from manuals are included in web-based interventions, the necessary therapeutic depth of the interventions is often not reached, and risk management is frequently lacking. There is a need for further research targeting general principles for the development and evaluation of therapeutically sound web-based interventions for depression.

(J Med Internet Res 2020;22(7):e16136) doi:10.2196/16136

KEYWORDS

web-based interventions; depression; mHealth; mental health; telemedicine; mobile phone; eHealth; electronic mental health; online therapy

XSL•FO
RenderX

Introduction

Background

Major depression is the leading cause of disability and affects more than 300 million people worldwide [1], with a lifetime prevalence up to 20.6% for different populations [2-4]. Even minor forms of depression are associated with reduced quality of life [5], increased mortality [6], and functional impairments [7]. Additionally, the economic consequences of depression are extensive, and the treatment costs are increasing [8]. Although psychotherapy is an established, evidence-based treatment for depression [9-12], many individuals in need of mental health services do not receive adequate treatment [13,14]. Access barriers to treatment that are often associated with mental illness include limited availability of psychotherapists, long waiting lists, high costs of treatment, and fear of stigmatization [15-19]. Web-based interventions are considered a promising model to address this treatment gap [13,14]. They provide immediate support at any time and place, are cost-effective, and can easily be administered to a vast majority of people, thus reducing barriers and enabling high accessibility to treatment [20-22].

The number of concepts for web-based interventions has been gradually increasing, including telemedicine, electronic health apps, smartphone health apps, or mobile health apps. Furthermore, web-based interventions can be classified according to various criteria, ranging from smartphone apps for mere self-management to guided programs that include the support of a therapist through either asynchronous communication (such as email) or synchronous communication (such as videoconferencing). In many cases, a clear and distinctive assignment of a web-based intervention to a single type is not possible [23]. In this study, we followed the definition of a web-based intervention as an intervention program that is (1) available via the internet, (2) provides interactive components, (3) comprises health-related material, (4) aims to create positive change, and (5) aims to improve knowledge and understanding of a disease [24]. Following this definition, we included interventions that provide initial access via the internet, whereas it was not mandatory for the intervention to remain connected to the internet throughout its course. Additionally, we distinguished between smartphone apps and desktop programs because of their differences in technological features (eg, use of push notifications in smartphones and better readability of text on desktop screens).

The increase of innovative technologies in the field of web-based interventions shows great promise for potential in improving quality of life for people suffering from depression. Web-based treatment for depression has proven to be effective in reducing symptoms [25-29], reducing stigma [30-32], and improving depression literacy [31,32]. Furthermore, web-based interventions can deliver therapy to users independent of location with 24-hour accessibility [33].

Consumers searching for mental health apps most likely use social media, online search engines, or web forums [34]. A recent study identified more than 380,000 health apps worldwide, 28% of which are targeted toward mental health and behavioral disorders [35]. For users in the United States, more than 500 apps for depression are currently available [36]. Nevertheless, only a small proportion of the web-based interventions available for depression are supported by evidence-based studies [37], and research on web-based interventions has thus far focused mainly on validating single products [38]. Criticisms of web-based interventions include poor methodological quality of effectiveness studies such as by comparing interventions with wait-list control groups or by showing evidence only through analysis of short-term effects [23,39].

When web-based interventions lack evidence or disregard standards and guidelines for psychotherapeutic treatment, they may offer ineffective care or cause direct harm to users, such as by providing adverse advice or diverting users from accessing effective treatment [40-42]. Another problem that many web-based interventions face is a high attrition rate because many users drop out after a few days or weeks of using the intervention [43,44]. Reasons for low user engagement, especially with smartphone-based mental health apps, were recently summarized in a review by Torous et al [45]. Low usability, lack of user-centric design, concerns about privacy, lack of trust by users and by clinicians, and concerns about unhelpfulness in emergencies were the most relevant issues identified. Thus, “building trust through standards” [45] is considered to be the most important, yet challenging, goal for developers of high-quality web-based interventions.

Clinicians and scientists are calling for standardized assessment criteria for web-based interventions to enable effective and trustworthy patient care [23,46]. As a result, professional associations such as the Anxiety and Depression Association of America have started to develop their own criteria [47], but further research is necessary.

Previous Research on Quality Measures for Web-Based Interventions

There are several approaches adopted in previous research to develop appropriate assessment tools for ensuring the therapeutic quality of web-based interventions for depression.

Renton et al [48] provided a scoping review and evaluated web-based interventions for depression on a 28-point rating system, covering aspects such as usability, accessibility, and type of tools used (eg, worksheets and assessments). They found high variability among the web-based interventions, and that most of these interventions used cognitive behavioral therapy (CBT) as the therapeutic approach with mood assessments and homework sheets implemented as the primary interactive tools. These results are supported by another recent review showing that 85% of the web-based interventions for depression implemented CBT techniques [49].

One prominent and often used scale is the Mobile App Rating Scale (MARS) [50], a multidimensional measure that includes 19 items on 4 objective quality indicators for apps: engagement, functionality, esthetics, and information quality. Additionally, 4 items measure the subjective quality of an app [50]. MARS has been validated with well-being apps and the instrument showed high internal consistency (Cronbach α=.90) [50]. The MARS has been used for a broad scope of apps, including

XSL•FO
RenderX

smoking cessation apps [51], fitness apps for cancer patients [52], and German depression apps [53].

However, Baumel et al [54] noted that current criteria-based rating scales such as MARS lack an assessment of therapeutic alliance principles. Therefore, the authors developed a detailed evaluation tool for mobile and web-based health interventions termed "Enlight," which includes a quality assessment section with 6 constructs: usability, visual design, user engagement, content, therapeutic persuasiveness, and therapeutic alliance (ie, basic acceptance and support). The instrument was validated with mobile apps and web-based programs for behavioral change in the case of medical illness or mental health, and showed high internal consistency (Cronbach α=.83 to .90). Although Enlight recognizes the therapeutic content of web-based interventions, only 4 items are dedicated to this issue.

Previous studies have measured the therapeutic quality of web-based interventions only at a general level, reflecting the overall impression of raters, without specific analysis of particular treatment components (eg, behavioral activation, cognitive restructuring) or therapeutic approaches. Although the majority of web-based interventions implement CBT as the therapeutic approach [48,49], we found only one study that examined the realization of a specific component of CBT in web-based interventions in detail (ie, behavioral activation) [55]. The authors found that the utility of these interventions is questionable, because only a few adhere to the core principles of CBT.

Qu et al [56] systematically examined the functionalities of depression apps and found that 31% of the apps evaluated offer depression screening, 66% offer tracking functionalities (eg, mood tracking), and 83% offer some form of therapeutic intervention (eg, psychoeducation or thought diaries). However, the therapeutic quality of these interventions was not the focus of their study.

To our knowledge, no studies published to date have aimed at specifically evaluating the therapeutic quality of web-based interventions at the level of several individual treatment components. As clinicians and scientists are calling for standardized assessment criteria, and previous measures did not examine the individual treatment components of web-based interventions for depression at an in-depth level, in the present study, we evaluated the therapeutic quality of currently available web-based interventions for depression.

Study Aim

The aim of this study was to evaluate the quality of unguided web-based interventions for depression at the level of individual treatment components based on their adherence to gold-standard treatment guidelines and manuals. Our primary research question was: How extensively do web-based interventions for depression adhere to established treatment standards? This includes (1) which core treatment components of established guidelines were realized in the web-based intervention, (2) how close did the treatment components follow the recommendations of the guidelines regarding their delivery mode and instruction, and (3) how potential risks in the treatment process are managed. We did not aim to provide recommendations for or against individual programs or apps. As a means to achieve the study aims, we examined current treatment guidelines and manuals for depression, and developed a questionnaire that comprises standardized testing criteria.

Methods

Search Strategy and Inclusion Criteria

In January 2018, three researchers independently searched for online treatment programs targeted at depression. Since we aimed at searching for web-based interventions from a patient perspective, the searches were carried out in three major app stores (Google Play, iTunes, Windows Store) and four broadly popular search engines (Google, Yahoo, Bing, Duckduckgo). Additionally, we investigated the Beacon website [57] (an Australian platform for health apps) and HealthOn [58] (a German platform for health apps) for web-based interventions meeting our inclusion criteria. We did not search published evidence in the scientific literature because it is not clear when a web-based intervention reported in the literature becomes available on the market. Further, we excluded web-based interventions that required participation in a study. We found that it is more likely for patients to participate in a noncommittal web-based intervention with easy access when they are looking for online treatment options, which is not the case in a clinical study. The search terms used were "depression" or the wildcard search term "depress*" either by itself or in combination with one or more of the following terms: "online," "web-based," "treatment," and "program."

Apps or desktop programs were included in our analysis based on the following criteria: (1) claimed to provide treatment or support for depression, (2) were accessible to the public via the internet (with or without fee), (3) had an interactive component (ie, were not purely educational) and required user participation or input (eg, homework, worksheets, mood assessments), (4) were available in English or German (because these were the fluently spoken languages by all raters), (5) provided a scientific basis for their treatment (eg, based on CBT), and (6) were targeted to adults (older than 18 years). We adapted inclusion criteria that can be found in previous scoping reviews on web-based interventions for depression in the scientific literature [48] to fit the purpose of this study (criteria 1-4), and added further criteria that we considered necessary to fulfill the particular objective of this study (criteria 5-6).

Apps or desktop programs were excluded from this study if they (1) only provided information regarding depression (ie, psychoeducation) and offered no further intervention, (2) did not claim to be based on a scientific background, (3) did not specifically target depression (ie, were targeted at other or multiple disorders), (4) were not accessible to the public (ie, programs for patients of a specific clinic), (5) targeted health care professionals for training purposes, (6) offered only mood tracking, (7) were only available for research purposes (ie, user must be enrolled in a study to access the program), (8) offered no treatment program, (9) could not be completed within the home or private setting (ie, must attend classes), (10) required the participant to get in touch with a counselor (ie, webcam

XSL•FO
RenderX

counseling, therapy sessions, chatroom counseling), and (11) refused participation in this study.

Programs were screened for relevance based on the title, description, and available further information given on the respective webpages (eg, frequently asked questions, videos, or screenshots).

In line with previous work that identified web-based intervention programs for depression [48], we ceased the search when no new programs could be identified within five consecutive pages of search results.

Testing Criteria

The original questionnaire used in this study was developed in German. A translated version of the complete questionnaire can be found in Multimedia Appendix 1.

To obtain objective testing criteria, we examined current guidelines and manuals that are established in the treatment of depression, including the S3 and National Health Care guideline on unipolar depression [59,60], Beck's manual for CBT [61], and the "Coping with Depression" course [62].

The treatment guidelines were examined by three researchers who each extracted treatment components and developed a set of objectively ratable criteria that could possibly be implemented by a web-based intervention. Subsequently, the researchers compared their criteria and solved discrepancies by consensus. The criteria were then reviewed by medical experts of the psychosomatic and psychosocial field, who provided comments and compared the criteria to the guidelines once more. The researchers used this information to revise the criteria conclusively, and any discrepancies were solved by consensus.

Following this process, we chose CBT components as superordinate categories in our questionnaire because CBT was identified as the most frequent approach in web-based interventions for depression [48,49]. Based on the guidelines and manuals, we expected the following components to occur in web-based interventions that claim to be evidence-based: behavioral activation, cognitive restructuring, psychoeducation, mood tracking, journal keeping, relaxation training, social skills training, resource activation, and crisis management.

The questionnaire includes a general and a specific part for each component, except for resource activation and crisis management. The general part comprises the same items for each component and aims to address basic principles such as transparency, understandability, or therapeutic potential. Example items for the general part are: "The therapeutic background for the intervention is presented," "The instruction can easily be understood," and "Possible difficulties regarding the intervention are addressed."

The specific part comprises items unique to the therapeutic tools or the theoretical rationale of each component. For example, in case of cognitive restructuring, items for the specific part were: "The principle of automatic thoughts is explained clearly," "The program suggests alternative thoughts for negative thoughts," and "The program offers the option to create a daily record of negative thoughts."

The questionnaire closes with crisis management and resource activation. For crisis management, items concerning the instruction, therapeutic background, digital implementation, and therapeutic meaningfulness were included from the general part. Further items asked for emergency contacts, relapse prevention, contact to support groups, and behavior in case of suicidality. Resource activation comprises 4 items concerning the identification of resources and their reflection in the course of the intervention.

The developed criteria do not claim to be exhaustive. Rather, we tried to find as many objectively ratable aspects as possible that can characterize a good and CBT-conforming psychotherapy. The absence of a treatment component or a particular criterion does not necessarily imply poor quality of the intervention. However, the presence of many criteria can indicate high quality.

Testing Procedure

The three independent researchers who conducted the search also rated the included web-based interventions following a standardized testing procedure. Raters followed the recommendations for usage given by each web-based intervention regarding the time and duration of usage (eg, daily usage for several minutes or weekly completion of session). When programs consisted of sessions, the programs were tested until all sessions relevant to the identified treatment components were completed. When provided, we followed the suggested order and instructions of each web-based intervention (eg, completing 12 sessions over 12 weeks). When a web-based intervention was designed to be used at will and no clear point of completion of the treatment could be identified, it was tested in an intuitive manner (ie, interventions were completed in the order in which they appeared to the user) until each intervention was completed at least once.

When symptom questionnaires, mood tracking, or progress tracking were included in a web-based intervention, raters responded with a set of previously defined symptoms. At the beginning of the web-based intervention, raters responded with a moderate symptom severity and changed the occurrence and severity of symptoms during the course of the intervention. Symptoms were chosen following a moderate depression score according to Beck's depression inventory II [63]. When repeated assessments were included during the progression of a web-based intervention, each researcher submitted a deterioration of symptoms, characterized by increased severity of core symptoms and the addition of suicidal thoughts, to test how the web-based intervention responded to suicidal tendencies and how it provided risk management. At later stages, each rater submitted an improvement of symptoms, characterized by mild degrees of core symptoms and loss of additional symptoms, to test how the program responded to treatment progress.

All raters have a degree in psychology and backgrounds in psychotherapy (CBT, psychodynamic therapy, and systemic therapy). Web-based interventions were randomly allocated to the raters and each intervention was independently tested by two raters.

XSL•FO
RenderX

No ethical approval was required because no human participants outside the researchers took part in this study.

Statistical Analysis

Descriptive statistical analysis (including means, SDs, and frequencies) and intraclass correlation coefficient (ICC) estimates were calculated using SPSS statistical package version 24 (IBM Corp Released 2016, IBM SPSS Statistics for Windows, Armonk, NY, USA).

To assess the interrater reliability for all reported metric items, ICC estimates and their 95% CIs were calculated for each rater pair based on a one-way random-effects model with absolute agreement.

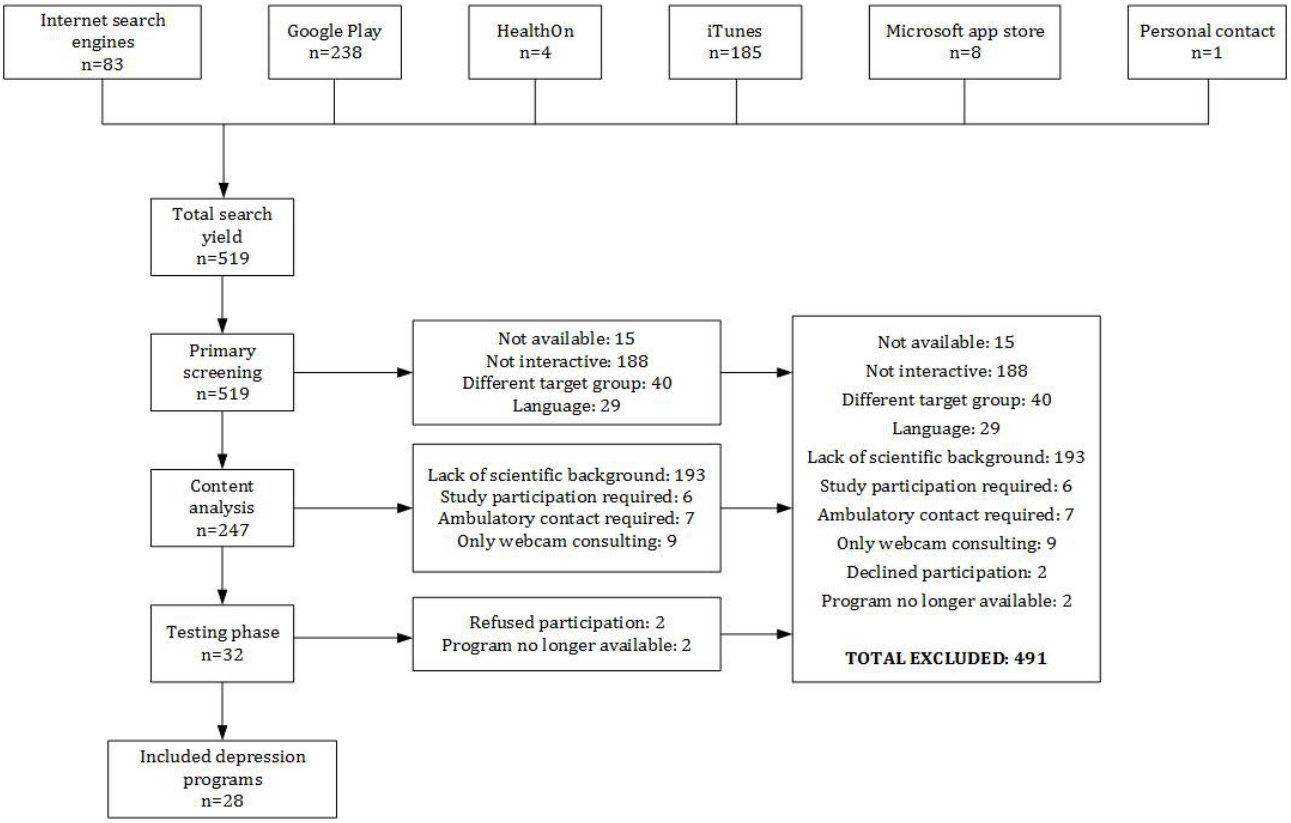
To assess the interrater reliability for all reported dichotomous items, Gwet AC1 statistic was calculated using formula 4.1 of Gwet [64] through the WINPEPI software developed by Abramson [65]. Gwet AC1 statistic is a chance-corrected measure of the extent of agreement between raters but has been recommended for use because it is less influenced by differences in individual rater tendencies to give positive ratings and by differences in the prevalence of the response categories, thus making it more robust and less biased compared to other metrics [66-68].

Results

Search Results

The initial search yielded 519 web-based interventions. Thirty-two of these were identified to meet our inclusion criteria. Discrepancies regarding the identified programs were resolved by consensus. During our testing period, 2 programs withdrew their participation from the study and 2 programs were no longer publicly available, leaving 28 programs that were ultimately tested (see Figure 1 for a flowchart of the screening and inclusion process). The 28 programs included 11 desktop programs and 17 smartphone apps. A list of the tested web-based interventions can be found in Multimedia Appendix 2.

Figure 1. Flowchart of the screening process to identify desktop programs and apps.



Interrater Reliability

Mean ICC estimates and Gwet AC1 statistics for all reported metric items are shown in Table 1. The overall range for ICC estimates was 0.66 to 0.79. According to the guideline developed by Koo and Li [69], mean ICC estimates in this study were indicative of moderate to good interrater reliability for all rating pairs. The overall range for the Gwet AC1 statistic was 0.75 to 0.84. Concordance for Gwet AC1 statistic is measured on the same scale as used for Landis and Koch's κ criteria [70,71]; therefore, interrater reliability for the dichotomous items in this study can be considered to be substantial.

XSL•FO
RenderX

Table 1. Intraclass correlation (ICC) estimates and Gwet AC1 statistics for all rating pairs.

Rating pair	ICC (95% CI)	Gwet AC1 (95% CI)
Rating pair 1 and 2	0.79 (0.72-0.87)	0.84 (0.79-0.89)
Rating pair 1 and 3	0.70 (0.58-0.83)	0.75 (0.68-0.83)
Rating pair 2 and 3	0.66 (0.51-0.81)	0.80 (0.74-0.87)
Overall	0.73 (0.67-0.79)	0.80 (0.76-0.84)

General Results

An overview of the identified treatment components for each web-based intervention can be found in Multimedia Appendix 3. Table 2 depicts the relative number of treatment components for each web-based intervention. We did not include resource activation in this summary because no web-based intervention assigned a specific module to resource activation, but rather implemented resource activation in several treatment components.

Table 2. Absolute numbers and respective availability of treatment components for desktop programs (N=11) and apps (N=17).

Treatment component	Desktop programs, n (%)	Apps, n (%)
Behavioral activation	9 (82)	10 (59)
Cognitive restructuring	8 (73)	13 (76)
Psychoeducation	10 (91)	11 (65)
Mood tracking	11 (100)	13 (76)
Journal keeping	4 (36)	8 (47)
Relaxation training	8 (73)	10 (59)
Social skills training	6 (55)	2 (12)
Crisis management	9 (82)	11 (65)

Taken together, the presentation of the therapeutic background of interventions received higher ratings in desktop programs (mean 4.13, SD 1.17) than in apps (mean 2.92, SD 1.46) and moderate ratings overall (mean 3.47, SD 1.46). Regarding the understandability of instructions, desktop programs received higher ratings than apps, and high ratings were given desktop programs and apps overall (mean 4.18, SD 1.15). The digital implementation of interventions was rated as moderate for both desktop programs and apps (mean 2.90, SD 1.17). Overall, interventions received moderate ratings for their therapeutic meaningfulness for desktop programs and apps (mean 3.54, SD 1.15). However, discrepancies regarding the therapeutic quality were identified between single interventions, with desktop programs receiving higher scores for behavioral activation, cognitive restructuring, and psychoeducation than apps, and similar scores for mood tracking and relaxation training (see Table 3 and Table 4 for detailed results).

XSL•FO
RenderX

Table 3. Mean (SD) scores of assessments of the therapeutic quality in interventions of desktop programs.

Item	Overall[a]	BA[b]	CR[c]	PE[d]	MT[e]	JK[f]	RT[g]	SST[h]	CM[i]
"The therapeutic background for the intervention is presented."	4.13 (1.17)	4.89 (0.32)	4.94 (0.25)	3.65 (1.40)	3.36 (1.40)	4.25 (0.71)	4.13 (1.15)	4.25 (0.87)	4.00 (1.19)
"The instruction can easily be understood."	4.70 (0.62)	5.00 (0.00)	4.81 (0.40)	4.70 (0.57)	4.59 (0.85)	4.63 (0.74)	4.56 (0.73)	4.58 (0.67)	4.67 (0.59)
"The (digital) implementation of the intervention is adequate."	3.04 (1.09)	3.28 (0.83)	3.63 (1.03)	3.20 (1.11)	2.55 (1.06)	3.00 (1.31)	2.56 (1.03)	3.25 (1.14)	3.11 (0.83)
"The intervention is therapeutically meaningful."	3.92 (1.07)	4.56 (0.62)	4.69 (0.60)	4.05 (1.05)	3.14 (1.04)	4.00 (0.93)	3.31 (1.08)	3.83 (1.47)	3.94 (0.73)

[a]Overall: includes all desktop programs.

[b]BA: behavioral activation.

[c]CR: cognitive restructuring.

[d]PE: psychoeducation.

[e]MT: mood tracking.

[f]JK: journal keeping.

[g]RT: relaxation training.

[h]SST: social skills training.

[i]CM: crisis management.

Table 4. Mean (SD) scores on assessments of the therapeutic quality in interventions of smartphone apps.

Item	Overall[a]	BA[b]	CR[c]	PE[d]	MT[e]	JK[f]	RT[g]	SST[h]	CM[i]
"The therapeutic background for the intervention is presented."	2.92 (1.46)	3.20 (1.32)	3.42 (1.42)	2.68 (1.59)	2.58 (1.45)	3.19 (1.33)	2.80 (1.47)	3.00 (2.31)	2.59 (1.40)
"The instruction can easily be understood."	3.74 (1.31)	3.80 (1.15)	3.77 (1.28)	3.64 (1.47)	4.04 (1.08)	4.06 (1.12)	3.80 (1.40)	3.50 (1.91)	3.18 (1.50)
"The (digital) implementation of the intervention is adequate."	2.83 (1.28)	2.80 (1.47)	2.73 (1.25)	2.68 (1.52)	3.04 (1.11)	3.19 (1.05)	2.75 (1.29)	2.00 (0.82)	2.36 (1.00)
"The intervention is therapeutically meaningful."	3.22 (1.12)	3.50 (1.15)	3.38 (1.17)	3.36 (1.26)	3.08 (1.06)	3.25 (0.78)	3.05 (1.05)	3.75 (0.96)	2.86 (1.28)

[a]Overall: includes all apps.

[b]BA: behavioral activation.

[c]CR: cognitive restructuring.

[d]PE: psychoeducation.

[e]MT: mood tracking.

[f]JK: journal keeping.

[g]RT: relaxation training.

[h]SST: social skills training.

[i]CM: crisis management.

Behavioral Activation

Interventions aimed at behavioral activation were included in 9 desktop programs and 10 apps. All behavioral activation interventions suggested activities, including activating activities and relaxing activities, whereas fewer interventions suggested social activities (16/19, 84%).

The suggested activities were rated to be easy to realize (mean 4.47, SD 0.89) and pleasant (mean 4.26, SD 0.89). Complex activities were rated as not divided into achievable intermediate steps (mean 1.74, SD 1.33). Difficulties that can occur in the realization of suggested activities were rated as not sufficiently addressed (mean 2.24, SD 1.52).

The relative share of web-based interventions realizing further features of behavioral activation can be found in Table 5.

XSL•FO
RenderX

Table 5. Absolute numbers and relative share of web-based interventions addressing specific features of behavioral activation in desktop programs (N=9) and apps (N=10).

Features of behavioral activation	Desktop programs, n (%)	Apps, n (%)
Difficulty levels for activities	1 (11)	1 (10)
Push notifications	9 (100)	2 (20)
Add individual activities	9 (100)	9 (90)
Schedule activities	9 (100)	5 (50)
Reminders for activities	9 (100)	4 (40)
Check completed activities	1 (11)	1 (10)
Reference to past activities	2 (22)	3 (30)

Cognitive Restructuring

Cognitive restructuring was included in 8 desktop programs and 13 apps. The explanation of the principle of automatic thoughts was rated to be more clear in desktop programs (mean 4.44, SD 1.15) than in apps (mean 2.92, SD 1.57). Beck's cognitive triad was addressed in desktop programs (mean 3.00, SD 1.41), but less often in apps (mean 1.50, SD 1.14).

Desktop programs and apps presented typical negative thoughts (8/8, 100% and 10/13, 77%, respectively), suggested alternative thoughts (8/8, 100% and 8/13, 62%, respectively), and suggested that the user write down individual negative thoughts (8/8, 100% and 13/13, 100%, respectively) as well as alternative thoughts (8/8, 100% and 10/13, 77%, respectively).

All cognitive restructuring interventions offered the option of keeping a daily thought protocol (21/21, 100%), which included the situation (19/21, 90%), feelings (17/21, 81%), negative thoughts (19/21, 90%), alternative positive thoughts (15/21, 71%), and the result of positive thoughts (16/21, 76%).

Cognitive distortions were addressed in both desktop programs and apps. An overview of the relative share of web-based interventions addressing cognitive distortions can be found in Table 6.

Table 6. Absolute numbers and relative share of web-based interventions addressing cognitive distortions for desktop programs (N=8) and apps (N=13).

Type of cognitive distortion	Desktop programs, n (%)	Apps, n (%)
Should statements	7 (88)	10 (77)
Focus on the negative	7 (88)	10 (77)
Disqualifying the positive	7 (88)	8 (62)
Personalization	3 (38)	11 (85)
Labeling and mislabeling	6 (75)	9 (69)
Catastrophizing	7 (88)	9 (69)
Magnification and minimization	6 (75)	8 (62)
Emotional reasoning	6 (75)	10 (77)
Jumping to conclusions	7 (88)	11 (85)
Polarized thinking	7 (88)	10 (77)
Overgeneralization	8 (100)	10 (77)

Psychoeducation

Psychoeducation was implemented in 10 desktop programs and 11 apps, which included explanations of depressive symptoms in both desktop programs (mean 3.90, SD 1.29) and apps (mean 3.35, SD 1.53). An explanatory model for the development of depression was provided in desktop programs (mean 3.95, SD 1.40) and apps (mean 3.48, SD 1.59). Low ratings were found for whether the user received support in the development of an individual explanatory model for desktop programs (mean 1.80, SD 1.24) and apps (mean 1.22, SD 0.52).

Both desktop programs and apps conveyed that depression is well treatable (mean 4.25, SD 1.12 and mean 3.30, SD 1.40, respectively) and that the user can overcome depression (mean 4.60, SD 0.75 and mean 3.30, SD 1.40, respectively).

Further evaluations regarding psychoeducation can be found in Table 7.

XSL•FO
RenderX

Table 7. Mean (SD) scores for evaluation of psychoeducation features in web-based interventions.

Item	Overall	Desktop programs	Apps
"The program appropriately conveys that there may be fluctuations in the course of treatment."	2.45 (1.35)	2.89 (1.49)	2.09 (1.13)
"The possible relationship between depression and anxiety is adequately conveyed."	2.16 (1.29)	2.45 (1.36)	1.91 (1.20)
"The role of avoidance behavior is explained clearly."	2.62 (1.58)	3.21 (1.40)	2.65 (1.58)
"The role of social isolation is explained clearly."	2.91 (1.44)	3.20 (1.24)	2.65 (1.58)
"The possibly increased perception of physical symptoms associated with depression is explained clearly.	1.84 (1.31)	1.75 (1.16)	1.91 (1.44)
"Possible somatic causes of physical symptoms are adequately addressed."	1.91 (1.49)	1.85 (1.63)	1.96 (1.40)

Mood Tracking

Mood tracking was implemented in 24 web-based interventions, including 11 desktop programs and 13 apps. When a web-based intervention reminded the user to report his or her mood on a regular basis, the reminders occurred less than every 6 hours in most cases (19/24, 79%). The queries of the mood were visualized in desktop programs (mean 3.41, SD 1.37) and apps (mean 3.85, SD 1.32), and took into account common diagnostic criteria for depression (desktop programs mean 3.45, SD 1.50; apps mean 2.50, SD 1.56). The mood queries fulfilled our criteria for scaling (ie, offering at least 5 options) in some desktop programs (8/11, 73%) and apps (11/13, 85%).

Low ratings were given for the explanation of why it can be important to track small periods of time (to detect triggers for specific moods; overall mean 1.27, SD 0.79) or large periods of time (to detect trends in the mood progression; overall mean 1.08, SD 0.35). When mood progression was visualized (eg, as a curve), small periods of time (eg, the last few hours) could be chosen in 10 web-based interventions, including 3/11 (27%) of desktop programs and 7/13 (54%) of apps, and longer periods of time (eg, more than 1 week) could be chosen in 12 web-based interventions, including 5/11 (45%) desktop programs and 7/13 (54%) apps. When mood improvements were given, 4 web-based interventions highlighted improvements (3/11, 27% of desktop programs and 1/13, 8% of apps) and 8 web-based interventions suggested possible relations between mood and current events (2/11, 18% of desktop programs and 6/13, 46% of apps).

Overall, 11 web-based interventions asked the user about anxiety symptoms (6/11, 55% of desktop programs and 5/13, 38% of apps) and 15 web-based interventions asked the user about physical ailments (8/11, 73% of desktop programs and 7/13, 54% of apps). When physical ailments were assessed, 7 web-based interventions advised the user to consult a physician (5/11, 45% of desktop programs and 2/13, 15% of apps).

Journal Keeping

Interventions that provided journal keeping were included in 4 desktop programs and 8 apps. Overall, the web-based interventions received low ratings for their explanation of which components the journal entries might contain (mean 1.96, SD 1.43) or for their explanation of how it can be helpful to note the positive aspects of the day (mean 2.38, SD 1.70). In most cases, journal entries could be entered through blank text boxes, although some web-based interventions suggested preset phrasings for particular aspects of the journal (7/12, 58%). Some web-based interventions took up journal entries in the progression of the program (2/12, 17%) or provided feedback on the content of journal entries (2/12, 17%).

Relaxation Training

Overall, relaxation trainings were realized in 8 desktop programs and 10 apps. In both types of web-based interventions, mindfulness was suggested as a relaxation technique and the concept of mindfulness was explained in desktop programs (mean 3.00, SD 1.63) and apps (mean 2.85, SD 1.63). Other relaxation techniques that were suggested in the web-based interventions are summarized in Table 8.

When mindfulness was discussed in an intervention, 12 web-based interventions suggested to accept perceptions without judging them (5/8, 63% of desktop programs and 7/10, 70% of apps). Overall, 12 web-based interventions suggested to distance oneself from thoughts (5/8, 63% of desktop programs and 7/10, 70% of apps) and 7 web-based interventions suggested to perform mindfulness exercises as part of a daily routine (2/8, 25% of desktop programs and 5/10, 50% of apps). Additionally, 14 web-based interventions offered audio- or video-based mindfulness exercises (6/8, 75% of desktop programs and 8/10, 80% of apps).

The explanation of typical stressors received low ratings in desktop programs (mean 2.00, SD 1.14) and apps (mean 1.45, SD 0.76). One desktop program and one app offered the user to add personal stressors. The explanation of possible risks in performing mindfulness exercises (eg, the occurrence of unpleasant feelings) received low ratings for desktop programs (mean 1.81, SD 1.17) and apps (mean 1.30, SD 0.57). Similar results were found for the explanation of possible difficulties in performing mindfulness exercises (eg, boredom or falling asleep) in desktop programs (mean 2.00, SD 1.10) and apps (mean 1.80, SD 1.15).

XSL•FO
RenderX

Table 8. Absolute numbers and relative share of relaxation techniques suggested in web-based interventions for desktop programs (N=8) and apps (N=10).

Relaxation technique	Desktop programs, n (%)	Apps, n (%)
Mindfulness	8 (100)	8 (80)
Progressive muscle relaxtion	7 (88)	5 (50)
Meditation	4 (50)	8 (80)
Guided imagery journeys	1 (13)	3 (30)
Imagination exercises	2 (25)	5 (50)

Social Skills Training

Social skills training was included in 6 desktop programs and 2 apps. When social skills training was included, users were encouraged to perform a change of perspective in desktop programs (mean 4.25, SD 1.22) and apps (mean 4.50, SD 0.58).

When nonverbal or verbal components of social interaction were discussed (eg, maintaining eye contact or paying attention to voice modulation), desktop programs received higher ratings for their explanations (mean 3.08, SD 1.38 for nonverbal components and mean 3.42, SD 1.44 for verbal components) than apps (mean 1.50, SD 0.58 and mean 1.25, SD 0.50, respectively).

Six desktop programs and one app encouraged the user to perform exercises of social interaction (eg, paying somebody a compliment). The perception of social cues was rated higher in desktop programs (mean 2.92, SD 1.38) than in apps (mean 1.75, SD 0.96).

Both desktop programs and apps addressed the establishment and maintenance of social contacts (mean 3.17, SD 1.33 and mean 3.00, SD 1.42, respectively) and the assertion of one's own wishes in social situations (mean 3.50, SD 1.62 and mean 3.25, SD 1.71, respectively).

Crisis Management

Crisis management was included in 9 desktop programs and 11 apps. Nine desktop programs and 9 apps provided an emergency contact to the user (eg, phone numbers or contact details that are accessible at any time; 9/9, 100% and 9/11, 82%, respectively). One desktop program and one app required the user to provide an emergency contact in order to use the program (1/9, 11% and 1/11, 9%, respectively). To prevent relapses, 3/9 (33%) desktop programs and 3/11 (27%) apps suggested creating an individual list of warning signs. Contact to support groups was offered by 2/9 (22%) desktop programs and 3/11 (27%) apps.

When the mood in the mood tracking dropped immensely, 7/9 (78%) desktop programs and 4/11 (36%) apps suggested contacting the emergency contact. Three desktop programs and 5 apps suggested creating an emergency plan with individual measures (3/9, 33% and 5/11, 45%, respectively). When the mood dropped, 1/9 (11%) desktop program and 1/11 (9%) app suggested resorting to the emergency plan.

Resource Activation

Although no web-based intervention evaluated in our study implemented a specific module for resource activation, this aspect was indirectly included within several other treatment components. A summary of evaluations regarding resource activation can be found in Table 9, including encouragement to identify individual resources, assistance in identifying resources, encouragement to reflect resources in their context, and uptake of resources in the progression of the web-based intervention. Among all web-based interventions included in this study, 3 desktop programs and 3 apps took up identified resources as the intervention progressed.

Table 9. Mean (SD) scores in the evaluation of resource activation features of web-based interventions.

Item	Overall	Desktop programs	Apps
"The user is encouraged to identify his/her own resources."	2.76 (1.56)	3.45 (1.36)	2.30 (1.54)
"The program provides adequate assistance in identifying individual resources (eg, thinking about past successes)."	2.32 (1.60)	3.20 (1.51)	1.73 (1.39)
"The user is encouraged to reflect resources in their context (eg, origin story, typical situations, promoting or inhibiting factors)."	2.02 (1.39)	2.75 (1.41)	1.53 (1.17)

Discussion

Principal Results

The aim of this study was to evaluate the quality of unguided web-based interventions for depression at the level of individual treatment components based on their adherence to established treatment guidelines. We tested 28 web-based interventions with a self-developed, standardized list of criteria, which were based on gold-standard treatment guidelines and manuals for depression. Despite the high number of programs and apps for depression available through the commercial market that claim to follow a scientifically sound methodology, we found varying degrees of adherence to established treatment guidelines and manuals.

All web-based interventions included some of the interventions we expected in a CBT-based treatment, indicating conformance

XSL•FO
RenderX

to treatment guidelines to some extent. As a result, our criteria were applicable, and we could gather insights about how the interventions were digitally realized. Overall, there is an evident trend for desktop programs to be rated as more adherent to treatment manuals than apps, which is in line with previous research. In a systematic review, Huguet et al [55] found that only 10% of depression apps seem to be consistent with evidence-based methods of CBT. However, their study focused on behavioral activation, which was only one part of our approach. In our study, cognitive restructuring, psychoeducation, and mood tracking were the most frequently realized components, which were included in more than 70% of the web-based interventions that we tested.

Our key findings reveal substantial variation regarding the therapeutic utility of these interventions, and interventions were diversely realized. For example, in treatment components targeting cognitive restructuring, the explanation of the principle of automatic thoughts was rated high in desktop programs, yet had a lower rating in apps. In contrast, Beck's cognitive triad was addressed less often, although it can provide therapeutic background as an explanatory model for depression. Nevertheless, all cognitive restructuring interventions offered the option to keep a protocol of negative thoughts, as recommended in the manuals, and most of them included details such as the situation, feelings, negative thoughts, alternative positive thoughts, and result of positive thoughts.

When considering the use of mood tracking, we found that 79% of the web-based interventions that implemented mood tracking in our study inquired the mood only once every 6 hours or less frequently. Since retrospective recall of mood in people with depression is biased toward the negative [72,73], web-based interventions could be improved by adding more frequent mood inquiries. Additionally, frequent mood tracking throughout the day can help patients identify triggers for negative moods. However, it may be difficult for patients to track their mood frequently throughout the day on a desktop program. This could be complemented by using worksheets or combining desktop programs with app features. In particular, apps that have the opportunity to make ideal use of ecological momentary assessment could tap into their full potential by suggesting multiple mood inquiries throughout the day [74].

Out of all of the web-based interventions that we tested, 61% did not provide sufficient risk and crisis management when the mood dropped immensely. The most common reaction to a sudden mood drop was the recommendation to call emergency contacts or a crisis line. Yet, some web-based interventions did not react at all to severe mood drops. Although all of the web-based interventions self-identified as not suitable for suicidal patients, strong mood changes and suicidality are common symptoms in depression [75] that should not be completely excluded.

Another aspect that we noted is that the majority of web-based interventions asked the user about physical ailments, but not about anxiety symptoms. Web-based interventions should pay attention to and educate the user about anxiety symptoms and somatic diseases related to depression because there are high rates of comorbidity among depression, anxiety, and physical illness [76-78]. It would be beneficial if a web-based intervention asked the user about anxiety symptoms and provided information about treatment opportunities for anxiety so that users can reach out to professional help when they need it. Additionally, web-based interventions should recommend that users receive a proper somatic examination when somatic symptoms are present because some symptoms of depression can also be caused by physical illnesses (eg, lack of energy, difficulty in concentrating).

Some web-based interventions did not mention possible difficulties regarding the interventions they suggested. This can result in feelings of failure or disappointment when the user does not manage to complete the intervention. As an example, we found that 19 web-based interventions included behavioral activation and suggested the uptake of specific activities, but the raters evaluated that complex activities were not sufficiently divided into achievable intermediate steps, difficulties for the realization of suggested activities were not sufficiently addressed, and difficulty levels for the suggested activities were not sufficiently provided. Additionally, only 2 out of these 19 web-based interventions asked if an activity was completed. Altogether, this may result in the user feeling invalidated.

As some components of web-based interventions can be text-intensive and do not achieve proper risk management, ongoing engagement and motivation may be required for the user to follow the intervention, which may be very challenging for a depressed individual. Previous research found dissatisfaction and a lack of motivation to be a possible explanation when no benefits of a web-based intervention could be found [79]. Additionally, lack of a user-centric design and concerns about unhelpfulness in emergencies were identified as reasons for low user engagement in smartphone-based mental health apps [45].

Taken all together, we suggest that lack of adherence to treatment guidelines might be a reason why some users find no sufficient benefit in web-based interventions and discontinue the use after a short time, a phenomenon that was identified in previous research [43,44]. The quality of reviewed interventions is variable and the lack of risk management and appropriate adherence to treatment guidelines make it questionable as to whether these products can be recommended to patients suffering from depression without professional oversight. However, many developers are aware of this issue and most web-based interventions do not claim to be a sole therapy tool, but rather a supplement to professional treatment. One reason for this is that high-quality web-based interventions require a wide range of psychotherapeutic, financial, legal, and technological support that cannot always be provided by the developers.

Given the rising demand for web-based treatment options, further research should aim to find principles for the development and evaluation of therapeutically sound web-based interventions for depression, to make is easier for both clinicians and patients to find a suitable product.

XSL•FO
RenderX

Limitations

This is the first study that investigated the quality of unguided web-based interventions for depression at the level of individual treatment components based on established treatment guidelines. However, some limitations should be regarded when interpreting our findings.

Although we used a standardized questionnaire and trained psychologists to achieve systematic and thorough evaluations, the questionnaire in this study has not yet been externally validated and all ratings are subjective to a certain degree. Further, some beneficial interventions might not be included in this questionnaire. For example, some web-based interventions offered interventions based on problem-solving therapy or systemic therapy, which we did not examine in this study. We consider this questionnaire to be an important first step, but there is room for improvement in the future. A future direction is the external verification of this questionnaire. Additionally, future research should aim to examine a wider range of interventions beyond those based on CBT. As we only used three psychologists to rate the web-based intervention in this study, future research should aim to include more raters with different areas of expertise.

Another limitation is that our raters had no depression diagnosis. Therefore, our raters were not the selected target group of the examined interventions and no conclusion on the effectiveness of web-based interventions for patients can be made. Although the criteria used in this study can be an indicator of high-quality interventions, they do not necessarily have a causal relation to symptom reduction or quality of life improvement, and they cannot be compared to randomized controlled trials involving patients. However, since the aim of this study was not to evaluate effectiveness or symptom reduction, but rather to evaluate the quality of interventions based on their adherence to psychotherapeutic guidelines, we considered the use of expert ratings to be suitable for the purpose of our study. Nevertheless, future research could compare the ratings of patients with the ratings of experts to determine if end users hold different opinions toward the quality of an intervention.

A further limitation of this study is that web-based interventions that required the user to get in touch with a counselor or psychotherapist were not included. Therefore, we can make no claim about how elaborated these interventions are. Additionally, we only evaluated interventions that were available in English or German, as these were the only languages that are fluently spoken by the researchers.

Finally, as we searched for web-based interventions in January 2018 and all our evaluations were performed in 2018, some interventions might no longer be available or may have changed through updates in the meantime.

As a side note, we did not evaluate web-based interventions that used sensor data (eg, through a smartwatch), because we could only identify one such app, which was not available in Germany. As the usage of sensor data is rising in research and practice and promises many advantages to patients, practitioners, and researchers, future research should aim to develop and evaluate web-based interventions that include sensor data.

Conclusion

Many unguided web-based interventions for depression claim to deliver therapeutic content of high quality, but there is high variability in their adherence to established treatment guidelines and manuals. Although many interventions from manuals are included in web-based interventions and developers offer a wide range of treatment components, the necessary level of therapeutic depth is seldom reached. Mental health professionals and developers should work together to implement current treatment guidelines in their interventions to close this gap, especially concerning the lack of risk management that we identified in our study. This could improve user experience and prevent adverse side effects such as users feeling overwhelmed, frustrated, or leaving the intervention altogether. When developers implement quality criteria of established treatment guidelines in their products, web-based interventions can be a valuable tool to supplement professional treatment. The use of web-based interventions in the treatment of depression enables patients and professionals to gather important information from the patients' everyday lives and to save financial and time resources because treatment components can be completed at home. There is a need for further research targeting general principles for the development and evaluation of therapeutically sound web-based interventions for depression, which includes treatment interventions beyond CBT.

Acknowledgments

This study was part of the project Self-administered PsychoTherapy SystemS (SELFPASS) funded by the Federal Ministry of Education and Research (FKZ 13GW0157B). We thank Dr. Simone Alvarez for providing linguistic support in translating the questionnaire for this publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English version of the questionnaire used in this study.
[DOCX File , 1895 KB - jmir_v22i7e16136_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Web-based interventions tested in this study.
[PDF File (Adobe PDF File), 195 KB - jmir_v22i7e16136_app2.pdf]

Multimedia Appendix 3
Treatment components that could explicitly be identified for each web-based intervention.
[PDF File (Adobe PDF File), 207 KB - jmir_v22i7e16136_app3.pdf]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1 [accessed 2019-05-15]
2. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
3. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861. [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
4. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. JAMA Psychiatry 2018 Apr 01;75(4):336-346 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
5. Nierenberg AA, Rapaport MH, Schettler PJ, Howland RH, Smith JA, Edwards D, et al. Deficits in psychological well-being and quality-of-life in minor depression: implications for DSM-V. CNS Neurosci Ther 2010 Aug;16(4):208-216 [FREE Full text] [doi: 10.1111/j.1755-5949.2009.00108.x] [Medline: 20236139]
6. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. J Affect Disord 2002 Dec;72(3):227-236. [doi: 10.1016/s0165-0327(01)00413-x] [Medline: 12450639]
7. Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritzer K. Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. Arch Gen Psychiatry 1995 Jan;52(1):11-19. [doi: 10.1001/archpsyc.1995.03950130011002] [Medline: 7811158]
8. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
9. Cuijpers P, Andersson G, Donker T, van Straten A. Psychological treatment of depression: results of a series of meta-analyses. Nord J Psychiatry 2011 Dec;65(6):354-364. [doi: 10.3109/08039488.2011.596570] [Medline: 21770842]
10. Karyotaki E, Smit Y, de Beurs DP, Henningsen KH, Robays J, Huibers MJH, et al. The long-term efficacy of acute-phase psychotherapy for depression: a meta-analysis of randomized trials. Depress Anxiety 2016 May;33(5):370-383. [doi: 10.1002/da.22491] [Medline: 27000501]
11. Karyotaki E, Smit Y, Holdt Henningsen K, Huibers MJH, Robays J, de Beurs D, et al. Combining pharmacotherapy and psychotherapy or monotherapy for major depression? A meta-analysis on the long-term effects. J Affect Disord 2016 Apr;194:144-152. [doi: 10.1016/j.jad.2016.01.036] [Medline: 26826534]
12. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5):e1001454 [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
13. Kazdin AE. Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
14. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
15. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
16. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
17. Barney LJ, Griffiths KM, Christensen H, Jorm AF. Exploring the nature of stigmatising beliefs about depression and help-seeking: implications for reducing stigma. BMC Public Health 2009 Feb 20;9:61 [FREE Full text] [doi: 10.1186/1471-2458-9-61] [Medline: 19228435]
18. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]

XSL•FO
RenderX

19. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring Attitudes towards Psychological Online Interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
20. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
21. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11. [doi: 10.1002/wps.20083] [Medline: 24497236]
22. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
23. Klein JP, Gerlinger G, Knaevelsrud C, Bohus M, Meisenzahl E, Kersting A, et al. Internetbasierte Interventionen in der Behandlung psychischer Störungen [Internet-based interventions in the treatment of mental disorders : Overview, quality criteria, perspectives]. Nervenarzt 2016 Nov;87(11):1185-1193. [doi: 10.1007/s00115-016-0217-7] [Medline: 27649987]
24. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
25. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
26. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
27. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
28. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
29. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety - a systematic review. Health Psychol Behav Med 2014 Jan 01;2(1):841-881 [FREE Full text] [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
30. Griffiths KM, Carron-Arthur B, Parsons A, Reid R. Effectiveness of programs for reducing the stigma associated with mental disorders. A meta-analysis of randomized controlled trials. World Psychiatry 2014 Jun;13(2):161-175. [doi: 10.1002/wps.20129] [Medline: 24890069]
31. Kiropoulos LA, Griffiths KM, Blashki G. Effects of a multilingual information website intervention on the levels of depression literacy and depression-related stigma in Greek-born and Italian-born immigrants living in Australia: a randomized controlled trial. J Med Internet Res 2011 Apr 19;13(2):e34 [FREE Full text] [doi: 10.2196/jmir.1527] [Medline: 21504872]
32. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185:342-349. [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
33. King R, Bickman L, Shochet I, McDermott B, Bor B. Use of the internet for provision of better counselling and psychotherapy services to young people, their families and carers. Psychotherapy in Australia 2010;17(1):66-74.
34. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and Interest in Health Apps Among Those With Mental Health Needs: Survey and Focus Group Study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
35. The growing value of digital health. IQVIA Institute for Human Data Science. 2017. URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-07-03]
36. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
37. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
38. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
39. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
40. Martinez-Martin N, Kreitmair K. Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]

XSL•FO
RenderX

41. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Jul 02;9:11 [FREE Full text] [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
42. Stevens WJM, van der Sande R, Beijer LJ, Gerritsen MG, Assendelft WJ. eHealth Apps Replacing or Complementing Health Care Contacts: Scoping Review on Adverse Effects. J Med Internet Res 2019 Mar 01;21(3):e10736 [FREE Full text] [doi: 10.2196/10736] [Medline: 30821690]
43. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
44. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. Psychol Med 2014 Mar;44(4):741-752 [FREE Full text] [doi: 10.1017/S0033291713001323] [Medline: 23795621]
45. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
46. BDP verabschiedet Gütesiegel für psychologische Gesundheitsangebote im Internet. Berufsverband deutscher Psychologinnen und Psychologen (BDP). 2017. URL: https://www.bdp-verband.de/binaries/content/assets/verband/aktuelles/2017/gutesiegel-des-bdp-fur-psychologische-angebote-im-internet.pdf [accessed 2019-07-03]
47. Mental Health Apps. Anxiety and Depression Association of America (ADAA). 2019. URL: https://adaa.org/mental-health-apps [accessed 2019-06-18]
48. Renton T, Tang H, Ennis N, Cusimano MD, Bhalerao S, Schweizer TA, et al. Web-based intervention programs for depression: a scoping review and evaluation. J Med Internet Res 2014 Sep 23;16(9):e209 [FREE Full text] [doi: 10.2196/jmir.3147] [Medline: 25249003]
49. Burger F, Neerincx MA, Brinkman W. Technological State of the Art of Electronic Mental Health Interventions for Major Depressive Disorder: Systematic Literature Review. J Med Internet Res 2020 Jan 20;22(1):e12599 [FREE Full text] [doi: 10.2196/12599] [Medline: 31958063]
50. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
51. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
52. Martín Payo R, Harris J, Armes J. Prescribing fitness apps for people with cancer: a preliminary assessment of content and quality of commercially available apps. J Cancer Surviv 2019 Jun;13(3):397-405. [doi: 10.1007/s11764-019-00760-2] [Medline: 31030308]
53. Terhorst Y, Rathner EM, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen [Help from the app store? A systematic overview and evaluation of apps for depression]. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
54. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
55. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
56. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of Top-Rated Mobile Apps for Depression: Systematic Search and Evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]
57. Online applications for mental and physical disorders (Beacon 2.0). Centre for Mental Health Research at the Australian National University. 2018. URL: https://beacon.anu.edu.au/ [accessed 2018-01-10]
58. Digital Health Knowledge Database. HealthOn e.V. 2018. URL: https://www.healthon.de/ [accessed 2018-01-11]
59. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression - Langfassung, 2. Auflage. Version 5. DGPPN, BÄK, KBV, AWMF für die Leitliniengruppe Unipolare Depression. 2015. URL: https://www.leitlinien.de/nvl/depression [accessed 2018-01-10]
60. Härter M, Klesse C, Bermejo I, Bschor T, Gensichen J, Harfst T, et al. [Evidence-based therapy of depression: S3 guidelines on unipolar depression]. Nervenarzt 2010 Sep;81(9):1049-1068. [doi: 10.1007/s00115-010-3084-7] [Medline: 20802992]
61. Beck JS. Cognitive behavior therapy: Basics and beyond. 2nd edition. New York, USA: Guilford Press; 2011.
62. Cuijpers P, Muñoz RF, Clarke GN, Lewinsohn PM. Psychoeducational treatment and prevention of depression: the "Coping with Depression" course thirty years later. Clin Psychol Rev 2009 Jul;29(5):449-458. [doi: 10.1016/j.cpr.2009.04.005] [Medline: 19450912]

XSL•FO
RenderX

63. Beck AT, Steer RA, Brown GK. Beck Depression Inventory-II (BDI-II): Manual. San Antonio, TX: Psychological Corporation; 1996.
64. Gwet KL. Handbook of inter-rater reliability. 2nd ed. Gaithersburg, MD: Advanced Analytics, LLC; 2010.
65. Abramson JH. WINPEPI updated: computer programs for epidemiologists, and their teaching potential. Epidemiol Perspect Innov 2011 Feb 02;8(1):1 [FREE Full text] [doi: 10.1186/1742-5573-8-1] [Medline: 21288353]
66. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008 May;61(Pt 1):29-48. [doi: 10.1348/000711006X126600] [Medline: 18482474]
67. Gwet K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. Statistical methods for inter-rater reliability assessment; 1(6). 2002. URL: http://agreestat.com/papers/papers/kappa_statistic_is_not_satisfactory.pdf [accessed 2019-06-18]
68. Gwet K. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. Statistical methods for inter-rater reliability assessment; 2(1). 2002. URL: http://agreestat.com/papers/papers/inter_rater_reliability_dependency.pdf [accessed 2019-06-18]
69. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
70. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977 Jun;33(2):363-374. [Medline: 884196]
71. Gwet KL. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. 4th ed. Gaithersburg, MD: Advanced Analytics, LLC; 2014.
72. Wenze SJ, Gunthert KC, German RE. Biases in affective forecasting and recall in individuals with depression and anxiety symptoms. Pers Soc Psychol Bull 2012 Jul;38(7):895-906. [doi: 10.1177/0146167212447242] [Medline: 22649114]
73. Urban EJ, Charles ST, Levine LJ, Almeida DM. Depression history and memory bias for specific daily emotions. PLoS One 2018;13(9):e0203574 [FREE Full text] [doi: 10.1371/journal.pone.0203574] [Medline: 30192853]
74. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
75. World Health Organization. International statistical classification of diseases and related health problems, ICD-10. Geneva: World Health Organization; 2012:1-2118.
76. Kang HJ, Kim SY, Bae KY, Kim SW, Shin IS, Yoon JS, et al. Comorbidity of depression with physical disorders: research and clinical implications. Chonnam Med J 2015 Apr;51(1):8-18 [FREE Full text] [doi: 10.4068/cmj.2015.51.1.8] [Medline: 25914875]
77. Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJLM, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). J Clin Psychiatry 2011 Mar;72(3):341-348. [doi: 10.4088/JCP.10m06176blu] [Medline: 21294994]
78. Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am J Psychiatry 2008 Mar;165(3):342-351. [doi: 10.1176/appi.ajp.2007.06111868] [Medline: 18172020]
79. Hoek W, Schuurmans J, Koot HM, Cuijpers P. Effects of Internet-based guided self-help problem-solving therapy for adolescents with depression and anxiety: a randomized controlled trial. PLoS One 2012;7(8):e43485 [FREE Full text] [doi: 10.1371/journal.pone.0043485] [Medline: 22952691]

Abbreviations

CBT: cognitive behavioral therapy
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale

Edited by G Eysenbach; submitted 04.09.19; peer-reviewed by CS Hung, M Subotic-Kerry, JJ Ding, V Strotbaum, F Lai; comments to author 06.02.20; revised version received 01.04.20; accepted 19.04.20; published 13.07.20.

Please cite as:
Bubolz S, Mayer G, Gronewold N, Hilbel T, Schultz JH
Adherence to Established Treatment Guidelines Among Unguided Digital Interventions for Depression: Quality Evaluation of 28 Web-Based Programs and Mobile Apps
J Med Internet Res 2020;22(7):e16136
URL: https://www.jmir.org/2020/7/e16136
doi:10.2196/16136
PMID:32673221

XSL•FO
RenderX

©Stefan Bubolz, Gwendolyn Mayer, Nadine Gronewold, Thomas Hilbel, Jobst-Hendrik Schultz. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of the e-Motivate4Change Program on Metabolic Syndrome in Young Adults Using Health Apps and Wearable Devices: Quasi-Experimental Study

Ji-Soo Lee[1], MSN, RN; Min-Ah Kang[1,2], MSN, RN; Soo-Kyoung Lee[1], PhD, RN

[1]Keimyung University, Daegu, Republic of Korea

[2]Gyeongnam Center for Infectious Disease Control and Prevention, Changwon, Republic of Korea

Corresponding Author:
Soo-Kyoung Lee, PhD, RN
Keimyung University
1095 Dalgubeol-daero, Dalseo-Gu
Daegu, 42601
Republic of Korea
Phone: 82 53 258 7665
Fax: 82 3 258 7616
Email: soo1005s@gmail.com

Abstract

Background: The health behaviors of young adults lag behind those of other age groups, and active health management is needed to improve health behaviors and prevent chronic diseases. In addition, developing good lifestyle habits earlier in life could reduce the risk of metabolic syndrome (MetS) later on.

Objective: The aim of this study is to investigate the effects of the e-Motivate4Change program, for which health apps and wearable devices were selected based on user needs. The program was developed for the prevention and management of MetS in young adults.

Methods: This experimental study used a nonequivalent control group. In total, 59 students from 2 universities in Daegu, Korea participated in the study (experimental group n=30; control group n=29). Data were collected over 4 months, from June 1 to September 30, 2018. The experimental group received a 12-week e-Motivate4Change program intervention, and the control group received MetS education and booklets without the e-Motivate4Change program intervention.

Results: After the program, the experimental group had significantly higher scores for health-related lifestyle (t=3.86; P<.001) and self-efficacy (t=6.00; P<.001) than did the control group. Concerning BMI, there were significant effects by group (F=1.01; P<.001) and for the group × time interaction (F=4.71; P=.034). Concerning cholesterol, there were significant main effects for group (F=4.32; P=.042) and time (F=9.73; P<.001).

Conclusions: The e-Motivate4Change program effectively improved participants' health-related lifestyle scores and self-efficacy, and significantly reduced their BMI and cholesterol levels. The program can be used to identify and prevent MetS among young adults.

(J Med Internet Res 2020;22(7):e17031) doi:10.2196/17031

KEYWORDS

metabolic syndrome; telemedicine; mobile apps; preventive care; wearable electronic devices

Introduction

Background

Chronic disease is a significant factor of disability as well as a major cause of death for nearly 60% of the global population [1]. Metabolic syndrome (MetS) is a serious risk factor for heart disease, stroke, and type 2 diabetes. There are numerous indicators of MetS, including hypertriglyceridemia, high blood sugar, and abdominal obesity; the incidence rate of MetS is steadily increasing [2]. Concurrently, MetS significantly lowers individuals' quality of life (QOL) [3-6], which is defined as one's level of acceptance in relation to the ideals, interests, expectations, and norms of one's culture and value system [7]. Health-related QOL, in particular, is a personal evaluation of the physical and psychological impact of a disorder [8].

XSL•FO
RenderX

Although patients use this indicator to demonstrate their satisfaction level with their own functionality, scholars use it to analyze the average health level and the impact of health care among a wide range of populations [6].

The World Health Organization (WHO) reported that the incidence rate of chronic disease is not only related to social and economic factors but also to individuals' eating habits and physical activities such as drinking and smoking [9]. Sedentary lifestyle, high-calorie diet, sweetened drinks, low economic status, and saturated fats are key risk factors that increase the risk of obesity; however, regular physical activity is known to lower individuals' risk [10,11]. People who are habitually active, less sedentary, and have better cardio-pulmonary functions are less obese and thus have lower incidence rates of MetS [12]. Further, one exemplary study reported that better knowledge of one's own disease increases medical accessibility and has a positive impact on the treatment process and self-care, thus improving individuals' lifestyles [13].

The incidence rate of MetS is notably high among children and adolescents in the United States [14]. Further, the morbidity rate is associated with the obesity rate among young adults, which is also highly related to numerous adult diseases including arteriosclerosis, cardiovascular disease, cancer, and diabetes. Thus, appropriate prevention and intervention is needed in all related fields [14]. One can significantly lower the risk of and prevent MetS by controlling diet habits, physical activity, smoking, drinking, and stress levels [10,15,16]. In particular, men aged 30 to 39 years and women aged 20 to 29 years have worse health activity compared to other age categories, and thus require more active and focused health management to improve their health habits and prevent chronic diseases [17]. Since most health habits become more solidified in mid to older age, it is important to develop a MetS prevention program and prepare the foundation for appropriate health habits among young adults.

Several studies have analyzed the relationship between the incidence rate of MetS and its risk factors such as smoking, drinking, physical activity, depression, and environmental and demographic factors [18]. Moreover, one study revealed that an unhealthy daily lifestyle increases the risk of cardiocerebrovascular diseases and that improved habits act as preventive measures [19]. Further, the Third Report of the National Cholesterol Education Program Expert Panel in Europe reported that it is important to identify patients with MetS and treat them to prevent cardiocerebrovascular diseases [20].

Previous studies on MetS intervention programs examined ubiquitous health (uHealth) nutrition education that focused on increasing fruit and vegetable intake [21] and community-based health education, both of which appeared effective [22]. The uHealth nutrition education initiative is an online self-monitoring program, and data are collected during each session, and are then analyzed cumulatively [21]. A previous cross-sectional study in rural Taiwan successfully reduced the mortality rate from MetS by applying an appropriate Health City plan [23]. The Health City plan for Miaoli County, Taiwan integrated public and civil departments, nongovernmental organizations, and community resources to develop a model of "must move" and "healthy diet" as part of a sustainable development strategy. Thus, an appropriate health promotion project may effectively reduce the threat of death by MetS [23]. However, most previous studies on MetS and its interventions focused on adults aged >30 years; consequently, more studies about young adults are needed.

With the rapid development of information and smart technology, the demand for mobile health (mHealth), which uses mobile technology such as health apps to prevent disease and improve health, has increased [24,25]. Currently, there are numerous apps that support users' health management. However, users report an unsatisfactory experience when they download the desired app and do not find the content of the app useful or relevant, making it difficult to use the app in the long run [26]. According to a user experience survey, 25.3% of users delete an app once they stop using it [27].

Wearable devices are a key subfield of mHealth. Wearable devices can constantly and easily monitor one's physical activity, and they can provide tailored feedback through synchronized apps [28]. According to previous international studies, wearable devices can improve the lifestyle of patients with chronic diseases [24,26,28]. Numerous companies including Fitbit, Xiaomi, and Samsung produce wearable devices, and the market for devices that primarily manage health is steadily increasing [29]. According to Wijsman and colleagues [30], health interventions using wearable devices increase the daily physical activities of older adults and improve their metabolism.

However, there are some problems with mHealth. First, there is a lack of entertaining aspects that can motivate sustained use. Second, the system that directly delivers the information and professional health care guidelines is incomplete [27]. Other reported problems include a lack of visual information, inconvenient data entry, and one-sided feedback [26]. Therefore, to motivate users and foster active participation, it is important to increase users' self-efficacy, which provides significant intrinsic motivation [31].

Self-efficacy is closely related to how well a person can control and perform specific activities under certain situations. Based on previous studies, health-related self-efficacy is a key motivator for one to continue physical activity; [32] thus, it serves as an important determinant of healthy lifestyle promotion [33]. Previous studies also reported a significant correlation between young adults' health activity and self-efficacy [34,35].

Therefore, this study aims to improve on existing intervention programs by actively using wearable devices and mobile apps and increasing entertaining factors and motivation. To support sustained health activity, we developed a tailored program based on the specific needs of young adults to increase their self-efficacy and general health-related QOL.

Objectives

The specific goal of our study is to develop a program (e-Motivate4Change) using health apps and wearable devices selected based on user's needs, and analyze its impact on MetS prevention among young people. In addition, we sought to determine the program's effect on participants' health-related lifestyle, self-efficacy, and QOL.

XSL•FO
RenderX

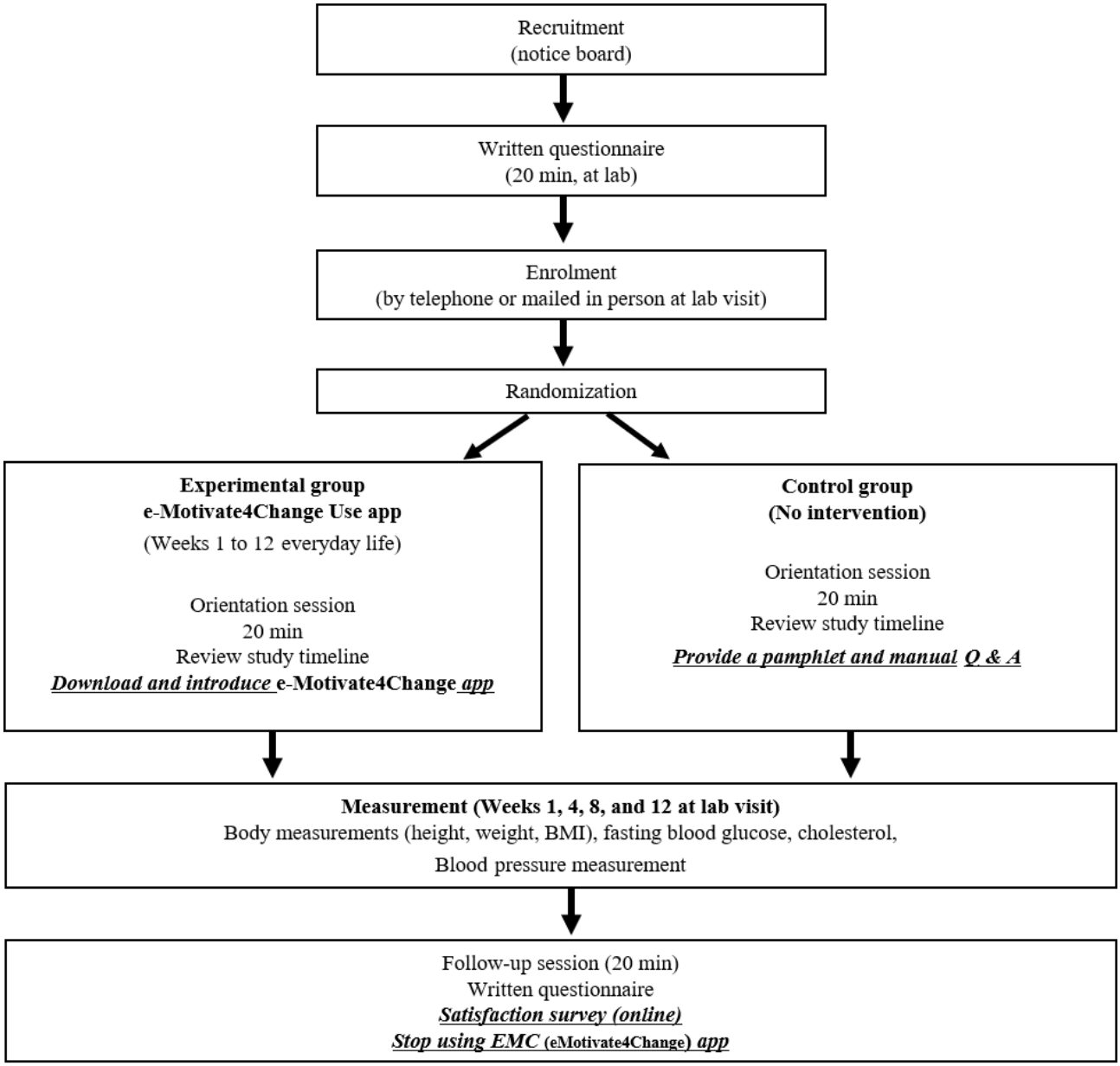
Methods

Study Design

This experimental study used a nonequivalent control group and pre- and post-test design. The experimental group completed a presurvey, the e-Motivate4Change program intervention, and three postsurveys. The control group received a pamphlet on MetS and completed three postsurveys. Data were collected over 4 months, from June 1 to September 30, 2018.

The questionnaire consisted of general characteristics, healthy lifestyle promotion, self-efficacy, and QOL. To collect physiological indicators, we measured participants' BMI, blood pressure, blood sugar, and cholesterol level. The general flow of the study is shown in Figure 1.

Figure 1. Flowchart of participants through each stage of the study.



Participants

Nursing students from Daegu University and Keimyung University who volunteered to participate and met the criteria were recruited. G*power (Version 3.1.7) was used to calculate the appropriate sample size with an effect size of 0.50, a significance level of .05, and a qualification level of 80%. The program indicated that at least 27 participants were required.

Anticipating a dropout rate of 20%, we recruited 60 participants (30 for each group) [36]. During the study period, one student from the control group opted out owing to a leave of absence; therefore, the data from 59 students were analyzed.

Research Procedure

Participants were allocated to either the experimental or control group and the questionnaires were distributed. To measure healthy lifestyle promotion, we used Kang's modified version [31] of Walker's health-promoting lifestyle profile [37]. We used a self-efficacy measurement tool developed by Sherer and colleagues [38] to measure self-efficacy. We also used the translated version of the WHO's QOL instrument (short version;

XSL•FO
RenderX

WHOQOL-BREF) to measure QOL. It took approximately 10 minutes to complete the questionnaire.

After completion, participants' height, weight, body composition, blood pressure, blood sugar level, and cholesterol level were measured (Figure 2). BMI was measured using a body composition analyzer (InBody 3.0; Biospace), which uses bioelectrical impedance analysis. According to the categorization of Asian adults' BMI, as reported by the Korean Society for the Study of Obesity, participants with a BMI from 18.5 to 24.9 kg/m^2 are "normal," while a BMI from 25.0 to 29.9 kg/m^2 indicates "slightly overweight," and a BMI over 30.0 kg/m^2 indicates "obese" [39].

Figure 2. Physical and physiological assessment tools.



(a) Body composition measurement (bioelectrical impedance analysis) - height, weight, body mass index



(b) Blood pressure measurement - systolic/diastolic blood pressure



(c) Blood sugar and cholesterol Measurement - fasting blood sugar and cholesterol

Participants were not provided with any written results of their body measurements since participants could freely check their status through the program. Additionally, the case report form, which is the electronic document of individuals' information as stated in the research design, provided the necessary data to the team. All participants were thoroughly educated on how to complete the case report form.

The e-Motivate4Change program was implemented for 10 minutes with the experimental group to increase their knowledge about MetS and promote health-related activities. Participants could also ask questions. In addition, 5 minutes were allowed for questions and answers. The control group was not provided with the e-Motivate4Change program; they only received a pamphlet. After the e-Motivate4Change program, the experimental group were asked to complete the questionnaire on healthy lifestyle promotion, QOL, and self-efficacy. We then measured their BMI, blood sugar level, cholesterol, and blood pressure again. Participants also completed an online satisfaction survey using a separate site on the e-Motivate4Change app. The control group completed the same surveys and health measurements.

Intervention Design

When participants entered their personal information, the e-Motivate4Change program generated and delivered tailored information and feedback. Such an interaction aimed to increase users' interest and self-efficacy. The e-Motivate4Change program framework is shown in Figure 3. The entity-relation of the e-Motivate4Change system is shown in Figure 4.

To develop an effective intervention program for young adults with a risk of MetS, we surveyed and analyzed users' demands, which could be used as foundational data for the development of further prevention and intervention programs that use mobile health. According to the needs analysis of professional program developers (N=40), nearly half had previously used health apps. The need for tailored prescription for individual users was rated the highest. Furthermore, 13.9% of the sample had used smart health bands. The most requested functions included a record of calorie consumption (88.9%), followed by a pedometer function, analysis of sleeping patterns, and a waterproof device [40].

We required a mobile app for data collection, a back-end server for storage, a server for real-time data analysis, a prediction server, and a message push server for a long-run comprehensive platform. To make exercise an entertaining factor, gamification was necessary. Therefore, we motivated users through the most foundational flow of gamification: mission, point, rewards [41]. After unifying all necessary factors, we developed a customized program for individual users [42].

XSL•FO
RenderX

Figure 3. E-Motivate4Change system architecture.

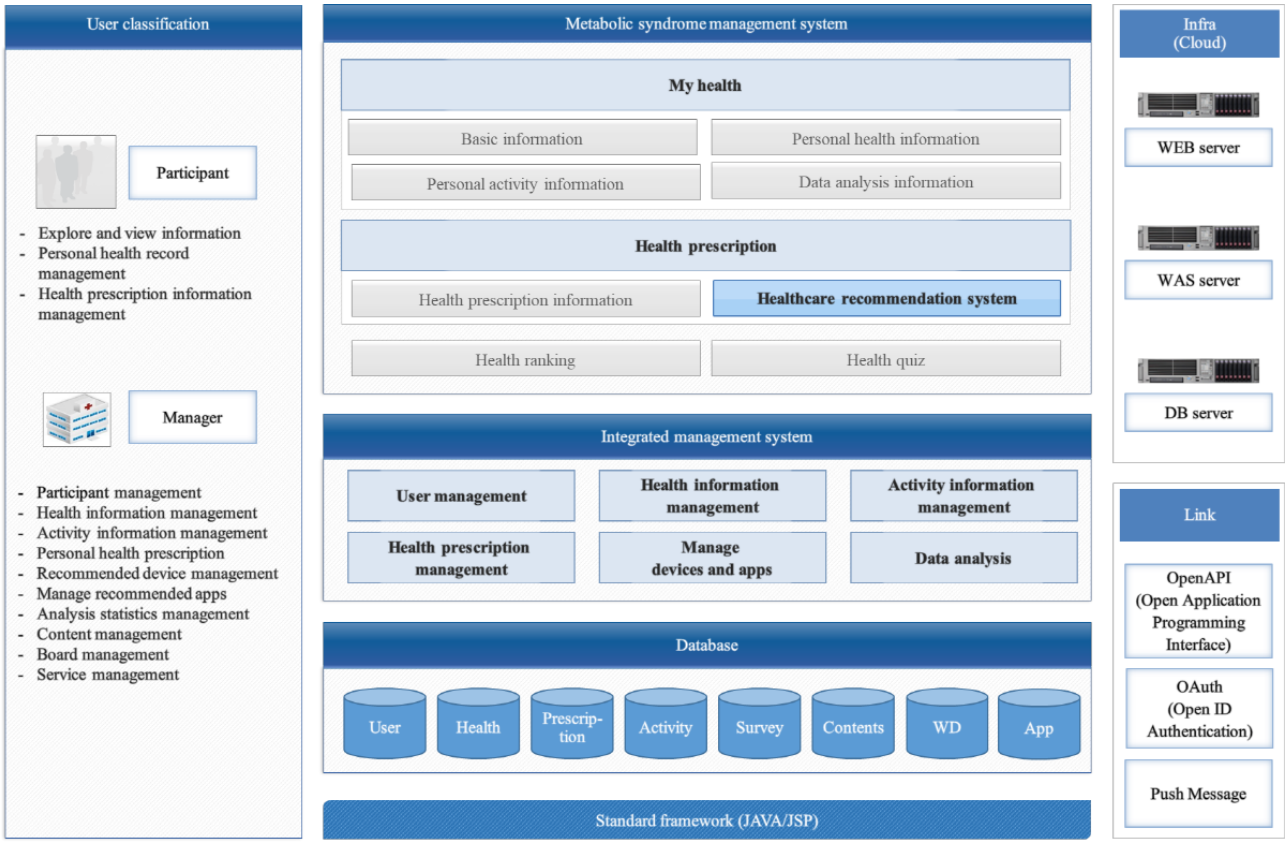
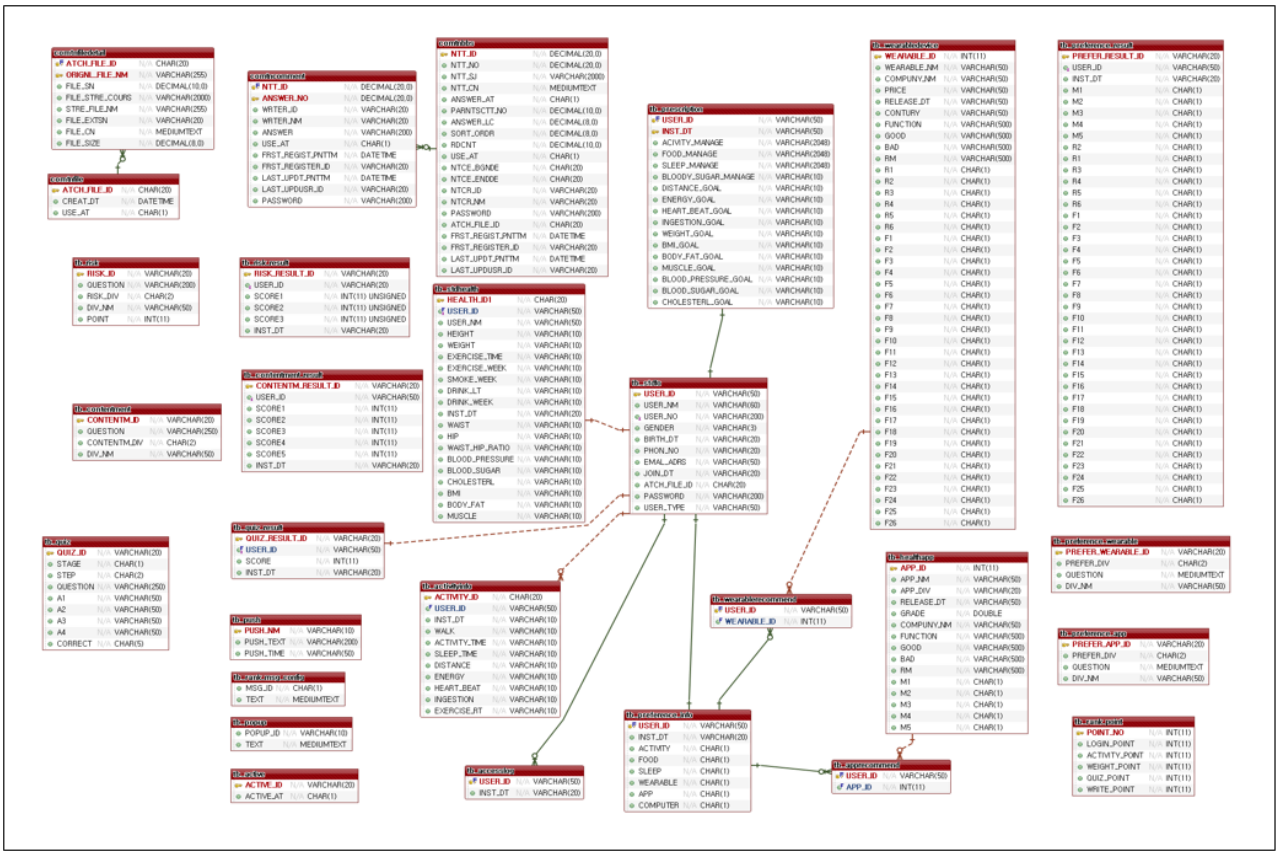


Figure 4. Entity-relation diagram of e-Motivate4Change system.



The e-Motivate4Change program was developed based on three modules (Figure 5). The first is a health behavior prediction and planning module that notified users of their current risk of MetS when they input their current health status, life habits,

XSL•FO
RenderX

and hobbies. It also provided customized health activity information (physical activity, diet, education, etc). In addition, it recommends a customized wearable device for users so that they can select one that they prefer. The second is an education module, which aimed to change users' health behavior based on the I-Change 2.0 model [43]. The education module included the definition of MetS and its causes, diagnosis criteria, incidence rate, further symptoms, prevention, and intervention. Such contents were also customized to maintain users' focus. The third module is an interactive communication module. Unlike previous one-sided feedback, the e-Motivate4Change program was designed to provide interactive feedback so that its users can share information and achieve the same goals together.

On the first page, participants could see a brief description about MetS, as well as a description of the program, types of MetS, and its risk factors. We also provided participants with links to websites (including YouTube) that provided information about effective dance moves for preventing MetS. Users could self-diagnosis their health status at the login page and check their risk of MetS as well as their body image. Risks of MetS were categorized into diet, disease, and lifestyle factors. Once the users answered a quick survey, the program provided scores on diet, disease, and lifestyle, indicating users' risk of MetS as "fair," "warning," or "danger."

Users were also asked to enter their basic health status, which provided an appropriate avatar with the same body image as the user (based on height and weight). There were five types of male and female avatars based on users' risk of MetS, and the avatar was designed to change as users' health status changed. Moreover, users could choose their preferred health management method (app, wearable, diet, or physical activity). The program reflected the needs of the users and recommended customized apps and wearable devices. The recommended algorithm is shown in Figure 6.

Concerning users' health management, we sought the advice of two professional clinical nurses who worked at the national health care company, "Meta Health." We regularly examined users' health information such as weight, obesity rate, body fat, and fasting cholesterol to provide individualized health prescriptions, which included detailed descriptions regarding physical activity, diet, and sleep management.

The community category consisted of a social networking service, questions, and games and quizzes for entertaining factors. Health rankings and a free bulletin board represented the community function. First, users could view their real-time health ranking information, which was determined by points ("mileage") gained from active participation in the program, such as writing on the bulletin or participating in a quiz. Users received mileage for getting the correct answer on a health quiz as well as writing and replying on the bulletin. To encourage participation and motivate the users, the users could check their real-time mileage rankings as well as their rewards. The program also included a "real-life alarm service" so that users could look at their avatars as their risk of MetS changed. Group messaging through a social network was also provided so that the participants could actively engage in information sharing about their health status. Lastly, we included a user preference survey so that we could receive feedback on their experience with the program at the end of the study.

XSL•FO
RenderX

Figure 5. Screenshots of the e-Motivate4Change program.



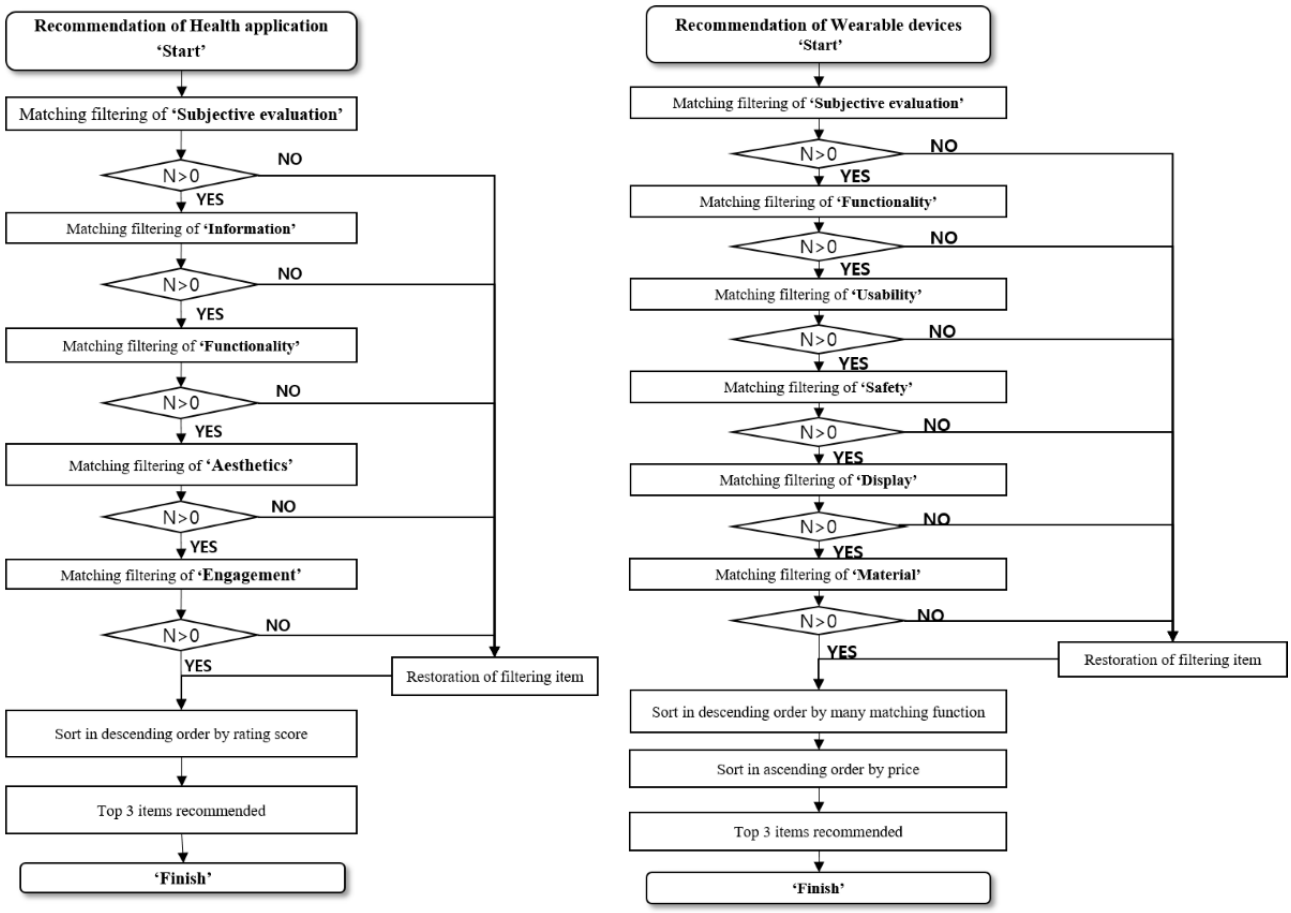
(a) Log-in page | (b) My page | (c) Metabolic syndrome risk analysis

(d) Metabolic syndrome(MetS) explanation | (e) Health quiz | (f) Preference survey

(g) Wearable device recommendation | (h) Health app recommendation | (i) Customized health prescription

XSL•FO
RenderX

Figure 6. Recommended algorithm of e-Motivate4Change system.



Statistical Analyses

All data collected through the study were analyzed using SPSS Statistics (Version 23.0; IBM Corp). To compare the general characteristics between groups, we used numbers, percentages, means, and standard deviations. We conducted chi-squared tests and Fisher exact tests to analyze the differences between groups, and we used *t* tests to determine if there were original differences in the dependent variable before the study. Subsequently, we used independent *t* tests to identify the effects of the program on participants' healthy lifestyle promotion, self-efficacy, and QOL. A repeated-measures analysis of variance was conducted for participants' physiological measurements (ie, weight, BMI, cholesterol, fasting blood sugar, and blood pressure level).

Ethical Considerations

This study was approved by the heads of two universities in D city and by the Institutional Review Board of Keimyung University (number 40525-201704-HR-020-02). Participants were informed about the research purpose and methods, as well as their rights, including the right to cease participation at any point without penalty. All questionnaires were completed anonymously, and participants were told that the results would not be used for anything other than research purposes. Participation was voluntary, and all respondents provided written informed consent.

Results

Participants' General Characteristics and Homogeneity Test

In total, 59 students (57 females) participated in the study (30 in the experimental group and 29 in the control group). Everyone agreed that lifestyle habits matter for the prevention of MetS. Further, 54.2% (26/48) answered that MetS was "a disease related to obesity and high cholesterol," 75% (36/48) answered that MetS causes "adult disease," 25% (12/48) answered that MetS results in "decreased physical strength and abnormal symptoms," and 52.1% (25/48) answered that exercise matters the most for preventing MetS, followed by a better diet (20/48; 41.7%). No significant differences were found in the homogeneity test conducted between the experimental and control groups regarding general characteristics; thus, the two groups were homogeneous.

Effect of the e-Motivate4Change Program on Psychosocial Indicators

We compared healthy lifestyle promotion, self-efficacy, and QOL between the two groups (Table 1). After the program, there were significant differences in healthy lifestyle promotion and self-efficacy, but not QOL. We also used Cronbach α to measure the reliability of the questionnaire.

XSL•FO
RenderX

Table 1. Comparison of psychosocial indicators between the experimental and control groups.

Variable and group	Prestudy, mean (SD)	Poststudy, mean (SD)	*t* value	*P* value
Healthy lifestyle promotion				
Experimental (n=30)	173.13 (21.89)	191.57 (19.19)	3.86	<.001
Control (n=29)	161.66 (26.12)	165.48 (24.76)	N/A[a]	N/A
Quality of life				
Experimental (n=30)	88.73 (9.79)	88.50 (11.23)	.08	.93
Control (n=29)	82.66 (9.28)	83.55 (9.80)	N/A	N/A
Self-efficacy				
Experimental (n=30)	977.97 (185.87)	1140.33 (99.74)	6.00	<.001
Control (n=29)	940.34 (178.11)	957.24 (134.66)	N/A	N/A

[a]N/A: not applicable.

Effect of the e-Motivate4Change Program on Physiological Indicators

Table 2 shows participants' change in BMI, cholesterol, blood sugar, and blood pressure based on different measurement periods. First, there were significant differences in BMI between groups and times; however, the interaction between group and time was nonsignificant. Second, there were significant differences in cholesterol scores per time, group, and the group × time interaction. Lastly, there were no significant differences concerning fasting blood sugar, systolic blood pressure, or diastolic blood pressure between groups.

Table 2. Comparison of physical and physiological indicators between the experimental and control groups.

Variable and group	Baseline, mean (SD)	Week 4, mean (SD)	Week 8, mean (SD)	Week 12, mean (SD)	*F* value (*P* value)		
					Group	Time	Group × time
Cholesterol							
Experimental (n=30)	219.47 (66.09)	193.67 (53.84)	183.23 (52.62)	166.30 (43.15)	4.32 (.42)	9.73 (.001)	6.15 (.01)
Control (n=29)	221.48 (61.01)	224.76 (63.41)	212.03 (60.30)	209.86 (58.98)	N/A[a]	N/A	N/A
Body mass index							
Experimental (n=30)	23.34 (3.57)	23.24 (3.62)	22.67 (3.55)	22.10 (3.47)	1.01 (<.001)	4.71 (.03)	2.66 (.10)
Control (n=29)	22.07 (3.28)	21.96 (3.14)	22.36 (3.62)	21.73 (3.34)	N/A	N/A	N/A
Fasting blood sugar							
Experimental (n=30)	87.97 (12.35)	81.50 (9.13)	80.73 (8.53)	84.47 (7.87)	1.13 (.29)	1.99 (.11)	1.07 (.34)
Control (n=29)	86.14 (6.36)	83.86 (22.65)	85.97 (9.80)	86.14 (6.36)	N/A	N/A	N/A
Blood pressure (systolic)							
Experimental (n=30)	110.80 (10.17)	111.23 (11.27)	109.37 (11.737)	110.80 (10.17)	.55 (.46)	.49 (.66)	.13 (.94)
Control (n=29)	109.62 (10.59)	109.34 (8.47)	108.52 (6.99)	108.52 (9.95)	N/A	N/A	N/A
Blood pressure (diastolic)							
Experimental (n=30)	73.07 (9.79)	73.47 (10.37)	73.47 (10.37)	70.63 (7.60)	1.26 (.26)	1.06 (.37)	1.79 (.16)
Control (n=29)	70.10 (8.40)	69.59 (6.81)	71.90 (6.82)	71.41 (6.06)	N/A	N/A	N/A

[a]N/A: not applicable.

Discussion

Principal Findings

To the best of our knowledge, this study was the first of its kind in Korea. This study was conducted to investigate the effectiveness of the e-Motivate4Change program using mobile apps and wearable devices developed to prevent and manage MetS in young adults. In total, 59 people from 2 universities in Korea participated. (experimental group n=30; control group n=29). The experimental group received a 12-week e-Motivate4Change program intervention, and the control group received MetS training and brochures without the e-Motivate4Change intervention. After the intervention, the experimental group scored significantly higher for health-related lifestyle (t=3.86; P<.001) and self-efficacy (t=6.00; P<.001) than the control group. With regard to BMI, there was a

XSL•FO
RenderX

significant effect by group (F=1.01; P<.001) and group × time interaction (F=4.71; P=.03). Regarding cholesterol, significant main effects for group (F=4.32; P=.04) and time (F=9.73; P<.001) were confirmed.

As the boundaries of traditional industry change owing to the fourth industrial revolution, scholars are anticipating an era of individually tailored services [44]. Such a trend also applies to the national medical policy. The paradigm of worldwide medical policy is gradually shifting from treatment to prevention. Concurrently, considering the rising incidence rate of MetS, the need for appropriate prevention is necessary, especially among young adults who benefit most from lifestyle changes [14,17,45]. Thus, in this study, we developed a MetS prevention program that used mobile apps, wearable devices, and the advice of professional nurses and developers. We then tested its effectiveness on voluntary participants for 12 weeks.

Among both groups, most participants answered that lifestyle was a key preventive factor of MetS, and most defined MetS as a disease related to obesity and cholesterol. Many answered that there was an abnormal physical change when one fails to prevent MetS, and that exercise and diet improvement matter the most. Such answers are consistent with the results of Cornier and colleagues [16] and Clark and colleagues [21].

Participants were asked to record their daily activities in the e-Motivate4Change program for 12 weeks, and the program provided a visual change in the form of their avatars to display their goals and achievements. When we analyzed the risk factors of MetS, there was a significant difference in the mean change in BMI and cholesterol level between the two groups. There was a significant decrease in mean BMI and cholesterol for the experimental group. Such results also showed that, despite their young age, participants had a relatively high BMI and cholesterol level and were at a high risk for MetS [34].

For the experimental group, continued participation in the e-Motivate4Change program decreased their cholesterol and BMI significantly. This result aligns with Aktas and colleagues' [46] study in Turkey, which tracked the lifestyles of participants for 12 weeks. These findings show that people with a chronic disease (including MetS) could lower their blood sugar and cholesterol by continuously improving their lifestyle. However, while fasting blood sugar and blood pressure levels decreased marginally for the experimental group, there was no significant difference between the two groups. One possible explanation is that all participants had average levels at baseline.

Concerning the psychosocial aspects of this study, we must note that the experimental group's self-efficacy and healthy lifestyle promotion scores increased significantly. This result aligns with Khalili and colleagues [22] and Shekari and colleagues [47], who showed that a health education program based on social recognition theory increased young adults' self-efficacy. However, when we look at QOL, the score of the experimental group slightly decreased, and there was no significant difference between the two groups. One potential explanation could be the short duration of the study. The study lasted for 12 weeks, which may have been too short to impact participants' QOL. Therefore, further studies should implement a longer intervention to determine if it could increase young adults' health-related QOL. Moreover, it is critical to increase young adults' self-efficacy and promote health-related lifestyles. Their needs and preferences should be properly understood so that more customized management programs can be developed. The e-Motivate4Change program is a customized interactive program that provides individualized information (ie, an avatar that reflects personal body shape and recommends health apps and wearable devices) and feedback based on users' input. The program also used big data technology with gamification functions to entertain users, ultimately promoting a healthy lifestyle and educating users about MetS.

Generally, MetS can be prevented through changing diet habits and lifestyle. However, it can be challenging to maintain regular physical activities and good diet habits at an individual level. Thus, by recommending appropriate mobile apps and wearable devices and encouraging their use, this study made it possible for the users to engage in sustained health management. The findings of this study align with those from previous studies that revealed the significance of health trackers and smart apps that are free of time and space limitations for a sustained increase in physical activities [30,48]. Previous recommendation systems were rather one-sided and manager-centered, which made it difficult for users to stay entertained. Consequently, most users stopped using the service.

To find a solution, we developed a needs-based customized recommendation system. Some of the most important features of this program can be summarized as follows. First, the program provides individualized health management prescriptions for the prevention of MetS. Based on the medical paradigm shift from treatment to prevention and management owing to an increasing risk of chronic disease, effective prevention is largely changing the clinical health sectors, as is the fourth industrial revolution [45]. Thus, the combination of data and artificial intelligence will enable individualized medical service. When supported with digital technology, the prevention and management of chronic disease will become much more effective and efficient [45,49]. Furthermore, it is critical for health managers to learn new competencies and skills in rapidly changing medical environments [49]. Thus, health managers for the e-Motivate4Change program analyzed the real-time data of participants and provided customized consulting and health prescriptions, which brought about positive changes in health-related lifestyle and self-care. An efficient system will increase the usability of medical services, decrease the symptoms of chronic disease, and increase cost efficiency, while also improving patients' medical experiences.

Clinical nurses play a critical role in enhancing health; therefore, more effort is needed to expand the role of nurses and support their competency [50]. Clinical nurses will pioneer a new field of expertise by using big data technology and artificial intelligence. Thus, it is necessary to combine the humane and technological aspects of nurse's expertise [45,49,51].

Second, previously one-sided services may not include gamification and entertainment factors. Such improvement provides individualized services for the users, maintaining their high level of interest and their motivation, encouraging them to keep using the program. For sustained management of chronic

XSL•FO
RenderX

diseases, including MetS, more user-centered content is necessary.

Limitations

The personalized service recommendation system proposed in this study provides appropriate dietary and physical activity recommendations to improve youth health to prevent MetS. The ranking system included in the program provides entertainment and motivation to encourage continued use of the program. Moreover, this study can be regarded as different from previous studies as the participants were aged <30 years, whereas the typical age range of MetS study participants is 40 to 60 years. However, this study had limitations. We used a portable body fat measurement device that may not be able to provide accurate and detailed data. At the same time, it was not possible to monitor the type, intensity, or length of physical activity. Subsequent studies should aim to achieve this. In addition, future studies could aim to recruit a participant group with a wider age range than this study. For example, future health intervention programs could include young working women and working mothers.

Conclusions

Despite its limitations, this study is meaningful in that we tried to increase accessibility and usability through smartphone apps while also bringing about changes in the perceptions and health behaviors of young adults by providing individualized recommendations from professional clinical nurses [47]. Moreover, unlike previous group-based education, this program encouraged users' voluntary participation in creating a healthy lifestyle, which allows for better long-term management and operation of the program.

In conclusion, the e-Motivate4Change program was associated with increased physical activity, decreased BMI, lower cholesterol, and increased self-efficacy among experimental group participants, thus effectively promoting a health-related lifestyle. Moreover, by providing programs to prevent and manage MetS, the program informed further studies. Therefore, this study provides foundational data to design further preventive strategies for relevant chronic disorders (eg, diabetes, high blood pressure, hyperlipidemia, cardiovascular disease).

Acknowledgments

The authors gratefully acknowledge Shin-Sil Kim (CEO) and Eun-Kyung Song (manager) of Meta Healthcare for their support and health consultations during this project. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) and funded by the Ministry of Education (number NRF-2015R1C1A1A01055753; 2018R1D1A3B07050652).

Authors' Contributions

All authors had full access to all data and take responsibility for the integrity of the data and the accuracy of the data analysis. JSL and SKL contributed to study concept and design. JSL and MAK contributed to the acquisition and statistical analysis of data. JSL and SKL contributed to the interpretation of data and drafting of the manuscript. SKL contributed critical revision of the manuscript for important intellectual content, obtained funding, and supervised the study. SKL provided administrative, technical, and material support.

Conflicts of Interest

None declared.

References

1. World Health Organization. General meeting of the WHO global coordination mechanism on the prevention and control of noncommunicable diseases: meeting report. 2018 Presented at: General meeting of the WHO global; November 5-6, 2018; Geneva, Switzerland.
2. Wu S, Jo E, Ji H, Kim K, Park J, Kim BH, et al. Exergaming Improves Executive Functions in Patients With Metabolic Syndrome: Randomized Controlled Trial. JMIR Serious Games 2019 Jul 31;7(3):e13575 [FREE Full text] [doi: 10.2196/13575] [Medline: 31368441]
3. Ford ES, Li C. Metabolic syndrome and health-related quality of life among U.S. adults. Ann Epidemiol 2008 Mar;18(3):165-171. [doi: 10.1016/j.annepidem.2007.10.009] [Medline: 18280918]
4. Corica F, Corsonello A, Apolone G, Mannucci E, Lucchetti M, Bonfiglio C, et al. Metabolic syndrome, psychological status and quality of life in obesity: the QUOVADIS Study. Int J Obes (Lond) 2008 Jan;32(1):185-191. [doi: 10.1038/sj.ijo.0803687] [Medline: 17653068]
5. Chedraui P, Hidalgo L, Chavez D, Morocho N, Alvarado M, Huc A. Quality of life among postmenopausal Ecuadorian women participating in a metabolic syndrome screening program. Maturitas 2007 Jan 20;56(1):45-53. [doi: 10.1016/j.maturitas.2006.05.008] [Medline: 16797890]
6. Han J, Park H, Shin C, Chang HM, Yun KE, Cho SH, et al. Metabolic syndrome and quality of life (QOL) using generalised and obesity-specific QOL scales. Int J Clin Pract 2009 May;63(5):735-741. [doi: 10.1111/j.1742-1241.2009.02021.x] [Medline: 19392923]

XSL•FO
RenderX

7. World Health Organization. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med 1995 Nov;41(10):1403-1409. [doi: 10.1016/0277-9536(95)00112-k] [Medline: 8560308]
8. Kruithof N, Haagsma J, Karabatzakis M, Cnossen M, de Munter L, van de Ree C, et al. Validation and reliability of the Abbreviated World Health Organization Quality of Life Instrument (WHOQOL-BREF) in the hospitalized trauma population. Injury 2018 Oct;49(10):1796-1804. [doi: 10.1016/j.injury.2018.08.016] [Medline: 30154022]
9. World Health Organization. Diet, nutrition and the prevention of chronic diseases. World Health Organ Tech Rep Ser 2003;916:i-viii. [Medline: 12768890]
10. Frates EP, Xiao RC, Sannidhi D, McBride Y, McCargo T, Stern TA. A Web-Based Lifestyle Medicine Curriculum: Facilitating Education About Lifestyle Medicine, Behavioral Change, and Health Care Outcomes. JMIR Med Educ 2017 Sep 11;3(2):e14 [FREE Full text] [doi: 10.2196/mededu.7587] [Medline: 28893726]
11. O'Brien WJ, Shultz SP, Firestone RT, George L, Kruger R. Ethnic-specific suggestions for physical activity based on existing recreational physical activity preferences of New Zealand women. Aust N Z J Public Health 2019 Oct;43(5):443-450. [doi: 10.1111/1753-6405.12902] [Medline: 31091346]
12. Bowden Davies KA, Sprung VS, Norman JA, Thompson A, Mitchell KL, Harrold JOA, et al. Physical Activity and Sedentary Time: Association with Metabolic Health and Liver Fat. Med Sci Sports Exerc 2019 Jun;51(6):1169-1177 [FREE Full text] [doi: 10.1249/MSS.0000000000001901] [Medline: 30694971]
13. Vilchis-Gil J, Klünder-Klünder M, Flores-Huerta S. Effect on the Metabolic Biomarkers in Schoolchildren After a Comprehensive Intervention Using Electronic Media and In-Person Sessions to Change Lifestyles: Community Trial. J Med Internet Res 2018 Feb 05;20(2):e44 [FREE Full text] [doi: 10.2196/jmir.9052] [Medline: 29402762]
14. Messiah S, Lebron C, Arheart K. Prevalence of the metabolic syndrome in US youth. In: Global Perspectives on Childhood Obesity. Cambridge, MA, USA: Academic Press; 2011:105-117.
15. Sankaran S, Dendale P, Coninx K. Evaluating the Impact of the HeartHab App on Motivation, Physical Activity, Quality of Life, and Risk Factors of Coronary Artery Disease Patients: Multidisciplinary Crossover Study. JMIR mHealth uHealth 2019 Apr 04;7(4):e10874 [FREE Full text] [doi: 10.2196/10874] [Medline: 30946021]
16. Driver S, Swank C, Froehlich-Grobe K, McShan E, Calhoun S, Bennett M. Weight Loss After Stroke Through an Intensive Lifestyle Intervention (Group Lifestyle Balance-Cerebrovascular Accident): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Oct 18;8(10):e14338 [FREE Full text] [doi: 10.2196/14338] [Medline: 31628790]
17. Orchard TJ, Temprosa M, Goldberg R, Haffner S, Ratner R, Marcovina S, Diabetes Prevention Program Research Group. The effect of metformin and intensive lifestyle intervention on the metabolic syndrome: the Diabetes Prevention Program randomized trial. Ann Intern Med 2005 Apr 19;142(8):611-619 [FREE Full text] [doi: 10.7326/0003-4819-142-8-200504190-00009] [Medline: 15838067]
18. Lotfaliany M, Bowe SJ, Kowal P, Orellana L, Berk M, Mohebbi M. Depression and chronic diseases: Co-occurrence and communality of risk factors. J Affect Disord 2018 Dec 01;241:461-468. [doi: 10.1016/j.jad.2018.08.011] [Medline: 30149333]
19. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and Acceptance of mHealth in Patients With Cardiovascular Diseases: A Cross-Sectional Study. JMIR mHealth uHealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
20. Athyros VG, Bouloukos VI, Pehlivanidis AN, Papageorgiou AA, Dionysopoulou SG, Symeonidis AN, MetS-Greece Collaborative Group. The prevalence of the metabolic syndrome in Greece: the MetS-Greece Multicentre Study. Diabetes Obes Metab 2005 Jul;7(4):397-405. [doi: 10.1111/j.1463-1326.2004.00409.x] [Medline: 15955126]
21. Clark RL, Famodu OA, Holásková I, Infante AM, Murray PJ, Olfert IM, et al. Educational intervention improves fruit and vegetable intake in young adults with metabolic syndrome components. Nutr Res 2019 Feb;62:89-100 [FREE Full text] [doi: 10.1016/j.nutres.2018.11.010] [Medline: 30803510]
22. Khalili D, Asgari S, Lotfaliany M, Zafari N, Hadaegh F, Momenan A, et al. Long-Term Effectiveness of a Lifestyle Intervention: A Pragmatic Community Trial to Prevent Metabolic Syndrome. Am J Prev Med 2019 Mar;56(3):437-446. [doi: 10.1016/j.amepre.2018.10.029] [Medline: 30777162]
23. Tai H, Tzeng I, Liang Y, Liao H, Su C, Kung W. Interventional Effects of Weight-Loss Policy in a Healthy City among Participants with Metabolic Syndrome. Int J Environ Res Public Health 2019 Jan 24;16(3):323-332 [FREE Full text] [doi: 10.3390/ijerph16030323] [Medline: 30682780]
24. Kim H, Park E, Lee S, Kim M, Park EJ, Hong S. Self-Management of Chronic Diseases Among Older Korean Adults: An mHealth Training, Protocol, and Feasibility Study. JMIR mHealth uHealth 2018 Jun 29;6(6):e147 [FREE Full text] [doi: 10.2196/mhealth.9988] [Medline: 29959109]
25. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of Smartphone App-Based Interactive Management on Glycemic Control in Chinese Patients With Poorly Controlled Diabetes: Randomized Controlled Trial. J Med Internet Res 2019 Dec 09;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
26. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable Sensor/Device (Fitbit One) and SMS Text-Messaging Prompts to Increase Physical Activity in Overweight and Obese Adults: A Randomized Controlled Trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]

XSL•FO
RenderX

27. Wang B, Park J, Choi I. Influencing Factors for the Adoption of Smartphone Healthcare Application. The Journal of the Korea Contents Association 2011 Oct 28;11(10):396-404. [doi: 10.5392/jkca.2011.11.10.396]
28. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer Perceptions of Wearable Technology Devices: Retrospective Review and Analysis. JMIR mHealth uHealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
29. Rodriguez-Villegas E, Iranmanesh S, Imtiaz SA. Wearable Medical Devices: High-Level System Design Considerations and Tradeoffs. IEEE Solid-State Circuits Mag 2018;10(4):43-52. [doi: 10.1109/mssc.2018.2867247]
30. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013 Nov 06;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
31. Kang Y. Health promoting lifestyle, self-efficacy, and life satisfaction of middle-aged Women. Journal of the Korean Data Analysis Society 2013;15(2):937-948.
32. Olson EA, McAuley E. Impact of a brief intervention on self-regulation, self-efficacy and physical activity in older adults with type 2 diabetes. J Behav Med 2015 Dec;38(6):886-898 [FREE Full text] [doi: 10.1007/s10865-015-9660-3] [Medline: 26162648]
33. Chen J, Allman-Farinelli M. Impact of Training and Integration of Apps Into Dietetic Practice on Dietitians' Self-Efficacy With Using Mobile Health Apps and Patient Satisfaction. JMIR mHealth uHealth 2019 Mar 04;7(3):e12349 [FREE Full text] [doi: 10.2196/12349] [Medline: 30829577]
34. Balgoon MJ, Al-Zahrani MH, Alkhattabi NA, Alzahrani NA. The correlation between obesity and metabolic syndrome in young female university students in the Kingdom of Saudi Arabia. Diabetes Metab Syndr 2019;13(4):2399-2402. [doi: 10.1016/j.dsx.2019.06.015] [Medline: 31405650]
35. Chen S, Chen S, Lai Y, Chen P, Huang T, Lin C, et al. Correlates Between Health Beliefs and Health-Promoting Lifestyle Profiles in the Development of Metabolic Syndrome in Taiwan. Asia Pac J Public Health 2019 Jan;31(1):30-38. [doi: 10.1177/1010539518816245] [Medline: 30525917]
36. Jun MK, Ha JY. [Effect of Smartphone Apps Applying BodyThink Program on Obesity in Adolescent Girls]. J Korean Acad Nurs 2016 Jun;46(3):390-399. [doi: 10.4040/jkan.2016.46.3.390] [Medline: 27411766]
37. Walker SN, Sechrist KR, Pender NJ. The health-promoting lifestyle profile II. Nursing Res 1995:1.
38. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The Self-Efficacy Scale: Construction and Validation. Psychol Rep 2016 Aug 31;51(2):663-671. [doi: 10.2466/pr0.1982.51.2.663]
39. Lee K. Relationship between uric acid and hepatic steatosis among Koreans. Diabetes & Metabolism 2009 Dec;35(6):447-451. [doi: 10.1016/j.diabet.2009.04.011]
40. Kang M, Lee S. Analysis of users' needs for developing mobile health based prevention and intervention programs for the metabolic syndrome in college students. AJMAHS 2017 Sep 30;7(9):429-442. [doi: 10.14257/ajmahs.2017.09.73]
41. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Reply: Some considerations on digital health validation. NPJ Digit Med 2019;2:103-104 [FREE Full text] [doi: 10.1038/s41746-019-0176-z] [Medline: 31637310]
42. Wang W, Zhang H, Lopez V, Wu VX, Poo DCC, Kowitlawakul Y. Improving awareness, knowledge and heart-related lifestyle of coronary heart disease among working population through a mHealth programme: study protocol. J Adv Nurs 2015 Sep;71(9):2200-2207. [doi: 10.1111/jan.12668] [Medline: 25879395]
43. Oosterom-Calo R, Te VSJ, Stut W, Brug J. Development of Motivate4Change Using the Intervention Mapping Protocol: An Interactive Technology Physical Activity and Medication Adherence Promotion Program for Hospitalized Heart Failure Patients. JMIR Res Protoc 2015;4(3):e88 [FREE Full text] [doi: 10.2196/resprot.4282] [Medline: 26195072]
44. Lee W. A study on Recommendation Service System for the Customized Convergence Wellness Contents. Journal of Korea Multimedia Society 2017 Feb 28;20(2):322-329. [doi: 10.9717/kmms.2017.20.2.322]
45. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
46. Aktas MF, Mähler A, Hamm M, Perger G, Simon F, Westenhöfer J, et al. Lifestyle interventions in Muslim patients with metabolic syndrome-a feasibility study. Eur J Clin Nutr 2019 May;73(5):805-808. [doi: 10.1038/s41430-018-0371-z] [Medline: 30538299]
47. Shekari R, Vahedian SM, Gholian aval M, Esmaily H, Attarian F. Evaluating the relationship between dietary pattern and dietary self-efficacy in teenage girl students. Journal of Midwifery and Reproductive Health 2019;7(4):1921-1928. [doi: 10.22038/jmrh.2019.27827.1301]
48. Kim DJ, Kim MS, Park NY, Lee SG, Lee YK. Effect of Activity Tracker & Application Use on Adherence, Metabolic Syndrome Factor and Body Composition in Middle and Senior-Aged. Asian J Kinesiol 2018 Apr;20(2):28-37. [doi: 10.15758/ajk.2018.20.2.28]
49. Aiello L, Calzone K. Precision medicine: preparing nurses in genomics for today's healthcare. Dean's Notes. 2018. URL: https://www.ajj.com/sites/default/files/services/publishing/deansnotes/fall2018.pdf [accessed 2020-07-20]
50. Kemppainen V, Tossavainen K, Turunen H. Nurses' roles in health promotion practice: an integrative review. Health Promot Int 2013 Dec;28(4):490-501. [doi: 10.1093/heapro/das034] [Medline: 22888155]

XSL•FO
RenderX

51. Kim KJ, Jang BY, Jung JY, Park OW. The Coming of the 4th Industrial Revolution and the HRD Issues for Nurses - Prospects and Challenges. Korean Journal of Resources Development 2018 Sep 30;21(3):137-159. [doi: 10.24991/kjhrd.2018.09.21.3.137] [Medline: 22888155]

Abbreviations

MetS: metabolic syndrome
QOL: quality of life
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.11.19; peer-reviewed by B Oh, A Cordola-Hsu; comments to author 30.03.20; revised version received 22.05.20; accepted 14.06.20; published 30.07.20.

Please cite as:
Lee JS, Kang MA, Lee SK
Effects of the e-Motivate4Change Program on Metabolic Syndrome in Young Adults Using Health Apps and Wearable Devices: Quasi-Experimental Study
J Med Internet Res 2020;22(7):e17031
URL: http://www.jmir.org/2020/7/e17031/
doi:10.2196/17031
PMID:32729838

©Ji-Soo Lee, Min-Ah Kang, Soo-Kyoung Lee. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Dose-Response Relationship of a Web-Based Tailored Intervention Promoting Human Papillomavirus Vaccination: Process Evaluation of a Randomized Controlled Trial

Mirjam Pot[1], PhD; Theo GWM Paulussen[1], PhD; Robert AC Ruiter[2], Prof Dr; Liesbeth Mollema[3], PhD; Miranda Hofstra[1], MSc; Hilde M Van Keulen[1], PhD

[1]Child Health, Netherlands Organization for Applied Scientific Research (TNO), Leiden, Netherlands

[2]Department of Work and Social Psychology, Maastricht University, Maastricht, Netherlands

[3]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, Netherlands

Corresponding Author:
Mirjam Pot, PhD
Child Health
Netherlands Organization for Applied Scientific Research (TNO)
Schipholweg 77
Leiden, 2316ZL
Netherlands
Phone: 31 643234293
Email: mirjampot90@gmail.com

Abstract

Background: In the Netherlands, human papillomavirus (HPV) vaccination uptake remains low. To improve informed decision making (IDM) and HPV vaccination acceptability, we systematically developed an interactive, web-based tailored intervention to which mothers of Dutch girls were invited to participate.

Objective: The aim of this study was to provide insight into the intervention's working mechanisms by evaluating (1) program use, (2) program acceptability, and (3) the relationship of program use with program acceptability and intervention effects (ie, dose-response relationship).

Methods: Only mothers from the intervention arm of a randomized controlled trial that assessed the effectiveness of the web-based, tailored intervention were included in this study. They were invited to visit the website of the web-based intervention between baseline (January 2015, just before access to the intervention) and follow up (March 2015, prior to the first HPV vaccination). Indicators for program use were time of website use (ie, duration of intervention exposure) and completeness (ie, proportion of all available web pages visited). HPV vaccination uptake registered by Praeventis was used as the primary outcome. Secondary outcomes were IDM, decisional conflict, and social-psychological determinants of HPV vaccination uptake.

Results: Among the 3995 invited mothers, 2509 (62.80%) logged on to the website, 2239 of whom (89.24%) visited at least one page of the intervention components. On average, mothers spent 21.39 minutes (SD 12.41) on the website and completed 50.04% (SD 26.18%) of the website components. Participants rated the website 7.64 (SD 1.39) on a 10-point scale. Program acceptability was significantly associated with completeness (β=4.36, P<.001), but not with time of website use (β=–.07, P=.77). Intention-to-treat analysis (N=3995) showed a significant positive effect of completeness on all outcome measures (all P<.003; Bonferroni-corrected α=.05/15 factors), including on HPV vaccination uptake. Time of website use had a significant positive effect on all outcomes (all P<.003), except for uptake (P=.20), risk perception when not vaccinated (P=.14), subjective norms (P=.03), and habit (P=.01).

Conclusions: Program use and acceptability of the intervention were adequate. Completeness was positively associated with acceptability. Furthermore, positive effects (ie, dose-response effects) were found for completeness and time of website use on the mothers' IDM, decisional conflict, and almost all of the social-psychological determinants of HPV vaccination acceptability. In addition, the extent to which mothers completed the intervention had a positive impact on their daughters' vaccination uptake. This indicates that the web-based, tailored intervention fits well with the mothers' needs, and that completeness of use is essential for improving HPV vaccination uptake, acceptability, and IDM. Program use should therefore be promoted.

Trial Registration: Netherlands Trial Register NTR4795; https://www.trialregister.nl/trial/4795

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e14822) doi:10.2196/14822

KEYWORDS

HPV vaccination; web-based tailored intervention; process evaluation; intervention use; dose-response relationship; randomized controlled trial; informed decision making

Introduction

Cervical cancer is the fourth most common cancer among women worldwide. Globally, in 2018, there were 569,847 new cases and 311,365 deaths caused by cervical cancer [1]. Persistent infection by human papillomavirus (HPV) appears to be the major cause of cervical cancer; nearly all cases (over 99%) are attributable to an HPV infection [2,3]. There is also an increasing body of evidence that strongly links HPV infection with cancers of the anus, vulva, vagina, penis, and head and neck (eg, [4,5]). In the Netherlands, on an annual basis since 2010, girls are invited by the Dutch National Immunization Program to receive the vaccine against HPV upon reaching 13 years of age. However, HPV vaccination uptake remains low (45.5% in 2017) [6]. The uptake needs to be higher to reduce the incidence of cervical cancer in women. Therefore, we developed an interactive, web-based, tailored intervention promoting HPV vaccination acceptability among Dutch mothers of invited girls for HPV vaccination [7]. Further, to improve vaccine acceptability, important intervention objectives were to improve informed decision making (IDM) and to decrease decisional conflict [7].

To date, only four studies have examined the effectiveness of tailored interventions in promoting HPV vaccination acceptance [8-11]; three of these studies showed positive effects [9-11]. To our knowledge, the only tailored intervention that was web-based turned out to be ineffective [8]. The advantage of web-based interventions is that they have the potential to enable large-scale application at relatively low cost [12] and therefore can have a substantial impact at the population level [13].

Furthermore, only two of the existing studies using tailored interventions to promote HPV vaccination acceptance incorporated a process evaluation [9,11]. Process evaluation provides insight into the extent to which an intervention was used, which is relevant for determining factors that influence the effectiveness of an intervention [14-16]. Specifically, this analysis can provide insight into whether an intervention is ineffective due to implementation failure or to the inferior quality of the intervention method(s), thereby avoiding the so-called type III error [17]. It also provides recommendations for intervention improvement.

For the use of online interventions, three aspects are relevant: a first visit, staying, and revisiting [18,19]. In the context of decision making about HPV vaccination, we consider a first visit and staying as the most relevant aspects. Revisiting seems less relevant, since getting the HPV vaccine is hardly a repetitive behavior (ie, it consists of two injections). Within the electronic health (eHealth) arena, limited intervention use in terms of a first visit and staying is a broadly recognized problem [20]. The extent to which one uses an intervention may positively impact the intervention effects (ie, the so-called "dose-response relation"). Positive dose-response relationships have been found for eHealth interventions (eg, [21-23]). In the context of promoting HPV vaccination acceptability, none of the studies that incorporated a process evaluation addressed the dose-response relationship (for a review, see [24]). Moreover, data on the use of web-based interventions have been poorly reported [20,25] or not reported at all [26].

Furthermore, assessing program acceptability may shed light on the causes of nonuse or incomplete use [27]. An intervention needs to be considered appropriate by the target group and to suit their needs in order to reach the intended outcomes [28]. Acceptability will improve interaction and the users' feelings of engagement, and, as a consequence, the likelihood of health-related behavior changes [29].

Results from the effect evaluation showed positive intervention effects on IDM, decisional conflict, and nearly all determinants of HPV vaccination uptake (for more details on the results of the effect evaluation, see [30]). This paper describes the process evaluation of the web-based tailored intervention promoting HPV vaccination acceptance [7]. The aim of this process evaluation was to provide insights into the intervention's working mechanisms by evaluating (1) program use, (2) program acceptability, and (3) the relationship of program use with program acceptability and intervention effects (ie, dose-response relationships).

Methods

Participants and Study Design

The study was approved by the Ethical Committee of the VU Medical Center in Amsterdam (Netherlands Trial Register: NTR4795). This process evaluation was part of a randomized controlled trial on the effectiveness of the intervention [30]. For the purpose of this study, only participants from the experimental condition (ie, those invited to use the web-based intervention; N=3995) were included in the analyses. Details of the trial design, such as power calculations, participant eligibility, and recruitment procedures, have been published elsewhere [30]. This study was conducted between January 2015 (baseline, just before access to the intervention) and March 2015 (follow up, prior to the first HPV vaccination), in line with the national HPV vaccination round of 2015. Mothers were invited to use the web-based intervention between baseline and follow up (mid-January 2015).

Intervention Components

We developed the web-based tailored intervention with virtual assistants using the 6-step Intervention Mapping protocol for developing theory- and evidence-based health promotion interventions [31]. In short, the web-based intervention provided mothers with tailored feedback about topics on HPV vaccination, delivered by two virtual assistants. Multimedia Appendix 1

XSL•FO
RenderX

presents a selection of screenshots of the website. The website consisted of 4 menu options: (1) 2-sided, tailored information about HPV vaccination, (2) a decisional balance in which mothers could weigh their perceived pros and cons, (3) practical information, and (4) frequently asked questions. Within menus 1-3, mothers could visit several components. In addition, mothers were given the opportunity to visit an "in-depth" information page within some of the components (eg, educational movies). Table 1 provides a brief overview of the intervention components. More detailed descriptions of the intervention, including its theoretical basis, are published elsewhere [7].

Table 1. Brief description of intervention components.

Menu and component	Description
Information about HPV[a] vaccination	
General information	Mothers are provided with general information about HPV, cervical cancer, and HPV vaccination.
Importance of vaccinating at a young age	Mothers are challenged to consider whether the age of their daughter is appropriate to get vaccinated against HPV. The relationship between sexual activity in relation to the HPV vaccine is also discussed.
HPV-related risks	Mothers are asked to estimate both the risk of their daughter getting infected with HPV and the risk of their daughter developing cervical cancer later in life, and are then provided with tailored feedback accordingly.
Methods to protect against cervical cancer	Mothers are asked to rate the effectiveness of alternative methods for protecting against cervical cancer and are then provided with tailored feedback according to their answers.
From HPV to cervical cancer	Mothers are provided with an explanation of how infection with HPV can lead to cervical cancer (eg, by viewing an educational video).
Facts and stories	Mothers are provided with several statements regarding HPV, cervical cancer, and HPV vaccination, and are asked to indicate whether these are true (a fact) or false (a story). They then receive tailored feedback accordingly.
Side effects	Mothers are presented with a variety of potential side effects of the HPV vaccine and are asked to indicate whether or not they are scientifically proven. They are then provided with tailored feedback, stating the correct responses.
Effectiveness	Mothers are asked about the effectiveness of the HPV vaccine in protecting both against their daughter getting infected with HPV and developing cervical cancer. They are then provided with tailored feedback about the effectiveness of HPV vaccination.
Other mothers	Mothers are asked to indicate what they think most mothers in their direct environment will decide regarding their daughters' HPV vaccination. They are then provided with tailored feedback and are shown the actual HPV vaccination uptake in different regions of the Netherlands in 2014.
Vaccine working mechanisms	Mothers are explained in a generic way how the HPV vaccine works in protecting against HPV and cervical cancer with an educational video.
Weighing the pros and cons	
Decisional balance	Mothers are presented with a list of pros and cons of HPV vaccination. Based on pros and cons mothers marked as most salient, a decisional balance reveals their current position on a scale ranging between wanting and not wanting to get their daughter vaccinated.
Value clarification	Mothers are invited to list their central values for life and link these to HPV vaccination.
Practical information	Practical information provides mothers with information such as how and where to receive the HPV vaccine and provides them with advice on how they discuss HPV vaccination with personal important others (eg, their daughter and partner).
Frequently asked questions	
About HPV vaccination	Frequently asked questions provides answers to known questions about the HPV vaccine (eg, "does my daughter know if she's infected with HPV?") or getting the HPV vaccine (eg, "where do I get the HPV vaccine?").
Problems with the website	Mothers are provided with possible solutions to problems with the website such as not being able to hear or see the virtual assistants.

[a]HPV: human papillomavirus.

Measurements

All data were derived from web-based self-report questionnaires and computer logs.

Program Use

Program use was measured according to time of website use and completeness. Both factors were constructed based on computer logs that registered the pages and components the

XSL•FO
RenderX

mothers had visited and the amount of time they spent on the website during their visit(s).

Time of website use represents the total amount of time spent on the website. Each time mothers logged on to the website, a record was created, in which both the date and time of the first and last page the mothers entered were registered. Time per session was calculated by subtracting the time of the last entered page from the time of the earliest entered page. The total time of website use was then calculated by adding up the amount of time spent in each session.

Completeness represents the proportion of all available web pages visited by the mother while she was logged on to the website, ranging from 0% (visited no pages) to 100% (visited all pages).

Program Acceptability

Program acceptability was assessed at the 6-week follow up by asking mothers to rate the intervention and the virtual assistants on a 10-point scale, ranging from 0 (very bad) to 10 (excellent). The grade for the website was used as an indicator for overall program acceptability to examine the association with program use. In addition, mothers evaluated the information provided by the website (eg, credibility, relevance), perceived user control (eg, perceived degree of autonomy), and how they perceived the virtual assistants (eg, reliability) (See Table 2 and Table 3 for an overview). These acceptability measures were assessed on a 7-point scale, and some were averaged into one scale. All composite scores showed sufficient internal consistency (Cronbach $\alpha \geq .71$ for scales with >2 items; Pearson $r \geq 0.84$ for scales with 2 items).

XSL•FO
RenderX

Table 2. Overview of the program acceptability measures for the website.

Measures	Items	Score and scale	Cronbach α or Pearson *r*	Reference
Rate	On average, how would you rate the website on a scale from 0 to 10?	0=very bad to 10=excellent	N/A[a]	[32]
Interest	In general, what did you think of the website?	1=very uninteresting to 7=very interesting; 1=boring to 7=engaging	0.84 (*r*)	[33]
Informative	In general, what did you think of the website?	1=very uninformative to 7=very informative; 1=very noneducational to 7 =very educational	0.86 (*r*)	[34]
Perceived user control	I felt that I had a lot of control over my visiting experiences at this website. While I was on the website, I could choose freely what I wanted to see. While surfing the website, I felt in control. While surfing the website, my actions decided the kind of experiences I got.	1=strongly disagree to 7=strongly agree	.71 (α)	[35]
Elaboration	How well did you read the information?	1=not carefully at all to 7=very carefully	N/A	[36]
Support	The website has helped me decide about my daughter's HPV[b] vaccination.	1=strongly disagree to 7=strongly agree	N/A	[37]
Recall	I can recall the information from the website.	1=strongly disagree to 7=strongly agree	N/A	[32]
Personal relevance	I considered the website to be personally relevant.	1=strongly disagree to 7=strongly agree	N/A	[32,37]
Tailoring	I considered the information on the website to be…	1=not tailored to me at all to 7=very tailored to me	N/A	[37]
Comprehensibility	I considered the information on the website to be…	1=not at all understandable to 7=very understandable	N/A	[37]
Reliability	I considered the information on the website to be…	1=very unreliable to 7=very reliable	N/A	[37]
Credibility	I considered the information on the website to be…	1=very incredible to 7=very credible	N/A	[37]
Usefulness	I considered the information on the website to be…	1=very useless to 7=very useful	N/A	[38]
Readability	I considered the information on the website to be…	1=very unreadable to 7=very readable	N/A	[39]
Sidedness	I considered the information on the website to be…	1=focused on the cons to 7=focused on the pros	N/A	[40]
Enjoyment	I considered the information on the website to be…	1=very unenjoyable to 7=very enjoyable	N/A	[41]
Novelty	The website contained new information for me.	1=strongly disagree to 7=strongly agree	N/A	[39]
Attitude toward website	How good or bad did you find… - the possibility to first answer a question and then receive information? - the possibility to weigh the pros and cons? - the speed of the website? - the layout of the website?	1=very bad to 7=very good	.81 (α)	[42]

[a]N/A: not applicable.

[b]HPV: human papillomavirus.

XSL•FO
RenderX

Table 3. Overview of the program acceptability measures for the virtual assistants.

Measures	Items	Score and scale	Pearson *r*	Reference
Overall rating	On a scale from 0 to 10, how would you rate: - the mother-like assistant - the doctor-like assistant	0=very bad to 10=excellent	0.94	[43]
Enjoyment	I considered the virtual assistants to be…	1=very unenjoyable to 7=very enjoyable	N/A[a]	[41]
Reliability	I considered the virtual assistants to be…	1=very unreliable to 7=very reliable	N/A	[43]
Credibility	I considered the virtual assistants to be…	1=very incredible to 7=very credible	N/A	[37]
Usefulness	I considered the virtual assistants to be…	1=very useless to 7=very useful	N/A	[37]

[a]N/A: not applicable.

Outcomes

Primary Outcome: HPV Vaccination Uptake

An objective measure for HPV vaccination uptake was derived from Praeventis, which is the national electronic vaccination register that monitors the vaccination status for all children up to 18 years of age living in the Netherlands [44]. Uptake was dichotomized into having received no HPV injection (0=not vaccinated) and having received 1 or 2 HPV injections (1=vaccinated), as data analyses showed that determinants of HPV vaccination uptake contrasted the most between these groups.

Secondary Outcomes: IDM, Decisional Conflict, and Social-Psychological Determinants of HPV Vaccination Uptake

Multimedia Appendix 2 provides the measurement details of the constructs IDM, decisional conflict, and the social-psychological determinants of HPV vaccination uptake [30,45-54]. Determinants accounted for were intention, attitude, risk perception of having received the HPV vaccine, risk perception without having received the HPV vaccine, anticipated regret about rejecting the HPV vaccine, beliefs, subjective norms, habit strength toward HPV vaccination, self-efficacy, knowledge, and perceived relative effectiveness of HPV vaccination compared to alternative methods.

Sociodemographics

Sociodemographics were included as background variables (ie, age, educational level, country of birth, and religion).

Level of education referred to the mothers' highest completed level of education. It was classified as low (less than secondary or vocational education), intermediate (secondary through preuniversity education), or high (professional or university education) [30,49,50].

Country of birth was dichotomized into "Netherlands" vs "other," because the percentage of mothers born in the latter category (276/3995, 6.91%) was too small for further subdivision in our sample.

Religion was measured by asking mothers about their religious convictions (Protestant, Roman Catholic, Muslim, Jewish, Buddhist, Hindu, other, or no religion). This was dichotomized into "Protestant" vs "not Protestant" as earlier research and data analyses showed that Protestants more frequently refrain from vaccination compared to other religious groups [30,49,50].

Statistical Analyses

The sample and intervention use data are summarized using descriptive statistics. For determining the dose-response relation, intention-to-treat (ITT) analyses were used, which increases power, while decreasing the risk of possible bias caused by selective dropout [55]. Missing data were imputed for uptake, determinants, and sociodemographics (not for use): we applied multiple imputation by chained equations [55,56]. In total, 15 imputed datasets were generated using the predictive mean matching algorithm in Statistical Package for the Social Sciences (SPSS, IBM Corp). Results from these datasets were pooled together using Rubin's rules [57]. Iteration plots were inspected to check convergence of the imputations.

The relationship between acceptability and use was examined using univariate regression analyses. The dose-response relation between program use and intervention effects was examined by logistic and linear regression analyses (for dichotomous and continuous variables, respectively) using the outcome score at follow up as the dependent variable and the outcome scores at baseline and program use (both completeness and time of website use) as the independent variables [58]. Effects were considered significant when the *P* value was lower than .003 (Bonferroni-corrected α=.05/15 factors). For the logistic regressions, the odds ratio (OR) was used as an index of effect sizes. These were interpreted as small (OR=1.5), medium (OR=3.5), or large (OR=9.0) [59]. For the linear regressions, effect sizes were calculated in R software (R Development Core Team, Vienna, Austria) using Cohen f^2 statistic, $(R^2AB - R^2A)/(1 - R^2AB)$, in which *B* is the variable of interest (ie, use), *A* is the set of all other variables (ie, the outcome at baseline), R^2AB is the proportion of variance accounting for *A* and *B* together, and R^2A is the proportion of variance accounted for by *A* [60]. These were classified as small (f^2=0.02), medium (f^2=0.15), and large (f^2=0.35) [59]. We performed complete case analyses as a sensitivity check for substantial differences with the results from ITT.

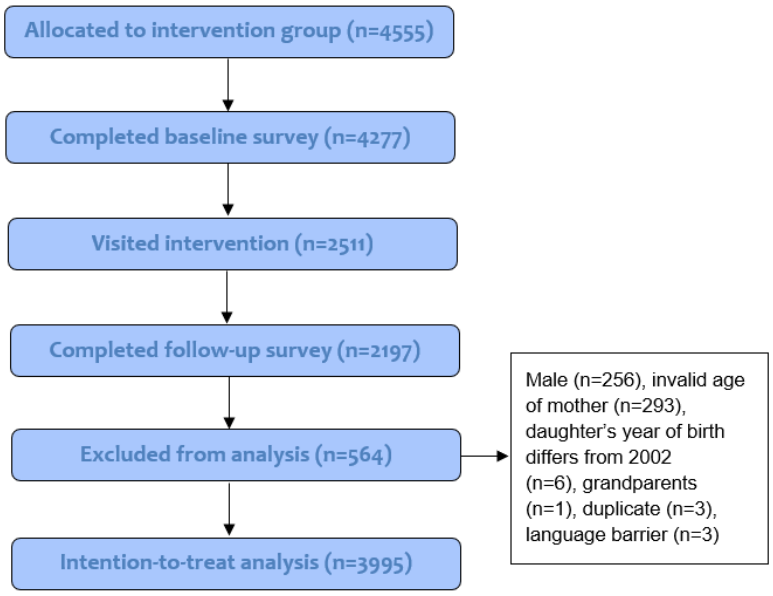
XSL•FO
RenderX

Results

Response Rates and Attrition

A flow diagram of the recruitment and response of participants in the experimental condition, as part of the process evaluation, is shown in Figure 1. From the 4555 participants who were randomized into the experimental condition at baseline (T0), 4277 (93.90%) completed the baseline questionnaire, 2511 (55.13%) visited the intervention, and 2197 (48.23%) completed the follow-up questionnaire 8 weeks later (T1). Dropout analysis showed that there was significantly more dropout with regard to participants not born in the Netherlands, those with a lower educational level, those having a daughter not being vaccinated, and those with low levels of risk perception. In total, 564 participants were excluded as they (a) did not meet the inclusion criteria (ie, being a mother of a daughter born in 2002 and aged 24-62 years), (b) were found to be duplicates, or (c) encountered a language barrier. The final sample for ITT analysis consisted of 3995 mothers.

Figure 1. Flow diagram of the recruitment and response of study participants from the experimental condition. Participants could be excluded based on multiple criteria (eg, male with an invalid age). Therefore, the total number of participants excluded differed from the sum of separate criteria for exclusion.



Sample Description

Table 4 shows the sample description of participants in the experimental condition. No data were available on sociodemographics of the population from which the sample was derived (ie, Dutch mothers of 12-year-old girls); hence, we were unable to assess whether the study sample was representative. The mothers' mean age was 43.70 years (range 27.00-62.00 years). On average, mothers had a positive intention toward their daughters' HPV vaccination at baseline (mean score 5.35, SD 1.69). Compared with the national HPV vaccination uptake at the time (ie, 2015), uptake was higher in the study sample (59,866/93,173; 60.98% vs 2923/3995; 73.17%) [61].

XSL•FO
RenderX

Table 4. Sample description of mothers in the intervention group (N=3995).

Variables	N (% missing)	Value[a]
Age in years, mean (SD)	3995 (0)	43.70 (4.27)
Country of birth, n (%)	3991 (0.10)	
The Netherlands		3715 (92.99)
Other		276 (6.91)
Religion, n (%)	3998 (0.18)	
Protestant		753 (18.85)
Not Protestant		3235 (80.98)
Educational level, n (%)	3991 (0.10)	
Low		588 (14.72)
Middle		1736 (43.45)
High		1660 (41.55)
HPV[b] vaccination uptake, n (%)	3986 (0.23)	
Yes		2923 (73.17)
No		1063 (26.60)

[a]By reporting 2 decimal points for the percentages, summing the percentages for each category may differ from 100%.

[b]HPV: human papillomavirus.

Program Use

From the 3995 mothers in the intervention group, a total of 2509 mothers (62.80%) logged in (ie, time of website use >0). Of these, 73.06% (1833/2509) logged in once, 19.89% (499/2509) logged in twice, 5.00% (125/2509) logged in three times, and 2.07% (52/2509) logged in four times or more. On average, mothers spent 21.39 minutes (SD 12.41) on the website.

Of the 2509 mothers that had initially logged in, 2239 (89.24%) visited at least one page of a component of the intervention (ie, completeness >0). Only these 2239 mothers were included for describing use of specific intervention components and program acceptability (see below). On average, the mothers completed 50.04% (SD 26.18%) of the website components (ie, 51/101 pages).

Use of Intervention Components

Table 5 shows the breakdown for the use of intervention components (N=2239). When mothers clicked on one or more links to receive in-depth information, it was marked as "visited." Mothers visited the page "Ways to Protect Against Cervical Cancer" the most frequently (1971/2239, 88.03%), followed by "Chance" (1945/2239, 86.87%), whereas they visited "Value Clarification" the least (293/2239, 12.53%). "General Information" had the highest completion rate (1622/2239, 72.44%), whereas "Side Effects" and "Effectivity" were completed the least (both 19/2239, 0.85%). Only a small percentage of the mothers (4.20%-13.22%; n=94-298) visited in-depth information (comprising educational movies or extra information).

XSL•FO
RenderX

Table 5. Program use of the various intervention components among those who visited at least one page of the intervention (N=2239).

Intervention component	Completed[a], n (%)[b]	Partly completed[c], n (%)	Not visited[d], n (%)	Visited in-depth information[e], n (%)
General Information	1622 (72.44)	71 (3.17)	546 (24.39)	N/A[f]
Ways to Protect against Cervical Cancer	130 (5.81)	1841 (82.22)	268 (11.97)	N/A
Chance	135 (6.03)	1810 (80.84)	294 (13.13)	142 (6.34)
From HPV[g] to Cervical Cancer	96 (4.29)	1110 (49.58)	1033 (46.14)	298 (13.22)
Age	154 (6.88)	1568 (70.03)	517 (23.09)	160 (7.15)
Side Effects	19 (0.85)	1686 (75.30)	534 (23.85)	299 (13.35)
Effectivity	19 (0.85)	1542 (68.87)	678 (30.28)	339 (15.14)
Other Mothers	1099 (49.08)	416 (18.58)	724 (32.34)	N/A
Working Mechanisms Vaccination	94 (4.20)	1015 (45.33)	1130 (50.47)	94 (4.20)
Facts and Stories	95 (4.24)	1220 (54.49)	924 (41.27)	99 (4.42)
Weighing Pros and Cons	615 (27.47)	911 (40.69)	713 (31.84)	N/A
Value Clarification	269 (12.01)	24 (1.07)	1946 (86.91)	N/A
Practical Information	556 (24.83)	1147 (51.23)	536 (23.94)	N/A

[a]A component was considered "completed" when mothers visited every page of the component.

[b]By reporting 2 decimal points for the percentages, summing the percentages for each category may differ from 100%.

[c]"Partly completed" indicates that the mothers had seen at least one, but not all pages.

[d]"Not visited" means that the mothers had not visited any of the component's pages.

[e]"Visited in-depth information" was also part of the total completeness, but is depicted separately here to provide a better overall view of the mothers' interest in this information. When one or more of the links to more in-depth information had been clicked on, the in-depth information was marked as "visited."

[f]N/A: not applicable as the component did not contain in-depth information.

[g]HPV: human papillomavirus.

Program Acceptability

The overall acceptability of the intervention was rated 7.64 (SD 1.39) on a 10-point scale (1513/2239, 67.57%). Acceptability was significantly and positively associated with completeness (β=4.36, *P*<.001); this did not account for time of website use (β=–.07, *P*=.77). The virtual assistants were rated 7.41 (SD 1.57) on a 10-point scale (1513/2239, 67.57%). The mean scores on all other program acceptability measures were moderate to high (range 4.34-5.71 on a 7-point Likert scale; see Table 6). As the data were not normally distributed but were positively skewed, the median scores are also included in Table 6.

XSL•FO
RenderX

Table 6. Mean (SD) and median scores on the program acceptability measures among those who visited at least one page of the intervention (N=2239)[a].

Variable	Website ($n_{missing}$=726-734)		Virtual assistants ($n_{missing}$=960)	
	Mean (SD)	Median	Mean (SD)	Median
Rate (1-10)	7.64 (1.39)	8.00	7.41 (1.57)	8.00
Interest (1-7)	5.41 (1.05)	5.50	N/A[b]	N/A
Informative (1-7)	5.67 (1.04)	6.00	N/A	N/A
Attitude toward website (1-7)	5.53 (.89)	6.00	N/A	N/A
Perceived user control (1-7)	5.26 (.97)	5.50	N/A	N/A
Elaboration (1-7)	5.70 (1.14)	6.00	N/A	N/A
Tailoring (1-7)	5.22 (1.21)	5.22	N/A	N/A
Comprehensibility (1-7)	5.40 (1.70)	6.00	N/A	N/A
Reliability (1-7)	5.27 (1.23)	6.00	5.04 (1.09)	5.00
Credibility (1-7)	5.35 (1.21)	6.00	5.01 (1.22)	5.00
Usefulness (1-7)	5.52 (1.10)	6.00	4.83 (1.33)	5.00
Readability (1-7)	5.71 (1.04)	6.00	N/A	N/A
Sidedness (1-7)	4.74 (1.09)	4.00	N/A	N/A
Enjoyment (1-7)	4.70 (1.08)	5.00	4.72 (1.37)	5.00
Novelty (1-7)	5.11 (1.39)	5.00	N/A	N/A
Support (1-7)	4.34 (1.80)	5.00	N/A	N/A
Recall (1-7)	5.16 (1.22)	5.00	N/A	N/A
Personal relevance (1-7)	4.87 (1.41)	5.00	N/A	N/A

[a]A higher score represents a higher score on the program acceptability measure.

[b]N/A: not applicable.

Dose-Response Relation Between Program Use and Intervention Outcomes

Table 7 provides an overview of the effects of program use (completeness and time of website use) on the primary and secondary outcomes according to the ITT analyses. Significant positive effects were found for completeness on all outcome measures (all P<.003). The more mothers had completed the intervention, the more likely they were to have their daughter vaccinated against HPV and make an informed decision, and were less likely to experience decisional conflict at follow up. In addition, positive effects of completeness were found with respect to all social-psychological determinants of HPV vaccination acceptability (eg, HPV vaccination intention, beliefs; see Table 7).

Time of website use had a positive effect on all outcomes (all P<.003), except for HPV vaccination uptake, risk perception when not vaccinated, subjective norms, and habit (Table 7). The more time mothers spent on the intervention, the more likely they were to make an informed decision, experience less decisional conflict, have a higher intention to vaccinate their daughter, have a more positive attitude, have more positive beliefs, have a lower risk perception of HPV vaccination, anticipate more feelings of regret about rejecting the HPV vaccine, report a higher relative effectiveness of the HPV vaccine, have higher self-efficacy expectations, and have more knowledge at follow up.

Effect sizes were small overall (see Table 7). Results from the complete case analyses were similar for both completeness and time of website use, except for a lack of effect of time of website use on intention.

XSL•FO
RenderX

Table 7. Effects of use on the outcome measures according to intention-to-treat analyses (N=3995).

Outcome[a]	Pretest, mean (SD)	Posttest, mean (SD)	Completeness		Time of website use	
			β (SEM)	Cohen f^2 or OR[b]	β (SEM)	Cohen f^2 or OR
Primary outcome: HPV[c] vaccination uptake[d]						
Has received no HPV injection (reference)	N/A[e]	26.67% (1066)	N/A	N/A	N/A	N/A
Has received one or two HPV injections	N/A	73.32% (2929)	.004 (.001)[f]	1.004	.003 (.003)[g]	1.003
Secondary outcomes						
IDM[h]: Not informed (reference)	67.31% (2689)	42.53% (1699)	N/A	N/A	N/A	N/A
IDM: Informed	32.69% (1306)	57.47% (2296)	.014 (.001)[f]	1.014	.021 (.003)[f]	1.021
IDM: continuous (0-48)	18.69 (11.21)	25.85 (12.30)	.087 (.006)[f]	0.075	.122 (.013)[f]	.027
Decisional conflict (1-7)	4.33 (1.75)	5.38 (1.36)	.007 (.001)[f]	0.043	.011 (.002)[f]	.018
Intention (1-7)	5.35 (1.69)	5.59 (1.87)	.004 (.001)[f]	0.014	.005 (.002)[f]	.004
Attitude (1- 7)	5.18 (1.45)	5.37 (1.51)	.005 (.001)[f]	0.027	.007 (.001)[f]	.010
Beliefs (1-7)	4.19 (.73)	4.47 (.81)	.004 (.000)[f]	0.051	.005 (.001)[f]	.016
Risk perception; not vaccinated (1-7)	3.74 (0.98)	3.77 (1.08)	.002 (.001)[f]	0.003	.002 (.001)[i]	.001
Risk perception; vaccinated (1-7)	2.77 (1.07)	2.64 (1.10)	-.003 (.001)[f]	0.008	–.005 (.001)[f]	.005
Anticipated regret (1-5)	3.71 (1.25)	3.59 (1.31)	.002 (.001)[f]	0.007	.004 (.001)[f]	.004
Subjective Norm (–20-20)	5.88 (7.81)	7.25 (9.20)	.020 (.004)[f]	0.009	.018 (.008)[j]	.002
Habit (1-7)	4.28 (1.78)	4.51 (1.83)	.004 (.001)[f]	0.011	.004 (.002)[k]	.003
Relative effectiveness (1-10)	-1.97 (2.22)	-1.35 (2.27)	.013 (.001)[f]	0.051	.016 (.003)[f]	.015
Self-efficacy (1-7)	6.27 (.73)	6.29 (.75)	.003 (.000)[f]	0.022	.004 (.001)[f]	.007
Knowledge (–8-8)	4.40 (2.14)	5.75 (2.09)	.017 (.001)[f]	0.095	.024 (.002)[f]	.034

[a]A higher score means a higher outcome (eg, more positive attitude) except for decisional conflict in which a higher score means less decisional conflict.

[b]OR: odds ratio; OR>1 means the higher the score on a factor, the higher the outcome of IDM and higher chance of the daughter being vaccinated; OR<1 means the higher the score on a factor, the lower outcome of IDM and lower chance of the daughter being vaccinated.

[c]HPV: human papillomavirus.

[d]HPV vaccination uptake was not assessed at baseline.

[e]N/A: not applicable.

[f]*P*<.003 (significant; Bonferroni: 0.05/15 factors).

[g]*P*=.20.

[h]IDM: informed decision making.

[i]*P*=.14.

[j]*P*=.03.

[k]*P*=.01.

Discussion

Principal Findings

The aim of this process evaluation was to examine (1) program use, (2) program acceptability, and (3) the relationship between program use and acceptability and intervention outcomes for the web-based tailored intervention about HPV vaccination.

Program Use

Almost two thirds (62.80%) of the mothers who were invited to visit the intervention logged in to the intervention. This reach is adequate, and is comparable to that reported for other eHealth interventions (eg, [62]). However, it still leaves room for improvement to increase the reach by employing strategies such as arousing interest in this eHealth intervention. Support for this notion comes from Crutzen et al [63], who showed that arousing interest successfully increased the intention to visit a

XSL•FO
RenderX

website about hepatitis A, B, and C virus and the likelihood of clicking on the link to visit the website. Interest was aroused in the invitation by challenging the potential visitor regarding their knowledge about hepatitis and it was emphasized that the website provided this information in a comprehensible manner. Not only could interest be aroused in the invitation for HPV vaccination for girls and their parents but also, for instance, via other channels by which the target group is reached (eg, the internet, social media). In addition, research has shown a positive recommendation by word of mouth to be an important trigger for visiting a web-based intervention for the first time [19]. Word-of-mouth recommendations could be encouraged by providing "tell-a-friend" services at the web-based intervention [19].

Of the mothers that logged in to the intervention, nearly all visited at least one component. However, a small portion of the mothers (10.76%) did not view any of the intervention's content after having logged in. This is likely due to technical difficulties because at follow up, some mothers (320, 12.75%) indicated that they were not able to see or hear the virtual assistants.

On average, time of website use, indicating the time mothers spent with the intervention, was adequate (21.39 minutes). This is longer than that reported by Brouwer and colleagues [26] in their review of web-based, interactive healthy lifestyle interventions, in which intervention exposure time varied from less than 10 minutes to 10-20 minutes. However, the present exposure time was lower than found by Sanders and colleagues [64] for their interactive, web-based tailored intervention promoting colorectal cancer screening (33 minutes). This difference could be explained by the controlled setting of their study [64], in which participants arrived 60 minutes prior to a medical appointment and viewed the intervention in the waiting room in the presence of a research assistant.

In addition, the mean completeness of the intervention was 50.04%. This indicates that the intervention fits well with the mothers' needs, which is also supported by the results regarding acceptability (see below). The mean completeness in this study is comparable to that found by Watts and colleagues [65], who reported a completeness of 49.09% for a web-based prostate cancer screening decision aid.

Use of Intervention Components

When looking at exposure to intervention components, we found that mothers were most interested in the effectiveness of (alternative) methods to protect against cervical cancer (88.03% visited) and the risks of their daughter getting infected with HPV and developing cervical cancer (86.87% visited). These appear to be essential components of communication about HPV vaccination.

The "Value Clarification" component was visited the least overall (13.09%). This may be attributed to the fact that it was not easily found by participants: this page was only accessible once mothers completed the component "Weighing Pros and Cons." Voncken-Brewster and colleagues [66] also found that participants overlooked a certain part of their intervention, and they successfully improved its visibility by repositioning it. In order to promote exposure to values clarification, we could enhance its visibility.

The "General Information" component was completed the most overall (72.24%). This is likely due to the fact that the mother-like virtual assistant recommended starting with this component when mothers first entered the main menu, and that it was brief (2 pages). "Side Effects" and "Effectivity" were completed the least (both 0.85%). This could be explained by the fact that these two components contained the most links to in-depth information. Only a small proportion of the mothers (≤13.22%) visited in-depth information.

Program Acceptability

The mothers evaluated the website as positive, as shown by the overall acceptability (7.64/10) and the scores on all acceptability measures (mean scores ranging from 4.34 to 5.71 on a 7-point Likert scale). The virtual assistants were also well appreciated (7.41/10). This is likely because it matched with the mothers' preferences for more interactive personalized feedback [40]. This adequate acceptability is similar to the findings of Paiva and colleagues [67] for a web-based tailored intervention to increase HPV vaccination among young adult women (3.27/4.0). We believe the high acceptability in the present study can be attributed to the systematic and user-centered development process of the intervention [7]. The target group was extensively involved throughout the development. Not only did we fine-tune the content of the intervention to the mothers' preferences and requirements but this also was considered in the design of the website, which was chosen by the mothers.

Furthermore, we found a positive relationship between acceptability and completeness (P<.001). This association was also found by others in a web-based intervention for breast cancer [68-70]. This finding underlines the importance of an intervention to be considered appropriate by the target group.

Effects of Program Use on Intervention Outcomes

In line with our expectations, completeness had a significant positive effect on all outcome measures (IDM, decisional conflict, social-psychological determinants of HPV vaccination uptake), including actual HPV vaccination uptake itself. Mothers who had completed more of the intervention were more likely to have their daughter vaccinated against HPV, and had higher levels of IDM and more positive scores on determinants of HPV vaccination acceptance. In particular, the effect on HPV vaccination uptake is an important finding, which was not found when simply contrasting the experimental to the control condition [30]. This stresses the importance of conducting a process analysis alongside such an effect evaluation. These results are very promising, considering the currently low HPV vaccination uptake rates in the Netherlands [6].

We believe that these positive effects can be attributed to the extensive tailoring throughout the intervention. Support for this notion comes from the study results indicating that mothers perceived the intervention to be well-tailored (5.22 on a 7-point scale). Not only did we tailor the content of the intervention to the mothers' personal interest, it was also used to guide the mothers' personal route through the intervention. The latter is likely to have improved the usability of the intervention.

XSL•FO
RenderX

Moreover, the intervention accounted for tailoring on a variety of themes. For example, it considered perceived barriers similar to the approach taken by Gerend and colleagues [9] as well as other beliefs, attitudes, subjective norms, habits, relative effectiveness, anticipated regret, risk perception, self-efficacy, and knowledge.

In contrast to our expectations, time of website use had a positive effect on all outcomes (P<.003), except for risk perception when not vaccinated (P=.14), subjective norms (P=.03), habit (P=.01), and vaccination uptake (P=.20). The latter is also in contrast to findings of a previous study [64]. An explanation for this lack of effects may be the measurement we used for time of website use, which may be less reliable for measuring exposure. We measured the total time spent on the website, but this does not indicate the specific pages the mothers had visited and for how long. In addition, we were unable to determine if there were any timeouts during a session and what the duration of a timeout was, because the intervention was web-based. For instance, they could have been distracted in the home environment during website use, which may have influenced the measured time of website use and may also have caused the lack of effect of acceptability on time of website use. An alternative, more accurate, measure can be found within the domain of education, namely time on task [71]. Future studies are needed to examine whether time on task (instead of total time of website use) has a positive influence on uptake.

Methodological Considerations

There are three methodological considerations. First, as mentioned above, the measurement of total time spent on the website seemed to be not entirely adequate to measure program use. In this study, completeness seemed to be a better indicator of use as it demonstrates the mother's navigation through the website, whereas time of website use did not. In addition, within (almost) every component, mothers were first asked a question about an HPV-related topic. The page was marked as "completed" if they provided an answer, which was necessary to obtain tailored feedback. Within the decisional balance, completeness was calculated based on the mother's answer to each statement. These are more complete indicators of actual use. Nevertheless, it is not an entirely complete indicator for use, since we could not measure whether they read generic information, listened to the tailored feedback from the virtual assistant, or saw a video. This may have influenced the dose-response effects found.

Second, considering the aim of initiating active processing of information, the positive effect found of completeness on habit may seem unwanted. However, at follow up, the mothers had already been exposed to the intervention, and were therefore likely to have actively processed the information. These results may indicate that the more mothers completed the intervention, the less they had to think about getting the HPV vaccine. The intervention may have helped them make a decision rather than inducing perceptions about HPV vaccination as something you take for granted, without thinking. In retrospect, the label "habit" appears to be misleading.

Third, the effect sizes found were small, which is in line with other web-based interventions targeting health behavior outcomes [72]. Despite this, we believe that even small effects are of relevance in public health as they become substantial at the population level. After all, in the Netherlands, approximately 100,000 girls are invited to receive the HPV vaccine on a yearly basis.

Strengths and Limitations

The most important strength of this study is that we, to our knowledge, are the first within the field of HPV vaccination to conduct an extensive process evaluation. According to a review [24], the process evaluation of an intervention promoting HPV vaccination acceptance by tailoring information to participants' perceived barriers, such as that performed by Gerend and colleagues [9], was limited to the evaluation of intervention information (eg, the extent to which the information was informative and convincing). Other notable strengths are the objective measurement of HPV vaccination uptake, the dose-response effect of completeness on actual HPV vaccination uptake, the broad focus on outcomes (ie, determinants of HPV vaccination acceptance, uptake, IDM, and decisional conflict), and the adequate level of both (objectively assessed) program use and acceptability.

However, some limitations need to be addressed. First, there appeared to be technical difficulties despite our best effort to minimize such issues (eg, by extensively pilot testing the website using various devices). For instance, it turned out that a portion of the mothers were unable to see or hear the virtual assistants and that certain components did not work adequately among some of the mothers. Such technical difficulties are likely to have influenced website usability and therefore may have had a negative impact on program use. Second, this eHealth intervention requires mothers to have eHealth literacy [73]. People with lower eHealth literacy have been shown to have a lower educational level and to spend less time online [74,75]. It could be interesting to examine the extent to which eHealth literacy might have influenced use, acceptability, and effects. Considering the latter, we do not expect substantial differences since we did not find any differences in intervention effects among mothers with a low vs high educational level [30]. Third, it was hard to compare the study results to other studies because data on program use are often very poorly reported or even completely lacking (for a review, see [20]). Furthermore, the study was subject to a considerable amount of attrition. This is a common finding in studies on eHealth interventions [20]. Dropout appeared higher, for example, among mothers with a lower educational level, which has also been reported for other web-based tailored interventions [76]. In this study, missing data and selective dropout were handled by using multiple imputation [55,56]. Fortunately, there were nearly no differences between the results from the complete case analyses and the ITT analyses. Therefore, it seems unlikely that the effects found are spurious or due to selective dropout. Finally, caution is needed when generalizing the results of this study to the general population (ie, Dutch mothers of 12-year-old girls) because we were unable to check the representativeness of the sample. Besides, it should be noted that a very homogenous sample was reached. An undiversified reach is a known phenomenon within web-based interventions aimed at health promotion [20]. In our study, the mothers were predominantly born in the Netherlands,

XSL•FO
RenderX

highly educated, and likely to have a daughter that was vaccinated against HPV. However, results from the effect evaluation showed no differences in intervention effects in specific subgroups of participants, as indicated by moderation analyses with sociodemographic variables [30].

Conclusions and Recommendations

This process evaluation has demonstrated that program use and acceptability of the web-based tailored HPV vaccination intervention were adequate. We found a positive association between completeness and acceptability. Furthermore, dose-response effects were found for completeness and time of website use on (nearly) all social-psychological determinants of the mothers' decision making about vaccination, levels of IDM, and decisional conflict. Importantly, the extent to which mothers completed the intervention positively influenced their daughters' actual HPV vaccination uptake. These results indicate that this web-based, tailored intervention fits well with the mothers' needs and has the potential to increase HPV vaccination uptake. Because of the intervention's adequate (dose-response) effects and acceptability, the intervention has recently been incorporated into the national HPV vaccination communication alongside the existing communication materials.

We recommend future interventions promoting HPV vaccination acceptance to (1) incorporate a process evaluation alongside the effect evaluation, (2), include strategies to arouse interest so as to expand reach, and to (3) include time on task. In addition, we recommend investigating the influence of different media types (eg, graphic vs nongraphic presentations) on decision making about HPV vaccination. For instance, Cox and colleagues [77] found that parents who viewed a graphic presentation of HPV-related risk information had a higher HPV vaccination intention compared to parents who viewed a nongraphic presentation. Furthermore, we recommend future research to examine differences in use and acceptability in specific subgroups of participants, for example by conducting moderation analyses with sociodemographics (eg, educational level). Finally, we recommend developing similar intervention models for (a) target groups other than mothers (eg, the daughter or sons in case they will become a next target group for HPV vaccination) and (b) other vaccinations (eg, maternal pertussis vaccination).

Acknowledgments

We would like to thank Petra Oomen (RIVM) and the panels (Veldkamp BV, Intromart GFK, and NGO Flycatcher) for their help with data collection. The Netherlands Organization for Health Research and Development (ZonMw) and Netherlands Organization for Scientific Research (NWO) support this study (grant-number: 50-51515-98-259).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the website.
[DOCX File , 2665 KB - jmir_v22i7e14822_app1.docx]

Multimedia Appendix 2
Measurement details (secondary outcomes).
[DOC File , 121 KB - jmir_v22i7e14822_app2.doc]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Schiffman M, Castle PE. Human papillomavirus: epidemiology and public health. Arch Pathol Lab Med 2003 Aug;127(8):930-934. [doi: 10.1043/1543-2165(2003)127<930:HPEAPH>2.0.CO;2] [Medline: 12873163]
3. Bosch FX, Lorincz A, Muñoz N, Meijer CJLM, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 2002 Apr;55(4):244-265 [FREE Full text] [doi: 10.1136/jcp.55.4.244] [Medline: 11919208]
4. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 2011 Nov 10;29(32):4294-4301 [FREE Full text] [doi: 10.1200/JCO.2011.36.4596] [Medline: 21969503]
5. Gillison ML, Chaturvedi AK, Lowy DR. HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. Cancer 2008 Nov 15;113(Suppl 10):3036-3046. [doi: 10.1002/cncr.23764] [Medline: 18980286]
6. van Lier E, Geraedts J, Oomen P, Giesbers H, van Vliet J, Drijfhout I. Vaccinatiegraad en jaarverslag Rijksvaccinatieprogramma Nederland. 2017. URL: https://www.rivm.nl/bibliotheek/rapporten/2018-0008.pdf [accessed 2019-05-19] [WebCite Cache ID 78UKJSwQr]

XSL•FO
RenderX

7. Pot M, Ruiter RAC, Paulussen TWGM, Heuvelink A, de Melker HE, van Vliet HJA, et al. Systematically Developing a Web-Based Tailored Intervention Promoting HPV-Vaccination Acceptability Among Mothers of Invited Girls Using Intervention Mapping. Front Public Health 2018;6:226. [doi: 10.3389/fpubh.2018.00226] [Medline: 30356852]
8. Bennett AT, Patel DA, Carlos RC, Zochowski MK, Pennewell SM, Chi AM, et al. Human Papillomavirus Vaccine Uptake After a Tailored, Online Educational Intervention for Female University Students: A Randomized Controlled Trial. J Womens Health (Larchmt) 2015 Nov;24(11):950-957 [FREE Full text] [doi: 10.1089/jwh.2015.5251] [Medline: 26488269]
9. Gerend MA, Shepherd MA, Lustria MLA. Increasing human papillomavirus vaccine acceptability by tailoring messages to young adult women's perceived barriers. Sex Transm Dis 2013 May;40(5):401-405. [doi: 10.1097/OLQ.0b013e318283c8a8] [Medline: 23588130]
10. Grandahl M, Rosenblad A, Stenhammar C, Tydén T, Westerling R, Larsson M, et al. School-based intervention for the prevention of HPV among adolescents: a cluster randomised controlled study. BMJ Open 2016 Jan 27;6(1):e009875 [FREE Full text] [doi: 10.1136/bmjopen-2015-009875] [Medline: 26817639]
11. Hopfer S. Effects of a narrative HPV vaccination intervention aimed at reaching college women: a randomized controlled trial. Prev Sci 2012 Apr;13(2):173-182. [doi: 10.1007/s11121-011-0254-1] [Medline: 21993613]
12. Neville LM, O'Hara B, Milat AJ. Computer-tailored dietary behaviour change interventions: a systematic review. Health Educ Res 2009 Aug;24(4):699-720 [FREE Full text] [doi: 10.1093/her/cyp006] [Medline: 19286893]
13. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
14. Moore G, Audrey S, Barker M, Bond L, Bonell C, Cooper C, et al. Process evaluation in complex public health intervention studies: the need for guidance. J Epidemiol Community Health 2014 Feb;68(2):101-102 [FREE Full text] [doi: 10.1136/jech-2013-202869] [Medline: 24022816]
15. Grimshaw JM, Zwarenstein M, Tetroe JM, Godin G, Graham ID, Lemyre L, et al. Looking inside the black box: a theory-based process evaluation alongside a randomised controlled trial of printed educational materials (the Ontario printed educational message, OPEM) to improve referral and prescribing practices in primary care in Ontario, Canada. Implement Sci 2007 Nov 26;2:38 [FREE Full text] [doi: 10.1186/1748-5908-2-38] [Medline: 18039362]
16. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005 Apr;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
17. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007 Nov 30;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
18. Van 't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010 Sep 02;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
19. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Internet-delivered interventions aimed at adolescents: a Delphi study on dissemination and exposure. Health Educ Res 2008 Jun;23(3):427-439. [doi: 10.1093/her/cym094] [Medline: 18209115]
20. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
21. Elfeddali I, Bolman C, Candel MJJM, Wiers RW, de Vries H. Preventing smoking relapse via Web-based computer-tailored feedback: a randomized controlled trial. J Med Internet Res 2012 Aug 20;14(4):e109 [FREE Full text] [doi: 10.2196/jmir.2057] [Medline: 22903145]
22. Rabius V, Pike KJ, Wiatrek D, McAlister AL. Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 2008 Nov 21;10(5):e45 [FREE Full text] [doi: 10.2196/jmir.1008] [Medline: 19033146]
23. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
24. Priest HM, Knowlden AP. Systematic review of primary prevention human papillomavirus interventions targeting college students. Int J Sex Health 2014 Aug 14;27(2):125-144. [doi: 10.1080/19317611.2014.945631]
25. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
26. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
27. Ludden GD, van Rompay TJL, Kelders SM, van Gemert-Pijnen JEWC. How to Increase Reach and Adherence of Web-Based Interventions: A Design Research Viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]

XSL•FO
RenderX

28. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
29. Miller L, Appleby P, Christensen J, Godoy C, Si M, Corsbie-Massay C. Virtual interactive interventions for reducing risky sex: adaptations, integrations, innovations. In: Noar SM, Harrington NG, editors. eHealth applications: promising strategies for behavior change. New York, NY: Routledge; 2012.
30. Pot M, Paulussen TG, Ruiter RA, Eekhout I, de Melker HE, Spoelstra ME, et al. Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of Invited Girls: Randomized Controlled Trial. J Med Internet Res 2017 Sep 06;19(9):e312 [FREE Full text] [doi: 10.2196/jmir.7449] [Medline: 28877862]
31. Bartholomew Eldredge LK, Markham CM, Ruiter RAC, Fernandez ME, Kok G, Parcel GS. Planning health promotion programs: An Intervention Mapping approach. 4th ed. Hoboken, NJ: Wiley; 2016.
32. van Keulen HM, Mesters I, Brug J, Ausems M, Campbell M, Resnicow K, et al. Vitalum study design: RCT evaluating the efficacy of tailored print communication and telephone motivational interviewing on multiple health behaviors. BMC Public Health 2008 Jun 19;8:216 [FREE Full text] [doi: 10.1186/1471-2458-8-216] [Medline: 18565222]
33. Turner SA, Silvia PJ. Must interesting things be pleasant? A test of competing appraisal structures. Emotion 2006 Nov;6(4):670-674. [doi: 10.1037/1528-3542.6.4.670] [Medline: 17144758]
34. Pauwels E, Van Hoof E, Charlier C, Lechner L, De Bourdeaudhuij I. Design and process evaluation of an informative website tailored to breast cancer survivors' and intimate partners' post-treatment care needs. BMC Res Notes 2012 Oct 03;5:548 [FREE Full text] [doi: 10.1186/1756-0500-5-548] [Medline: 23034161]
35. Liu Y. Developing a scale to measure the interactivity of websites. J Advert Res 2003 Jun 01;43(2):207-216. [doi: 10.2501/jar-43-2-207-216]
36. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Person Soc Psychol 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
37. Friederichs SAH, Oenema A, Bolman C, Guyaux J, Van Keulen HM, Lechner L. Motivational interviewing in a web-based physical activity intervention: questions and reflections. Health Promot Int 2015 Sep;30(3):803-815. [doi: 10.1093/heapro/dat069] [Medline: 24101160]
38. Samoocha D, Snels IAK, Bruinvels DJ, Anema JR, van der Beek AJ. Effectiveness of an interactive website aimed at empowerment of disability benefit claimants: results of a pragmatic randomized controlled trial. J Occup Rehabil 2011 Sep;21(3):410-420 [FREE Full text] [doi: 10.1007/s10926-010-9283-z] [Medline: 21258848]
39. Mevissen FEF, Meertens RM, Ruiter RAC, Feenstra H, Schaalma HP. HIV/STI risk communication: the effects of scenario-based risk information and frequency-based risk information on perceived susceptibility to chlamydia and HIV. J Health Psychol 2009 Jan;14(1):78-87. [doi: 10.1177/1359105308097948] [Medline: 19129340]
40. Van Keulen H, Fekkes M, Otten W, Van DPS, Kocken P, Ruiter R. Onderzoek naar de HPV-vaccinatiebereidheid bij moeders en dochters naar aanleiding van de inhaalcampagne in Nederland. Leiden: TNO; 2010.
41. Crutzen R, Cyr D, Larios H, Ruiter RAC, de Vries NK. Social presence and use of internet-delivered interventions: a multi-method approach. PLoS One 2013;8(2):e57067 [FREE Full text] [doi: 10.1371/journal.pone.0057067] [Medline: 23437310]
42. Spook J, Paulussen T, Kok G, van Empelen P. Evaluation of a Serious Self-Regulation Game Intervention for Overweight-Related Behaviors ("Balance It"): A Pilot Study. J Med Internet Res 2016 Sep 26;18(9):e225 [FREE Full text] [doi: 10.2196/jmir.4964] [Medline: 27670222]
43. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
44. van Lier A, Oomen P, de Hoogh P, Drijfhout I, Elsinghorst B, Kemmeren J, et al. Præventis, the immunisation register of the Netherlands: a tool to evaluate the National Immunisation Programme. Euro Surveill 2012 Apr 26;17(17):20153 [FREE Full text] [doi: 10.2807/ese.17.17.20153-en] [Medline: 22551495]
45. van der Pal S, Otten W, Detmar S. Evaluatie van de voorlichting aan ouders over de hielprik. Tijds Gezondheids Wetenschappen 2012 Jun 19;88(8):449-453. [doi: 10.1007/s12508-010-0679-7]
46. Michie S, Dormandy E, Marteau TM. The multi-dimensional measure of informed choice: a validation study. Patient Educ Couns 2002 Sep;48(1):87-91. [doi: 10.1016/s0738-3991(02)00089-7] [Medline: 12220754]
47. Marteau TM, Dormandy E, Michie S. A measure of informed choice. Health Expect 2001 Jun;4(2):99-108 [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00140.x] [Medline: 11359540]
48. O'Connor AM, Jacobsen MJ, Stacey D. An evidence-based approach to managing women's decisional conflict. J Obstet Gynecol Neonatal Nurs 2002;31(5):570-581. [doi: 10.1111/j.1552-6909.2002.tb00083.x] [Medline: 12353737]
49. van Keulen HM, Otten W, Ruiter RAC, Fekkes M, van Steenbergen J, Dusseldorp E, et al. Determinants of HPV vaccination intentions among Dutch girls and their mothers: a cross-sectional study. BMC Public Health 2013 Feb 06;13:111 [FREE Full text] [doi: 10.1186/1471-2458-13-111] [Medline: 23388344]

XSL•FO
RenderX

50. Pot M, van Keulen HM, Ruiter RAC, Eekhout I, Mollema L, Paulussen TWGM. Motivational and contextual determinants of HPV-vaccination uptake: A longitudinal study among mothers of girls invited for the HPV-vaccination. Prev Med 2017 Jul;100:41-49 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.005] [Medline: 28389328]
51. Paulussen T, Lanting C, Buijs G, Hirasing R. Ouders over het Rijksvaccinatieprogramma: Tevredenheid en vaccinatiebereidheid van ouders van jonge kinderen in Nederland Parents about the governments immunization program: Satisfaction and vaccine intention of Dutch parents of toddlers. Leiden: TNO; 2000.
52. Reiter PL, Brewer NT, Gottlieb SL, McRee A, Smith JS. Parents' health beliefs and HPV vaccination of their adolescent daughters. Soc Sci Med 2009 Aug;69(3):475-480. [doi: 10.1016/j.socscimed.2009.05.024] [Medline: 19540642]
53. Gerend MA, Weibley E, Bland H. Parental response to human papillomavirus vaccine availability: uptake and intentions. J Adolesc Health 2009 Nov;45(5):528-531. [doi: 10.1016/j.jadohealth.2009.02.006] [Medline: 19837361]
54. Verplanken B, Orbell S. Reflections on Past Behavior: A Self-Report Index of Habit Strength1. J Appl Social Pyschol 2003 Jun;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
55. Van Buuren S. Flexible Imputation of Missing Data. Boca Raton: Chapman Hall/CRC; 2012.
56. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
57. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
58. Van Breukelen GJP. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
59. Cohen J. Statistical power analysis for the behavioral science. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988.
60. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A Practical Guide to Calculating Cohen's f(2), a Measure of Local Effect Size, from PROC MIXED. Front Psychol 2012;3:111. [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
61. van Lier E, Oomen P, Giesbers H, van Vliet J, Drijfhout I, Zonnenberg-Hoff I. Vaccinatiegraad Rijksvaccinatieprogramma Nederland: verslagjaar. 2016. URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0064.pdf [accessed 2019-04-29] [WebCite Cache ID 77zmB1UMu]
62. Kelders SM, Van Gemert-Pijnen JEWC, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
63. Crutzen R, Ruiter RA, de Vries NK. Can interest and enjoyment help to increase use of Internet-delivered interventions? Psychol Health 2014 May 28;29(11):1227-1244. [doi: 10.1080/08870446.2014.921300] [Medline: 24798208]
64. Sanders M, Fiscella K, Veazie P, Dolan JG, Jerant A. Does patient time spent viewing computer-tailored colorectal cancer screening materials predict patient-reported discussion of screening with providers? Health Educ Res 2016 Aug;31(4):555-562 [FREE Full text] [doi: 10.1093/her/cyw032] [Medline: 27343254]
65. Watts KJ, Meiser B, Wakefield CE, Barratt AL, Howard K, Cheah BC, et al. Online prostate cancer screening decision aid for at-risk men: a randomized trial. Health Psychol 2014 Sep;33(9):986-997. [doi: 10.1037/a0034405] [Medline: 24274808]
66. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013 Jan 16;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
67. Paiva AL, Lipschitz JM, Fernandez AC, Redding CA, Prochaska JO. Evaluation of the acceptability and feasibility of a computer-tailored intervention to increase human papillomavirus vaccination among young adult women. J Am Coll Health 2014;62(1):32-38. [doi: 10.1080/07448481.2013.843534] [Medline: 24313694]
68. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MFM, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. J Med Internet Res 2013 Aug 19;15(8):e170 [FREE Full text] [doi: 10.2196/jmir.2566] [Medline: 23958584]
69. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an Internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012 Jul;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]
70. De Cocker K, Charlier C, Van Hoof E, Pauwels E, Lechner L, Bourgois J, et al. Development and usability of a computer-tailored pedometer-based physical activity advice for breast cancer survivors. Eur J Cancer Care (Engl) 2015 Sep;24(5):673-682. [doi: 10.1111/ecc.12225] [Medline: 25132146]
71. Kovanovic´ V, Gas˘evic´ D, Dawson S, Joksimovic´ S, Baker RS, Hatala M. Does time-on-task estimation matter? Implications for the validity of learning analytics findings. J Learn Analyt 2015;2(3):81-110. [doi: 10.18608/jla.2015.23.6]
72. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
73. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]

XSL•FO
RenderX

74. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a Population With Moderate-to-High Cardiovascular Risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
75. Xesfingi S, Vozikis A. eHealth Literacy: In the Quest of the Contributing Factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
76. Reinwand DA, Crutzen R, Elfeddali I, Schneider F, Schulz DN, Smit ES, et al. Impact of Educational Level on Study Attrition and Evaluation of Web-Based Computer-Tailored Interventions: Results From Seven Randomized Controlled Trials. J Med Internet Res 2015 Oct 07;17(10):e228 [FREE Full text] [doi: 10.2196/jmir.4941] [Medline: 26446779]
77. Cox DS, Cox AD, Sturm L, Zimet G. Behavioral interventions to increase HPV vaccination acceptability among mothers of young girls. Health Psychol 2010 Jan;29(1):29-39. [doi: 10.1037/a0016942] [Medline: 20063933]

Abbreviations

eHealth: electronic health
HPV: human papilloma virus
IDM: informed decision making
ITT: intention-to-treat
OR: odds ratio

Edited by G Eysenbach; submitted 26.05.19; peer-reviewed by A Peine, R Mpofu; comments to author 19.07.19; revised version received 04.11.19; accepted 14.05.20; published 17.07.20.

Please cite as:
Pot M, Paulussen TGWM, Ruiter RAC, Mollema L, Hofstra M, Van Keulen HM
Dose-Response Relationship of a Web-Based Tailored Intervention Promoting Human Papillomavirus Vaccination: Process Evaluation of a Randomized Controlled Trial
J Med Internet Res 2020;22(7):e14822
URL: https://www.jmir.org/2020/7/e14822
doi:10.2196/14822
PMID:32520718

©Mirjam Pot, Theo G W M Paulussen, Robert A C Ruiter, Liesbeth Mollema, Miranda Hofstra, Hilde M Van Keulen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Implementation Strategies for Web-Based Apps for Screening: Scoping Review

Chor Yau Ooi[1,2*], MMed, MD; Chirk Jenn Ng[1*], MBBS, MMed, PhD; Anne E Sales[3,4*], RN, PhD; Hooi Min Lim[1*], MBBS, MMed

[1]University of Malaya eHealth Initiative, Department of Primary Care Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[2]Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

[3]Center for Clinical Management Research, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, MI, United States

[4]University of Michigan Medical School, Ann Arbor, MI, United States

[*]all authors contributed equally

Corresponding Author:
Chirk Jenn Ng, MBBS, MMed, PhD
University of Malaya eHealth Initiative, Department of Primary Care Medicine
Faculty of Medicine
University of Malaya
Kuala Lumpur,
Malaysia
Phone: 60 79492306
Email: ngcj@um.edu.my

Abstract

Background: Screening is an effective primary prevention strategy in health care, as it enables the early detection of diseases. However, the uptake of such screening remains low. Different delivery methods for screening have been developed and found to be effective in increasing the uptake of screening, including the use of web-based apps. Studies have shown that web-based apps for screening are effective in increasing the uptake of health screening among the general population. However, not much is known about the effective implementation of such web-based apps in the real-world setting. Implementation strategies are theory-based methods or techniques used to enhance the adoption, implementation, and sustainability of evidence-based interventions. Implementation strategies are important, as they allow us to understand how to implement an evidence-based intervention. Therefore, a scoping review to identify the various implementation strategies for web-based apps for screening is warranted.

Objective: This scoping review aims to identify (1) strategies used to implement web-based apps for health screening, (2) frameworks used for implementing web-based apps for health screening, (3) outcome measures of implementation strategies, and (4) effective implementation strategies.

Methods: This scoping review was conducted based on Arksey and O'Malley's framework. After identifying the review question, two researchers independently screened and selected relevant literature from PubMed, Embase, Cochrane, Cumulative Index of Nursing and Allied Health Literature, PsycINFO, International Standard Randomised Controlled Trial Number Registry, OpenGrey, ClinicalTrials.gov, World Health Organization International Clinical Trials Registry Platform, and Web of Science. This was followed by charting the data using a standardized form. Finally, we collated, summarized, and reported the results quantitatively and qualitatively based on the review objectives.

Results: A total of 16,476 studies were retrieved, of which 5669 were duplicates. From a total of 10,807 studies, 10,784 studies were excluded based on their titles and abstracts. There were 23 full-text articles reviewed, and 4 articles were included in the final analysis. Many studies were excluded because they focused on the effectiveness and not on the implementation of web-based apps. Facilitation was the most cited implementation strategy used, followed by reminders, clinical champions, and educational meetings and materials. Only 2 studies used implementation frameworks to guide the evaluation of their studies. Common outcome measures for implementation strategies were feasibility, fidelity, and penetration. Implementation strategies reported to be effective were quality improvement meetings, facilitation, educational meetings, and clinical champions.

XSL•FO
RenderX

Conclusions: There is a dearth of literature on the implementation of web-based apps for health screening. Implementation strategies were developed without any reported use of implementation theories or frameworks in most studies. More research on the development and evaluation of web-based screening app implementations is needed.

(J Med Internet Res 2020;22(7):e15591) doi:10.2196/15591

KEYWORDS

internet; mHealth; eHealth; mass screening; implementation strategies

Introduction

Background

Screening is one strategy for health promotion and disease prevention. Screening is the process of identifying healthy people who may be at an increased risk of a particular disease or health problem [1]. Screening enables early detection of disease so that steps can be taken quickly, when relatively little damage has been done, to prevent the disease from progressing [2].

Prior Research

Many studies have shown screening to be effective in the prevention of diseases. A study in Korea reported that participation in cardiovascular disease (CVD) health screening was associated with lower CVD and all-cause mortality from CVD. There was also an increase in the early detection of CVD and a reduction in health care use and cost [3]. A systematic review of colorectal cancer (CRC) screening using fecal occult blood test showed that it is effective in reducing CRC mortality [4]. However, not all health screening tests are beneficial [5]. Studies have shown, for example, that prostate-specific antigen testing has led to overdiagnosis in up to 40% of the patients screened [6]. Screening tests are not uniformly accurate and are liable to false-positive results and overdiagnosis, which can cause harm to patients [5]. Therefore, care is needed in discussing the pros and cons of screening tests with patients, and to come to shared decision making about screening [5].

Methods of Screening Delivery

However, even when screening is beneficial and effective, its uptake remains suboptimal [7]. One reason is that screening cannot be done using just a single method, a “one size fit all” solution [7]. Different delivery methods must be used to address patient needs. A systematic review looking at improving health screening uptake in men showed that there was a variety of delivery methods for screening, ranging from use of education materials or video decision aids to web-based approaches [8].

Information and communication technology (ICT) has become an important platform to improve health care in the general population [9,10]. ICT use in health is sometimes also known as electronic health (eHealth) [11]. Many forms of eHealth interventions have been developed to improve public health care, in particular, web-based apps [12]; although, the use of electronic health records is another approach using ICT. Our focus in this study was on web-based approaches that do not require the use of an electronic health record, which is often not available, especially in low-resource settings.

A web-based app is defined as a program that is accessed over a network connection using HTTP rather than existing within a device’s memory. Web-based apps often run inside a web browser. However, web-based apps may also be client-based, where a small part of the program is downloaded to a user’s desktop, but processing is done over the internet on an external server [13]. With the proliferation of the internet and the accessibility of Wi-Fi and mobile data, web-based apps have become highly accessible to the general population, even in low-resource settings. The availability of smartphones enables the general population to have easy access to many web-based apps instantaneously from anywhere around the world. Web-based apps addressing different health needs like health screening, health promotion, and health advice have proliferated over the last decade. Many studies have shown that these web-based apps can be effective in changing health behaviors and improving health status [12]. For example, a completed systematic review showed that computer-based interventions are effective in improving sexual health knowledge [14].

Web-based apps for screening use questionnaires to accomplish screening. This screening can include mental and behavioral health as well as risk factor assessment. Examples include screening for depression risk, alcoholic addiction, and smoking. Users (patients) enter their responses based on the questions, and the web app sometimes provides appropriate advice to the patient. Several studies have shown that web-based apps for screening are effective in increasing the uptake of health screening among the population [15-17]. However, how successful implementation of these web-based health apps is accomplished has not yet been studied.

Rationale for Implementation Research

Implementation research involves the study of “the use of strategies to adopt and integrate evidence-based health interventions and change practice patterns within specific settings” [18]. Implementation science approaches use a systematic process of getting an evidence-based intervention (EBI) to reach the target population. Many frameworks, theories, and models are available to guide the implementation process [19]. By using frameworks appropriately to guide the implementation process, including development and design of strategies, *implementation strategies* can be developed to facilitate the uptake of EBIs in a specific health care setting. *Implementation strategies* are activities to enhance the adoption, implementation, and sustainability of EBIs [20]. Explicit implementation increases the likelihood of getting the EBI to the target population compared to leaving it to usual processes. For example, a study done in a clinic found that by attaching a reminder to each patient’s chart about guidelines on the management of dyslipidemia, 94% of patients received the

XSL•FO
RenderX

recommended treatment compared to just 35% in the control group [21].

Nonetheless, studies have shown that many EBIs do not reach their target population [22]. There is an obvious evidence-practice gap. Considerable effort is expended on the discovery of new EBIs but not on examining how these EBIs can be successfully implemented in health care settings [18]. A previous scoping review showed that limited research has been done to implement internet interventions for depression [23]. Therefore, there is a need to increase the amount of implementation to ensure that this evidence-practice gap can be reduced.

Studies show that implementing ICT in health care settings can be complex [23-25]. Systematic reviews demonstrate that organizational factors, setting, integration of the EBI into the workflow, and contextual and societal factors are important influences on the success of implementation [23,24]. However, most studies have focused more on the features of the ICT app and characteristics of end-users than on implementation factors or determinants [24]. Studies of the process of implementation and the factors associated with it are also lacking in the context of ICT in health care [24]. A systematic review demonstrated that studying the implementation process and factors influencing it are important in determining the success of implementation [23]. It is important to evaluate barriers and facilitators to successful implementation of an EBI [24].

Studying *implementation strategies* allows us to understand how to implement a particular EBI [26]. As noted in the literature, developing implementation strategies to overcome barriers to implementation is an important research agenda. Implementation strategies can be critical in bridging the evidence-practice gap when appropriately designed and deployed. Recognizing the importance of implementation strategies, various studies have been done to identify effective implementation strategies for various types of EBI [27]. However, there is a scant of literature on implementation strategies for web-based screening apps. A search of the PROSPERO (International Prospective Register of Systematic Reviews) database of systematic review protocols found no systematic reviews in this area. Therefore, a scoping review to identify the various implementation strategies for web-based screening apps appeared appropriate.

Methods

Overview

This scoping review uses the framework by Arksey and O'Malley [28] and Levac et al [29]. This methodology was selected because our aim is to explore the variation in implementation strategies used to implement web-based health screening apps. The Arksey and O'Malley framework is comprised of six stages: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consultation exercise.

Stage 1: Identifying the Research Question

Among studies exploring the implementation of web-based screening apps:

1. What implementation strategies were used and proven effective?
2. What studies were informed by implementation theories, models, and frameworks, and which ones in particular?
3. What implementation outcomes were measured?
4. What factors were identified as influencing implementation effectiveness?

Stage 2: Identifying Relevant Studies

The literature search was conducted using PubMed, Embase, Cochrane, CINAHL (Cumulated Index to Nursing and Allied Health Literature) via EBSCOHost, PsycINFO via OvidSP, ISRCTN (International Standard Randomised Controlled Trials Number) registry, OpenGrey, ClinicalTrials.gov, World Health Organization International Clinical Trials Registry Platform, and Web of Science for peer-reviewed scientific literature. The search was conducted in August 2018 for 2 weeks and concluded on August 28, 2018. Other search methods used to supplement the literature include reference or footnote tracking, using the "related articles" function in PubMed, citation tracking, personal knowledge and personal contacts, and contacting experts in the field. The search terms and search strategies are included in Multimedia Appendix 1.

Stage 3: Study Selection

The inclusion criteria were studies that described any implementation strategies used to implement web-based apps for screening. Definition of specific terms are outlined as follows:

- Implementation strategies: methods or techniques used to enhance the adoption, implementation, and sustainability of EBIs [20]
- Web-based app: any program accessed over a network connection using HTTP rather than existing within a device's memory [13]
- Frameworks: A framework is a structure, overview, outline, system, or plan consisting of various descriptive categories (eg, concepts, constructs, or variables) and the relations between them that are presumed to account for a phenomenon [19].

The study must also include all age groups and genders; quantitative, qualitative, and mixed-method study designs will be included, and studies must be in the English language.

The exclusion criteria were as follows:

1. Non–web-based apps such as desktop-based, computer-based, CD-ROM interventions, mobile apps, and electronic health records
2. Nonempirical references such as trial protocols, book reviews, editorials, magazine articles, and theoretical or methodological articles
3. Non-English studies

Two independent reviewers screened the titles and the abstracts. The full texts of relevant articles were retrieved and screened

XSL•FO
RenderX

by the two independent reviewers. Any disagreements were discussed among the two reviewers, and a third reviewer was consulted if the two reviewers were unable to resolve any disagreements.

Stage 4: Charting the Data

Data were extracted using a standardized form. Two reviewers independently extracted the data. Any disagreements were resolved with a third reviewer. Data extracted from individual studies include author(s), year of publication, origin or country of origin (where the study was published or conducted), aims or purpose, frameworks, study population and sample size (if applicable), methodology or methods, intervention type, comparator and details of these (eg, duration of the intervention; if applicable), duration of the intervention (if applicable), outcomes and details of these (eg, measures; if applicable), and how the key findings relate to the scoping review question(s). The two reviewers met after extracting data from two articles to determine if the data extraction process was consistent with the research questions and purpose. Any inconsistencies were discussed until a consensus was reached. A third researcher was consulted when the two researchers were unable to reach a consensus.

Stage 5: Collating, Summarizing, and Reporting the Results

The results of this review were divided into three distinct steps based on recommendations by Levac et al [29]: (1) analysis (descriptive numerical summary analysis and qualitative thematic analysis); (2) reporting the results according to the research questions; and (3) considering the meaning of the findings as they relate to the overall study purpose and discuss implications for future research, practice, and policy.

Stage 6: Consultation Exercise

The consultation exercise was conducted among experts in primary care, eHealth, and implementation science. The experts discussed the selection of articles and helped troubleshoot the issues that occurred during the review process. This exercise helped to validate and shape the study outcomes of this review.

Quality Assessment

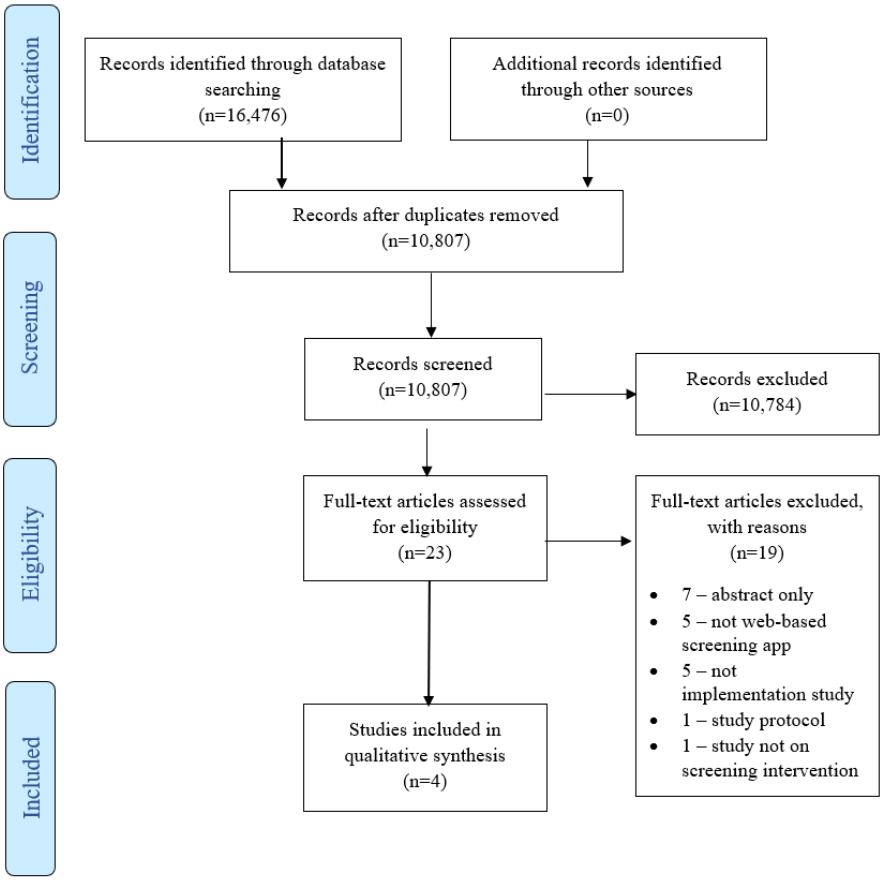
Quality assessment of the studies was done using the Mixed Methods Appraisal Tool (MMAT). MMAT was used as it is a validated tool to quickly assess qualitative, quantitative, and mixed methods studies [30]. This was done to assess the quality of the studies and not used as criteria for the inclusion of studies. The table reporting the quality assessment is included in Multimedia Appendix 2.

Results

Included Studies

The searching databases yielded a total of 16,476 records. After the initial screening and removal of duplicates, 4 studies were included in this review (Figure 1).

Figure 1. Flow diagram of selection of studies.



XSL•FO
RenderX

Most studies were excluded at the title or abstract screening phase because the aim of this review was to identify implementation strategies for web-based apps. Most studies found through the database search evaluated the effectiveness of the web-based apps rather than focusing on their implementation. Only 4 studies fulfilled the inclusion criteria [31-34]. Of these studies, 3 were conducted in primary care settings and 1 in an emergency department. All the studies were descriptive in nature, and the study duration ranged from 2 to 15 months. The characteristics of each study are summarized in Table 1.

Table 1. Characteristics of included studies.

Authors	Year	Intervention	Setting	Study population	Methods	Implementation duration
Webb et al [31]	2018	Health and lifestyle screening tool ("app"), Check Up GP[a] for young people in an Australian general practice	• Australia • General practice	• Patients: young people 14-25 years of age • HCP[b]: GP, practice manager, reception coordinator, and receptionists	• Qualitative: semistructured interviews and focus groups • Quantitative: cross-sectional survey	2 months
Diez-Canseco et al [32]	2018	A web-based mental health screening app	• Peru • Primary care	• Patients: adults ≥18 years of age • HCP: midwives, nurses, and nurse assistants	• Mixed methods • Qualitative: face-to-face structured interviews with HCPs and patients • Quantitative: web-based data collection platform	9 weeks
Krist et al [33]	2014	MOHR[c], a web-based health risk assessment tool	• United States • Primary care	• Patients: adults • HCP: clinicians, front desk staff, practice rooming staff, and medical assistants	• Cluster-randomized, mixed methods implementation trial • Qualitative: interviews with HCPs • Quantitative: data from research networks and the MOHR tool	10 months
Scribano et al [34]	2011	Computerized intimate partner violence screening	• United States • Hospital emergency department	• Patients: adults • HCP: emergency department staffs	• Qualitative: direct observation of patient use, feedback from patients, one-on-one feedback from emergency department staff, and team meetings • Quantitative: questionnaire survey through kiosk.	15 months

[a]GP: general practitioner.

[b]HCP: health care provider.

[c]MOHR: My Own Health Report.

Quality Assessment

The single randomized controlled trial included in this review was of moderate quality [33]. The 2 arms were not comparable at baseline, and the outcome assessors were not blinded [33]. The other 3 studies contained both qualitative and quantitative components [31,32,34]. For the qualitative component, 2 studies were of good quality [31,32]. As there was no clear descriptions on how the data were collected, analyzed, and interpreted, 1 study was considered to be of poor quality [34]. For the quantitative component, 2 studies were of moderate quality, as the sample population and the risk of nonresponse bias of the studies were not clearly described [32,34]. The other study was of good quality [31].

Implementation Frameworks and Strategies for Web-Based Screening Apps

Only 2 studies used frameworks to guide and evaluate the process and outcomes of the study. However, in both studies, a framework and theory were used to guide the evaluation of the studies and not on the development of the implementation strategies. Table 2 summarizes the implementation strategies and frameworks used in each study. The types of frameworks were categorized based on Nilsen's [19] taxonomy of implementation frameworks. The implementation strategies are categorized based on the Effective Practice and Organisation of Care taxonomy [35].

XSL•FO
RenderX

Table 2. Implementation frameworks and strategies used for web-based apps.

Authors	Frameworks	Types of framework [19]	Implementation strategies [35]	Implementation activities
Webb et al [31]	NPT[a]	Implementation theories		
			Reminder	Phone call and SMS as reminder to patients to complete the app
			Reminder and facilitation	Receptionist to prompt patient to complete app in the waiting room using tablet
			Continuous quality improvement	Quality improvement meetings with HCPs[b]
			Educational materials	Provision of educational documents to support HCPs
Diez-Canseco et al [32]	None	N/A[c]		
			Leadership engagement/buy-in	Engaging policy makers for support and buy-in
			Educational meetings	Training for HCPs
			Technical assistance	Telephone and face-to-face support, and supervision for HCPs throughout the implementation period
			Coaching	Regular supervision meetings with HCPs to troubleshoot problems encountered throughout the implementation period
Krist et al [33]	RE-AIM[d]	Evaluation framework		
			Clinical champions	Appointment of practice champions
			Educational meetings	Training for HCPs
			Facilitation	Mailed invitations to complete the app
			Facilitation	Kiosk provided in the clinic waiting room to complete the app with help from researcher
			Facilitation	Completion of the app via phone call by researcher
			Facilitation	Completion of the app via tablet with help from either researcher, practice rooming staff, or medical assistant in the clinic waiting room
Scribano et al [34]	None	N/A		
			Patient education and facilitation	Nurses and receptionists provided instruction forms to patients for the screening kiosks
			Environment	Placing screening kiosks in strategic locations
			Clinical champions	Appointment of practice champions

[a]NPT: Normalization Process Theory.

[b]HCP: health care provider.

[c]N/A: not applicable.

[d]RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance.

Facilitation was the most cited strategy used. Other strategies that were cited more than once included reminders, clinical champions, and educational meetings.

Outcome Measures

The implementation outcomes for the studies included in this review were categorized based on the taxonomy of implementation outcomes by Proctor et al [36]. The implementation outcomes can be broadly categorized into acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability. Table 3 outlines the outcome measures reported in the studies.

XSL•FO
RenderX

Table 3. Outcome measures.

Authors, Outcomes measured	Implementation outcome
Webb et al [31]	
Number of times needed to provide support to staff on the use of the app	Feasibility
Location completed Check Up GP[a]	Feasibility
If patients received an SMS with a link to Check Up GP before attending the practice	Fidelity
If patients felt that they had sufficient privacy when completing the app	Fidelity
Postimplementation staff interviews and focus group discussions based on NPT[b]	Acceptability, adoption, appropriateness, feasibility, fidelity, sustainability
Diez-Canseco et al [32]	
Number of screenings	Penetration
Fidelity to screening	Fidelity
Integration of screening into routine clinical service	Sustainability
Krist et al [33]	
Reach: proportion of eligible patients approached who completed a MOHR[c] assessment	Penetration
Adoption: percentage of practices approached for study participation who agreed to participate	Adoption
Implementation: how practices integrated MOHR into their workflow and the time and staff needed to carry out implementation steps	Fidelity
Maintenance: whether early intervention practices continued to use MOHR after the study	Sustainability
Scribano et al [34]	
Number of screenings completed by patients	Penetration
Computer technology failure rate	Feasibility

[a]GP: general practitioner.

[b]NPT: Normalization Process Theory.

[c]MOHR: My Own Health Report.

Effective Implementation Strategies

Different studies revealed different implementation strategies that were reported to be effective in their settings (Textbox 1).

For Webb et al [31], the quality improvement meetings using the Plan-Do-Act-Study framework were reported to be effective. The study also reported that facilitation by researchers on-site helped in the implementation process. Comparing different sites using the web-based app, the authors found that using the web-based app in the waiting room was the most effective compared to other settings (84%) [31].

Textbox 1. Effective implementation strategies.

Health care providers

- Quality improvement meetings
- Facilitation
- Educational meetings
- Clinical champions

Patients

- Environment
- Facilitation

In the study conducted by Diez-Canseco et al [32], the authors stated that training health care providers was the key to promote use of the web-based app. Supervision was another important strategy to troubleshoot issues when implementing the web-based app.

For Krist et al [33], the authors compared implementation strategies used in nine different primary care clinics. They found that the strategy with the highest screening completed was to complete the app in the clinic waiting room with facilitation from the researcher (94.4% completion rate). Overall, they found

XSL•FO
RenderX

that facilitation from staff in the clinic to help the patient complete the web-based app was the key to successful implementation. There was a higher completion rate with facilitation from staff compared to no facilitation (71.2% vs 30.3%) [33].

In the study by Scribano et al [34], appointing nurse champions helped to increase the screening rate. Another strategy that improved the screening rate was moving the kiosk to a more strategic location in the emergency department.

Discussion

Principal Findings

This review revealed that there is a very small set of literature reporting on the implementation of web-based apps for screening. Many studies evaluated the effectiveness or application of web-based screening apps, but very few described how to implement them [10,25]. The lack of literature found in this review suggests that there is a need for more implementation studies to be conducted in this area. Many studies have shown that implementation will not happen on its own, and implementation has to be planned to succeed [37].

Integrating an implementation study with an effectiveness study may be one way forward. Hybrid study designs may yield more information, as both the effectiveness of the intervention and the implementation can be assessed in a single study [38]. This may also reduce the time needed to translate an EBI into practice. However, these study designs require more expertise, resources, and manpower to conduct [38]. Therefore, it may not be feasible to conduct a hybrid study in settings with limited resources.

In this review, the most common implementation strategies used in implementing web-based apps were facilitation, education, and clinical champions. Most of these strategies targeted only patients and health care providers. Only 1 study mentioned strategies related to leadership engagement [32]. The strategies found in this review were limited to smaller scales and local contexts.

There was also no clear discussion on how strategies were developed or designed for implementation. Using a systematic process and determinant frameworks to help guide the development of strategies can be important for successful implementation and evaluation [19]. Having a systematic process for developing strategies will also help others replicate the process if the implementation was successful. We found that only 2 of the 4 studies used implementation frameworks or theories. This is consistent with the implementation research literature [19]. Without a solid theoretical foundation, it is challenging to explain why an implementation fails or succeeds [19]. Currently, many theories and frameworks are available for implementation research [19]. By using a combination of these theories or frameworks, researchers may better understand how to plan an implementation study systematically. By testing strategies and frameworks, new knowledge can be gained for future implementation studies.

In this review, different outcome measures were used in each study. However, there was no clear explanation of how the authors of the studies determined which outcome to measure. Which implementation strategies contributed to the outcomes were not clear because several different strategies were used in most studies, and their individual effects could not be evaluated. Therefore, we were unable to conclude which implementation strategies were useful. The lack of implementation frameworks or theories used in the studies may have contributed to this issue. The process of developing implementation strategies should be done systematically, ideally to link each strategy to different outcomes. If this is feasible, it could allow us to assess which strategies work. This may also allow researchers to further explore why a particular strategy works for a specific outcome. Lack of effect may lie with the implementation strategy itself or may be related to how the implementation strategy was implemented. Therefore, it is crucial that researchers give adequate thought to the implementation outcomes they measure and how an implementation strategy might affect each outcome. In addition, the field of implementation research is shifting from an exclusive focus on the effectiveness of implementation strategies to a focus on how strategies achieve their effect, if any, and on the factors that make implementation, across a spectrum of evidence-based practices, including ICT approaches, either more or less likely to be successful. Future studies will hopefully include this broader focus, as the current literature does not go beyond a simplistic examination of the implementation strategies' effectiveness.

Strengths and Limitations

This review may be limited because only English language studies were included. Another limitation could have been related to screening titles and abstracts. This may have limited the inclusion of studies if implementation strategies were not discussed in the abstract. However, based on our review, we found that most of the implementation strategies could be identified from abstracts alone. To this end, this scoping review revealed that only a few studies addressed the implementation of web-based health screening apps, and therefore, more research is needed in this area.

Conclusion

This scoping review shows that the study of implementation of web-based apps for screening is still in its infancy. Many studies have assessed the effectiveness of web-based apps, but only a few have focused on how to implement them. We were able to identify different implementation strategies for implementing web-based apps for screening. However, there is little evidence that the strategies reported were systematically developed using theories or frameworks. The lack of frameworks and theories used in these studies was also evident. There is a need to study not only the effectiveness of implementation strategies but the process of implementation and how this affects outcomes. This review shows that more work is still needed to study implementation of web-based apps for screening in a systematic process based on implementation theories and frameworks.

XSL•FO
RenderX

Acknowledgments

This study is funded by University of Malaya Specialist Centre C.A.R.E research grant PV-063-2018. The authors would also like to thank Universiti Malaysia Sarawak for financial support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and search strategy for scoping review.
[DOCX File , 17 KB - jmir_v22i7e15591_app1.docx]

Multimedia Appendix 2
Mixed Method Appraisal Tool Table.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v22i7e15591_app2.xlsx]

References

1. NHS population screening explained. Public Health England. URL: https://www.gov.uk/guidance/nhs-population-screening-explained [accessed 2019-03-04]
2. Wilson J, Jungner G. Principles and practice of screening for disease. World Health Organization. 1968. URL: https://apps.who.int/iris/bitstream/handle/10665/37650/WHO_PHP_34.pdf?sequence=17&isAllowed=y
3. Lee H, Cho J, Shin DW, Lee S, Hwang S, Oh J, et al. Association of cardiovascular health screening with mortality, clinical outcomes, and health care cost: a nationwide cohort study. Prev Med 2015 Jan;70:19-25. [doi: 10.1016/j.ypmed.2014.11.007] [Medline: 25445334]
4. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008 Jun;103(6):1541-1549. [Medline: 18479499]
5. Bell NR, Grad R, Dickinson JA, Singh H, Moore AE, Kasperavicius D, et al. Better decision making in preventive health screening: balancing benefits and harms. Can Fam Physician 2017 Jul;63(7):521-524 [FREE Full text] [Medline: 28701439]
6. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, ERSPC Investigators. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014 Dec 06;384(9959):2027-2035 [FREE Full text] [doi: 10.1016/S0140-6736(14)60525-0] [Medline: 25108889]
7. Dryden R, Williams B, McCowan C, Themessl-Huber M. What do we know about who does and does not attend general health checks? Findings from a narrative scoping review. BMC Public Health 2012 Aug 31;12:723 [FREE Full text] [doi: 10.1186/1471-2458-12-723] [Medline: 22938046]
8. Teo CH, Ling CJ, Ng CJ. Improving health screening uptake in men: a systematic review and meta-analysis. Am J Prev Med 2018 Jan;54(1):133-143. [doi: 10.1016/j.amepre.2017.08.028] [Medline: 29254551]
9. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
10. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
11. eHealth. World Health Organization. 2017. URL: https://www.who.int/ehealth/en/ [accessed 2019-02-22]
12. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
13. What is a Web-Based Application? Techopedia. 2018. URL: https://www.techopedia.com/definition/26002/web-based-application [accessed 2019-02-22]
14. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010 Sep 08(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
15. van Bergen JEAM, Fennema JSA, van den Broek IVF, Brouwers EEHG, de Feijter EM, Hoebe CJPA, et al. Rationale, design, and results of the first screening round of a comprehensive, register-based, Chlamydia screening implementation programme in the Netherlands. BMC Infect Dis 2010 Oct 07;10:293 [FREE Full text] [doi: 10.1186/1471-2334-10-293] [Medline: 20925966]
16. Harrington JW, Bai R, Perkins AM. Screening children for autism in an urban clinic using an electronic M-CHAT. Clin Pediatr (Phila) 2013 Jan;52(1):35-41. [doi: 10.1177/0009922812463957] [Medline: 23117237]

XSL•FO
RenderX

17. Brooks BA, Haynes K, Smith J, McFadden T, Robins DL. Implementation of web-based autism screening in an urban clinic. Clin Pediatr (Phila) 2016 Sep;55(10):927-934 [FREE Full text] [doi: 10.1177/0009922815616887] [Medline: 26581361]
18. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: current and future directions. Am J Public Health 2012 Jul;102(7):1274-1281. [doi: 10.2105/AJPH.2012.300755] [Medline: 22594758]
19. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
20. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
21. Stamos TD, Shaltoni H, Girard SA, Parrillo JE, Calvin JE. Effectiveness of chart prompts to improve physician compliance with the National Cholesterol Education Program guidelines. Am J Cardiol 2001 Dec 15;88(12):1420-3, A8. [doi: 10.1016/s0002-9149(01)02125-7] [Medline: 11741565]
22. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. Implement Sci 2012 May 31;7:50 [FREE Full text] [doi: 10.1186/1748-5908-7-50] [Medline: 22651257]
23. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of internet interventions for depression: a scoping review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
24. Abildso CG, Zizzi SJ, Reger-Nash B. Evaluating an insurance-sponsored weight management program with the RE-AIM Model, West Virginia, 2004-2008. Prev Chronic Dis 2010 May;7(3):A46 [FREE Full text] [Medline: 20394685]
25. Christie HL, Bartels SL, Boots LM, Tange HJ, Verhey FR, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interv 2018 Sep;13:51-59 [FREE Full text] [doi: 10.1016/j.invent.2018.07.002] [Medline: 30206519]
26. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 01;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
27. Brouwers MC, De Vito C, Bahirathan L, Carol A, Carroll JC, Cotterchio M, et al. What implementation interventions increase cancer screening rates? A systematic review. Implement Sci 2011 Sep 29;6:111 [FREE Full text] [doi: 10.1186/1748-5908-6-111] [Medline: 21958556]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
29. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
30. Pluye P, Hong QN. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45. [doi: 10.1146/annurev-publhealth-032013-182440] [Medline: 24188053]
31. Webb MJ, Wadley G, Sanci LA. Experiences of general practitioners and practice support staff using a health and lifestyle screening app in primary health care: implementation case study. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e105 [FREE Full text] [doi: 10.2196/mhealth.8778] [Medline: 29691209]
32. Diez-Canseco F, Toyama M, Ipince A, Perez-Leon S, Cavero V, Araya R, et al. Integration of a technology-based mental health screening program into routine practices of primary health care services in Peru (The Allillanchu Project): development and implementation. J Med Internet Res 2018 Mar 15;20(3):e100 [FREE Full text] [doi: 10.2196/jmir.9208] [Medline: 29588272]
33. Krist AH, Phillips SM, Sabo RT, Balasubramanian BA, Heurtin-Roberts S, Ory MG, MOHR Study Group. Adoption, reach, implementation, and maintenance of a behavioral and mental health assessment in primary care. Ann Fam Med 2014;12(6):525-533 [FREE Full text] [doi: 10.1370/afm.1710] [Medline: 25384814]
34. Scribano PV, Stevens J, Marshall J, Gleason E, Kelleher KJ. Feasibility of computerized screening for intimate partner violence in a pediatric emergency department. Pediatr Emerg Care 2011 Aug;27(8):710-716. [doi: 10.1097/PEC.0b013e318226c871] [Medline: 21811196]
35. EPOC taxonomy. Cochrane Effective Practice and Organisation of Care. URL: https://epoc.cochrane.org/epoc-taxonomy [accessed 2019-02-23]
36. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
37. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]
38. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]

XSL•FO
RenderX

Abbreviations

CINAHL: Cumulated Index to Nursing and Allied Health Literature
CRC: colorectal cancer
CVD: cardiovascular disease
EBI: evidence-based intervention
eHealth: electronic health
ICT: information and communication technology
ISRCTN: International Standard Randomised Controlled Trials Number
MMAT: Mixed Methods Appraisal Tool
PROSPERO: International Prospective Register of Systematic Reviews

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by E Hermes, M Barwick, A Ekeland, C Lu; comments to author 03.10.19; revised version received 11.12.19; accepted 14.05.20; published 20.07.20.

Please cite as:
Ooi CY, Ng CJ, Sales AE, Lim HM
Implementation Strategies for Web-Based Apps for Screening: Scoping Review
J Med Internet Res 2020;22(7):e15591
URL: http://www.jmir.org/2020/7/e15591/
doi:10.2196/15591
PMID:32706655

©Chor Yau Ooi, Chirk Jenn Ng, Anne E Sales, Hooi Min Lim. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Internet-Based Self-Management Support After High-Altitude Climate Treatment for Severe Asthma: Randomized Controlled Trial

Thijs Beerthuizen[1], MD; Lucia H Rijssenbeek-Nouwens[2], MD, DPhil; Sophia M van Koppen[1], BSc; Rishi J Khusial[1], BSc; Jiska B Snoeck-Stroband[1], MD, DPhil; Jacob K Sont[1], DPhil

[1]Department of Biomedical Data Sciences, Section Medical Decision Making, Leiden University Medical Center, Leiden, Netherlands
[2]Dutch Asthma Centre Davos, Davos, Switzerland

Corresponding Author:
Jacob K Sont, DPhil
Department of Biomedical Data Sciences
Section Medical Decision Making
Leiden University Medical Center
P.O. Box 9600, 2300 RC, Leiden, The Netherlands
Leiden, 2300 RC
Netherlands
Phone: 31 71 5264578
Email: j.k.sont@lumc.nl

Abstract

Background: In patients with severe asthma, high-altitude climate treatment has been shown to improve asthma control. However, asthma symptoms and limitations may increase after finishing inpatient rehabilitation programs and returning to sea level.

Objective: We assessed the effectiveness of a patient-tailored, internet-based, self-management strategy in addition to usual care after finishing high-altitude climate treatment.

Methods: We performed a randomized controlled trial with a 1-year follow-up in patients from a high-altitude asthma center in Davos, Switzerland. At the end of a 12-week multidisciplinary rehabilitation program, 62 adults with asthma were randomized to receive either internet-based self-management support in addition to usual care (n=33) or usual care only after discharge (n=29). The endpoints were changes in asthma-related quality of life according to the Asthma Quality of Life Questionnaire (AQLQ) (a higher score is better) and asthma control according to the Asthma Control Questionnaire (ACQ) (a lower score is better), with a minimally important difference of 0.5 points for both.

Results: Asthma-related quality of life and asthma control declined over time in the usual care strategy group, whereas there was a slower decline in the internet-based strategy group. For both endpoints, mixed-model analysis showed a significant positive effect in favor of internet-based self-management during follow-up (mean AQLQ score difference 0.39, 95% CI 0.092-0.69; *P*=.01 and ACQ score difference −0.50, 95% CI −0.86 to −0.15; *P*=.006), which was prominent among patients with uncontrolled asthma at discharge (AQLQ score difference 0.59, 95% CI 0.19-0.99; *P*=.003 and ACQ score difference −0.73, 95% CI −1.18 to −0.28; *P*=.002).

Conclusions: Internet-based self-management support was associated with a smaller decline in quality of life and asthma control as compared with usual care, especially in patients with lower asthma control, after completion of high-altitude climate treatment. Internet-based self-management support in adults with severe asthma seems feasible and effective to maintain quality of life and asthma control.

Trial Registration: The trial is registered in the Netherlands Trial Register (NTR1995).

(J Med Internet Res 2020;22(7):e13145) doi:10.2196/13145

KEYWORDS

self-management; pulmonary rehabilitation; severe asthma; quality of life; asthma control; internet; monitoring

XSL•FO
RenderX

Introduction

In patients with severe refractory asthma, standard treatment according to international guidelines [1] is often not sufficient for achieving adequate asthma control [2]. Patients with persistent uncontrolled asthma can be referred to specialized asthma high-altitude clinics [3,4]. The population referred for in-patient pulmonary rehabilitation comprises patients with complex somatic and psychosocial problems [5]. The effectiveness of pulmonary in-patient rehabilitation in severe asthma has been established in a previous study that reported relevant improvements in asthma control, quality of life, and clinical outcomes with a 12-week multidisciplinary treatment program involving environmental trigger avoidance in an alpine climate among patients with severe refractory asthma [6]. Pulmonary rehabilitation should be part of an integrated care process and include self-management support. Changing patient behavior and ensuring maintenance are complex processes and require time. Currently, most self-management programs take between 6 and 12 weeks, and a longer duration is associated with greater improvements in physical and psychological functioning [6]. At present, there are sparse data on whether the benefits are sustained beyond the completion of high-altitude climate treatment, and patients are not always amenable to optimal self-management on returning to sea level in their own environment. A recent prospective uncontrolled observational cohort study suggested reduced exacerbation and sustained improvements in asthma control 12 months after high-altitude climate treatment for severe asthma as compared to the findings before admission, with only a slight relapse after discharge [7,8]. To sustain these improvements in long-term outcomes, appropriate self-management behavior during and after discharge is considered essential [9]. In order to achieve a sustained optimal long-term quality of life, we need a dependable system of coordinated health care intervention and communication, and components that include self-management support. Innovative forms of self-management support, including an online community, monitoring, communication, an action plan, and motivational feedback via the internet, have high potential to improve long-term outcomes.

Several recent studies have shown the feasibility and effectiveness of telemonitoring and self-management support [10-14]. Earlier studies of our group have demonstrated that internet-based self-management support improves quality of life, the number of symptom-free days, and clinical outcomes [15,16] in patients with partly controlled asthma.

However, the long-term effectiveness of sustaining self-management support via the internet in patients with severe asthma, who have completed in-patient pulmonary rehabilitation, has not been determined yet.

We hypothesize that self-management support via an internet-based service in addition to usual care will improve health-related quality of life and asthma control in patients with severe asthma who have completed pulmonary rehabilitation in a specialized high-altitude asthma clinic. We conducted a pragmatic randomized controlled trial to assess the effectiveness of an internet-based self-management support app in patients who completed an in-patient pulmonary rehabilitation program at high altitude. In this trial, the primary outcome was asthma-related quality of life. In addition, we assessed asthma control and evaluated whether the degree of asthma control impacts the changes in quality of life. The process outcomes were engagement with the electronic health (eHealth) intervention and aspects of self-management and health education impact.

Methods

Study and Approval

The PRACTISS (Pulmonary Rehabilitation of Asthma: a Trial of sustained Internet-based Self-management Support) study is a randomized parallel group trial with 12 months of follow-up. The trial was performed from December 2012 through January 2016 in The Netherlands. The study was approved by the Medical Ethics Committee of the Leiden University Medical Center. All participants provided written informed consent upon entering the trial. The trial is registered in the Netherlands Trial Register (NTR1995) and conforms to the Helsinki Declaration.

Patients

Patients with a diagnosis of severe asthma according to the American Thoracic Society criteria were recruited from a specialized asthma clinic. All patients completed pulmonary rehabilitation in the Dutch Asthma Centre Davos (Switzerland). They were referred to the rehabilitation clinic because previous maximal treatment did not lead to adequate asthma control. The exclusion criteria that may interfere with compliance or reliability of the measurements were relevant somatic comorbidity, serious psychological problems, and psychiatric comorbidity assessed by the Hospital Anxiety and Depression Scale [17].

Design

The study was started during a high-altitude treatment program for severe asthma in a specialized asthma clinic. Upon enrolling in the study, approximately 1 week before completing the rehabilitation, patients were randomized via a computer-generated permuted block scheme by an independent researcher to either an internet-based strategy or usual care strategy group, with stratification for age and gender. Patients received explanations about the study procedures during the last days of their rehabilitation program, in addition to the usual preparation prior to discharge. After completing the program and finishing the baseline questionnaires, patients returned home and were treated by their own physicians who had referred them to the rehabilitation program. These physicians were notified about the participation of the patients in the study. Participants were asked to fill out questionnaires at 3, 6, 9, and 12 months.

Intervention

Patients in the internet-based strategy group had access to various tools that supported them to achieve optimal self-management skills. For effective self-management, asthma patients need to understand the condition, the purpose of medication, and the environmental influences. Furthermore, they need to recognize worsening respiratory function and know

XSL•FO
RenderX

when to seek medical attention [18]. The PatientCoach self-management support app was designed to promote attributes attained in the rehabilitation program. The self-management support modules could be customized by the health care professionals, allowing the program to be tailored to the specific needs of each patient. The modules included a written action plan, personalized asthma control questions, a daily asthma control questionnaire with lung function measurement, an actometer (Fitbit Ultra; Fitbit), patient health goals, a calendar, education provided by asthma centers, and a helpdesk. The app did not provide an option to communicate with a care provider.

The individual action plan was written by a physician from the rehabilitation clinic and included patient-tailored alarm signals, physician recommendations, and directives on when to contact health care professionals. Each patient together with the rehabilitation clinic physician formulated a patient-tailored question that served as a red flag for the patient to act upon when answered positively. In the daily asthma control questionnaire, patients could fill out the Asthma Control Questionnaire and the forced expiratory volume in one second (FEV_1) measurement. Patients in the internet-based strategy group were provided with a small personal device (PIKO-1, NSpire Health) to measure the FEV_1. The asthma control cutoff values that patients used to act upon were patient-tailored by the rehabilitation clinic physician. Internet support patients received an actometer (Fitbit) for daily use to measure activity. Each patient had a personal goal for daily steps. Patients formulated personal health goals together with the rehabilitation clinic physician for motivational purposes.

The calendar could be used by patients to list their health-related appointments. Patients received access to the educational material used by the rehabilitation clinic. The helpdesk was accessible for information and communication technology purposes. Several screenshots of the current version of the PatientCoach app are presented in Multimedia Appendix 1. As this was a pragmatic trial, participants were free to use the intervention as often as they liked.

Health-Related Quality of Life

The primary outcome of this trial was health-related quality of life, which was assessed every 3 months using the validated and standardized Asthma Quality of Life Questionnaire (AQLQ) by Juniper et al [19]. A minimally clinically important difference of 0.5 was reported for the overall score [20]. The AQLQ consists of the following four subdomains: activity limitation, symptoms, emotional function, and environmental stimuli, and the overall score is determined. A higher score represents a more favorable outcome.

Clinical Control

Every 3 months, all participants in this study were asked to fill out a questionnaire assessing clinical control (six-question Asthma Control Questionnaire [ACQ6]) as an outcome parameter for study purposes. The ACQ6 questionnaire is a validated questionnaire to measure clinical control in patients with asthma [21]. For the ACQ6, a lower score represents a more favorable outcome.

eHealth User Engagement

User engagement was based on the amount of times a patient logged into the PatientCoach web app, as registered in the log files during the study period of 12 months.

Health Education

The aspects of self-management and impact of health education were measured using the Health Education and Impact Questionnaire (heiQ) [22]. The heiQ is a validated questionnaire consisting of eight independent scales, providing a rich range of information about the value of patient self-management programs. The domains comprise health-directed behavior, positive and active engagement in life, negative affect, self-monitoring and insight, constructive attitudes and approaches, skill and technique acquisition, social integration and support, and health service navigation [23]. Scores range from 1 to 4, with a higher score representing favorable outcomes, except for negative affect, where a lower score represents favorable outcomes.

Statistical Analysis

Power analysis showed that with the use of 36 patients per arm and an SD of changes in the AQLQ score of 0.75, we could expect to detect a minimally important difference of 0.5 points between AQLQ changes in the two arms (alpha=.05, power=0.80). Patients were analyzed according to the intention-to-treat approach. To correct for possible selective nonresponse, missing data were imputed using linear regression modelling. We created 100 sets of imputations using disease severity, age, gender, center, strategy, and available quality of life data as regression variables. Rubin rules were used to congregate the imputations [24]. After imputation, we constructed a linear mixed-effects model assessing the impact of the intervention on quality of life and asthma control. In this mixed model, we used strategy, follow-up, strategy × follow-up interaction, and a random intercept at the patient level to adjust for repeated measurements, as data from 3, 6, 9, and 12 months were imputed. These mixed models provided estimates for changes in asthma-related quality of life and asthma control during the follow-up period per strategy.

Patients were stratified for clinical asthma control at baseline using the ACQ6. Patients with an ACQ6 score <1.5 at baseline were considered as controlled, whereas those with an ACQ6 score ≥1.5 were considered as uncontrolled.

To define user engagement, we established a cutoff point of 12 logins per year. Patients with <12 logins were included in a low user engagement group, and those with ≥12 logins were included in a high user engagement group. We assessed whether there were differences in the AQLQ score between the user engagement groups.

To obtain insights into which self-management domains lead to improved quality of life, we used mixed models with the improvement in the AQLQ score with the internet-based strategy during follow-up as a dependent variable and the improvement in separate heiQ domains as the independent variable. All analyses were performed with Stata/IC 11.0 (Stata Corp).

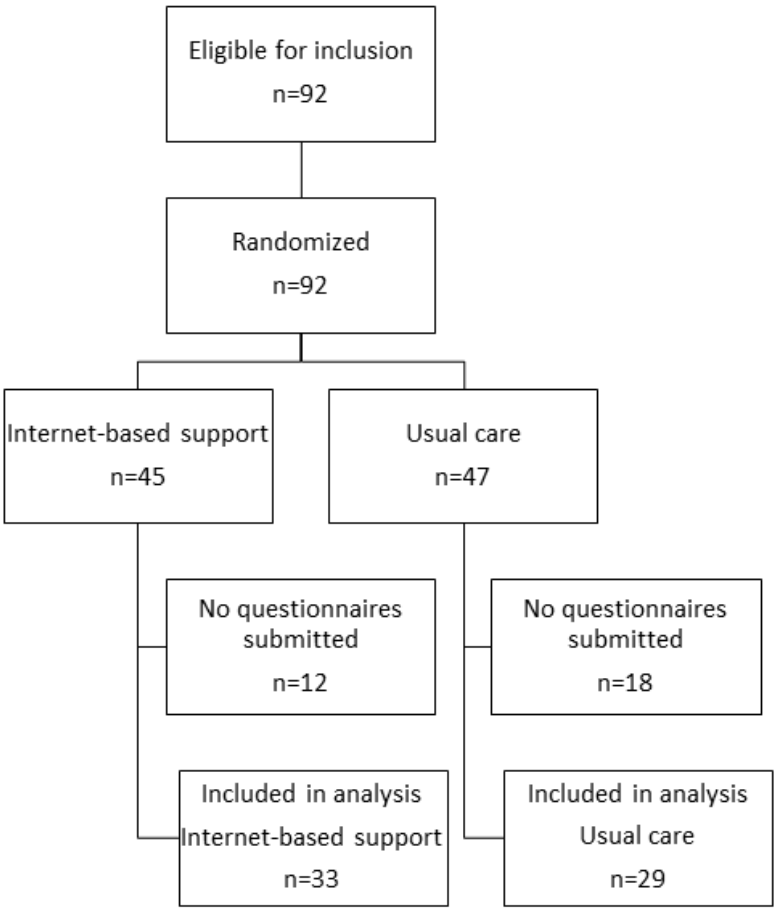
XSL•FO
RenderX

Results

Recruitment and Baseline Characteristics

In total, 92 patients from the Dutch Asthma Centre Davos were eligible for enrollment and were randomized after consent. However, 30 patients did not fill out the baseline questionnaire required to be included in the analysis. In our analysis, the usual care strategy group included 29 patients and the internet-based strategy group included 33 patients. Figure 1 shows the patient flow diagram. Out of 310 records (five questionnaires for 62 patients), 91 (27%) were missing, with the internet-based strategy group missing 49 questionnaires (29%) and the usual care strategy group missing 34 questionnaires (23%) (*P*=.22). All baseline questionnaires were completed.

Figure 1. Flow diagram for the study patients.



The population characteristics at baseline are shown in Table 1. There were no relevant differences in age, sex, quality of life, and asthma between the two strategy groups. It should however be noted that clinically relevant [20], but statistically insignificant, baseline differences were found in quality of life (*P*=.11) and asthma control (*P*=.17). For both parameters, the usual care strategy group appeared to score slightly better than the internet-based strategy group. In the domain negative affect, the internet-based strategy group scored worse than the usual care strategy group.

XSL•FO
RenderX

Table 1. Baseline characteristics.

Characteristic	Usual care strategy (n=29), n (%) or mean (SD)	Internet-based strategy (n=33), n (%) or mean (SD)	*P* value
Age, years	44.0 (2.4)	46.7 (2.3)	.47
Male sex	9 (31%)	8 (24%)	.56
AQLQ[a] score	5.6 (0.2)	5.2 (0.2)	.11
AQLQ-activity limitation score	5.3 (0.2)	4.8 (0.2)	.17
AQLQ-symptom score	5.7 (0.2)	5.2 (0.2)	.10
AQLQ-emotional function score	6.4 (0.1)	6.0 (0.2)	.14
AQLQ-environmental stimuli score	5.5 (0.2)	5.1 (0.2)	.31
ACQ6[b] score	1.5 (0.2)	1.9 (0.2)	.17
heiQ[c]-behavior score	3.3 (0.1)	3.1 (0.1)	.91
heiQ-positive engagement score	3.3 (0.1)	3.3 (0.1)	.86
heiQ-negative affect score	1.4 (0.1)	1.8 (0.1)	.006
heiQ-monitoring/insight score	3.4 (0.1)	3.3 (0.1)	.52
heiQ-constructive attitude score	3.3 (0.1)	3.2 (0.1)	.53
heiQ-skill acquisition score	3.2 (0.1)	3.0 (0.1)	.06
heiQ-social support score	3.2 (0.1)	3.0 (0.1)	.07
heiQ-navigation score	3.3 (0.1)	3.3 (0.1)	.87

[a]AQLQ: Asthma Quality of Life Questionnaire.

[b]ACQ6: six-question Asthma Control Questionnaire.

[c]heiQ: Health Education and Impact Questionnaire.

Health-Related Quality of Life

The changes in asthma-related quality of life for both strategies are shown in Figure 2. Quality of life declined in both groups during follow-up, showing less decline in the internet-based strategy group. There were no relevant between-group differences in quality of life at 12 months. Analysis by a linear mixed-effects model, with treatment, strategy, follow-up × strategy interaction, and a random intercept at the patient level to adjust for repeated measurements, showed that the decline in quality of life according to the 3-month AQLQ scores (a higher AQLQ score is better) was significantly smaller during 12 months of follow-up in the internet-based strategy group than in the usual care strategy group (mean between-group AQLQ score difference 0.39, 95% CI 0.092-0.69; *P*=.01).

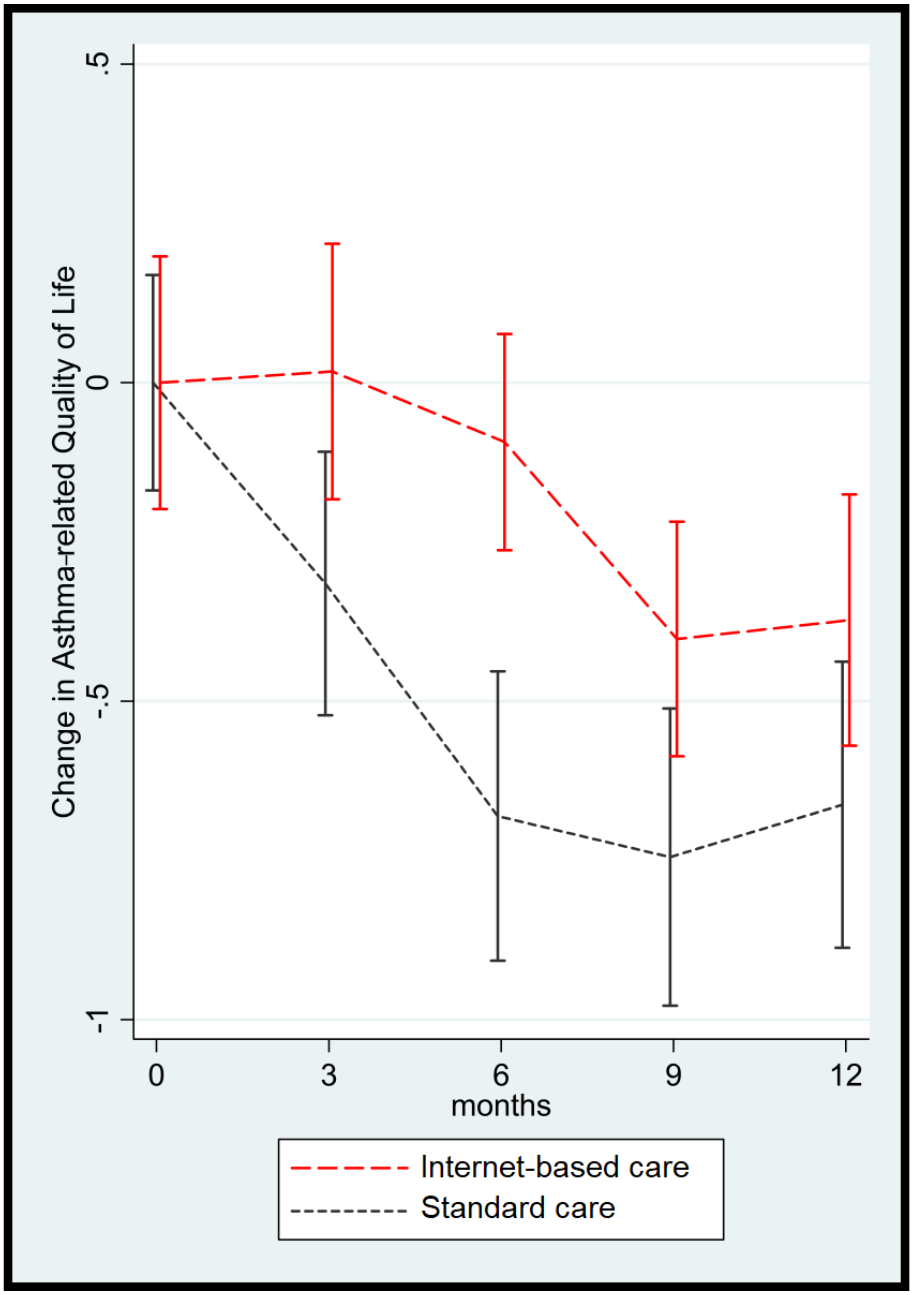
A similar graph showing quality of life during follow-up using absolute values instead of changes is shown in Multimedia Appendix 2. As the baseline values for the ACQ6 (*P*=.17) and AQLQ (*P*=.11) were slightly different between the groups, but not statistically significant, we analyzed the results separately in strata according to well-established cutoff points for asthma control at baseline. This approach using stratification was also adopted in a previous study [25]. Patients with an ACQ6 score <1.5 were considered as controlled, whereas those with an ACQ6 score ≥1.5 were considered as uncontrolled. In the usual care strategy group, 16 patients were considered as controlled and 13 were considered as uncontrolled. In the internet-based strategy group, 12 patients were considered as controlled and 21 were considered as uncontrolled. Among patients with controlled asthma, a comparable decline was observed with both strategies (mean AQLQ score difference 0.07, 95% CI −0.38 to 0.53; *P*=.76) (Table 2). However, among patients with uncontrolled asthma at baseline, the decline in quality of life was relatively smaller in the internet-based strategy group than in the usual care strategy group (mean between-group AQLQ score difference 0.59, 95% CI 0.19-0.99; *P*=.003) (Figure 3).

The results of the AQLQ domains are shown in Table 3. Patients in the internet-based strategy group showed less worsening in activity limitation during follow-up as compared with the usual care strategy group (AQLQ score difference 0.44, 95% CI 0.06-0.82; *P*=.02). Overall, patients in the internet-based strategy group reported significantly less impact of asthma symptoms on quality of life during follow-up (AQLQ score difference 0.48, 95% CI 0.12-0.83; *P*=.009). Among patients with controlled asthma at baseline, no differences between the study arms were found. However, among patients with uncontrolled asthma at baseline, patients in the usual care strategy group reported significantly more impact of symptoms on quality of life, whereas those in the internet-based strategy group maintained better levels of symptom-related quality of life throughout the follow-up (AQLQ score difference 0.83, 95% CI 0.34-1.31; *P*=.001).

Similar results were found in the environmental stimuli AQLQ domain in favor of the intervention, specifically among patients with uncontrolled asthma. Overall, no relevant differences were found in the emotional functioning domain, except within the uncontrolled group, which had less impact of emotional functioning on quality of life with the internet-based strategy as compared with the usual care strategy.

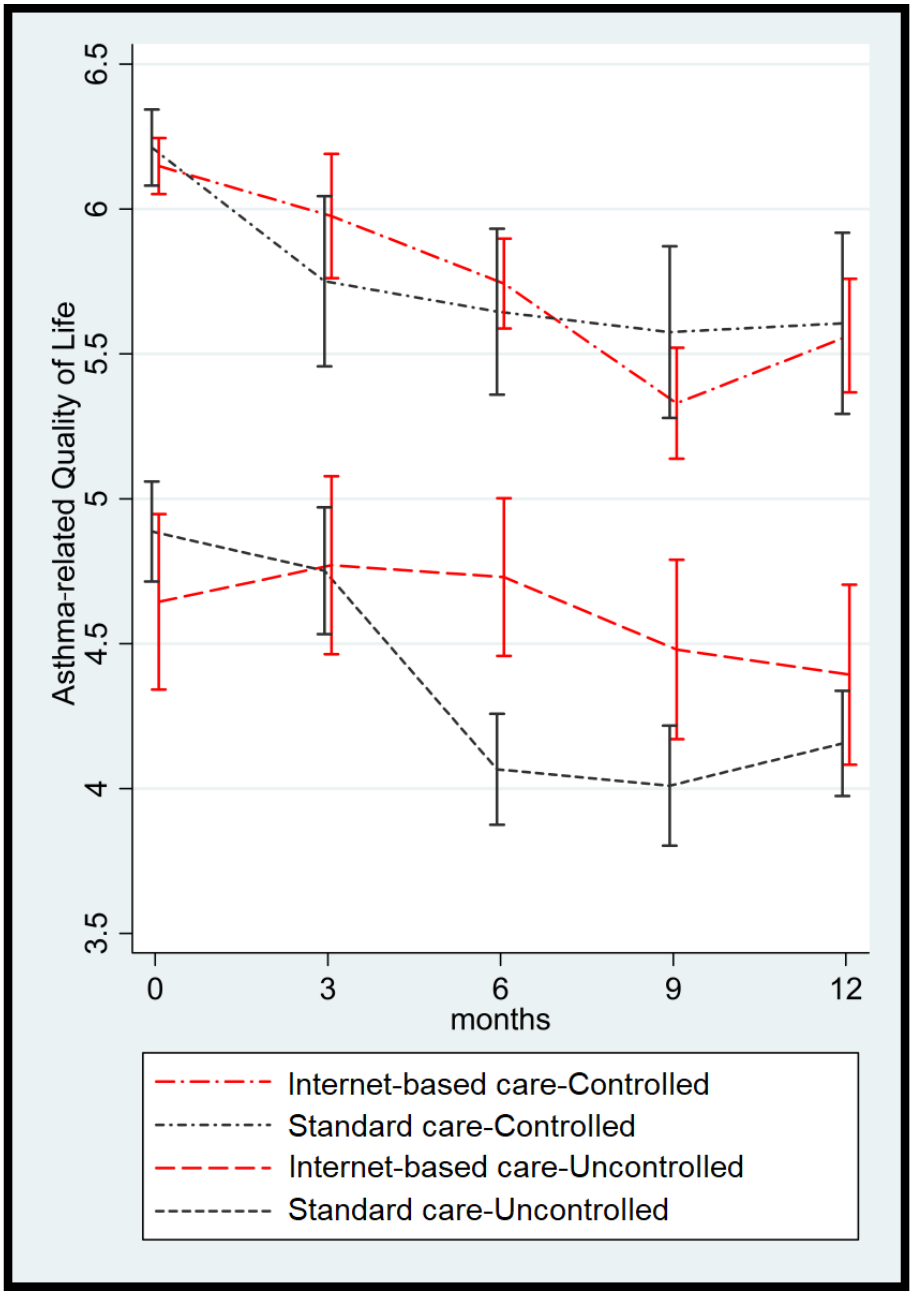
XSL•FO
RenderX

Figure 2. Quality of life according to the Asthma Quality of Life Questionnaire (a higher score is better) during follow-up.



XSL•FO
RenderX

Figure 3. Quality of life according to the Asthma Quality of Life Questionnaire (a higher score is better) during follow-up in controlled and uncontrolled asthma patients.



XSL•FO
RenderX

Table 2. Asthma-related quality of life and asthma control between strategies.

Variable	Difference between baseline and follow-up, mean (95% CI)		Difference between strategies[a], mean (95% CI)	*P* value
	Usual care strategy	Internet-based strategy		
AQLQ[b]	−0.60 (−0.82 to −0.38)	−0.21 (−0.42 to −0.01)[c]	0.39 (0.092 to 0.69)[c]	.01
AQLQ, controlled patients	−0.57 (−0.87 to −0.27)	−0.50 (−0.84 to −0.15)[c]	0.07 (−0.38 to 0.53)	.76
AQLQ, uncontrolled patients	−0.64 (−0.95 to −0.33)	−0.05 (−0.30 to 0.19)	0.59 (0.19 to 0.99)[c]	.003
ACQ6[d]	0.67 (0.41 to 0.93)	0.17 (−0.08 to 0.41)	−0.50 (−0.86 to −0.15)[c]	.006
ACQ6, controlled patients	0.74 (0.37 to 1.11)[c]	0.70 (0.27 to 1.13)[c]	−0.038 (−0.60 to 0.53)	.89
ACQ6, uncontrolled patients	0.59 (0.24 to 0.94)[c]	−0.14 (−0.41 to 0.14)	−0.73 (−1.18 to −0.28)[c]	.002

[a]An AQLQ score difference >0 and an ACQ6 score difference <0 reflect a positive effect of the intervention.

[b]AQLQ: Asthma Quality of Life Questionnaire.

[c]Statistically significant.

[d]ACQ6: six-question Asthma Control Questionnaire.

Table 3. Asthma-related quality of life subdomains between strategies.

Variable	Difference between baseline and follow-up, mean (95% CI)		Difference between strategies, mean (95% CI)	*P* value
	Usual care strategy	Internet-based strategy		
AQLQ[a]-activity limitation	−0.73 (−1.00 to −0.45)[b]	−0.29 (−0.54 to −0.03)[b]	0.44 (0.06 to 0.82)[b]	.02
AQLQ-activity limitation, controlled patients	−0.86 (−1.26 to −0.46)[b]	−0.50 (−0.96 to −0.03)[b]	0.36 (−0.25 to 0.97)	.24
AQLQ-activity limitation, uncontrolled patients	−0.57 (−0.95 to −0.18)[b]	−0.17 (−0.47 to 0.14)	0.40 (−0.09 to 0.89)	.11
AQLQ-symptoms	−0.65 (−0.91 to −0.39)[b]	−0.17 (−0.41 to 0.07)	0.48 (0.12 to 0.83)[b]	.009
AQLQ-symptoms, controlled patients	−0.56 (−0.91 to −0.22)[b]	−0.61 (−1.00 to 0.21)	−0.04 (−0.57 to 0.48)	.88
AQLQ-symptoms, uncontrolled patients	−0.75 (−1.13 to −0.37)[b]	0.08 (−0.22 to 0.38)	0.83 (0.34 to 1.31)[b]	.001
AQLQ-environmental stimuli	−0.63 (−0.86 to −0.40)[b]	−0.32 (−0.53 to −0.10)[b]	0.32 (0.01 to 0.63)[b]	.047
AQLQ-environmental stimuli, controlled patients	−0.54 (−0.84 to −0.24)[b]	−0.42 (−0.76 to −0.07)[b]	0.12 (−0.34 to 0.58)	.60
AQLQ-environmental stimuli, uncontrolled patients	−0.75 (−1.10 to −0.41)[b]	−0.26 (−0.53 to 0.01)	0.49 (0.05 to 0.93)[b]	.028
AQLQ-emotional function	−0.19 (−0.40 to 0.02)	−0.07 (−0.27 to 0.13)	0.12 (−0.17 to 0.41)	.43
AQLQ-emotional function, controlled patients	0.04 (−0.22 to 0.29)	−0.30 (0.60 to 0.00)	−0.33 (−0.73 to 0.06)	.10
AQLQ-emotional function, uncontrolled patients	−0.46 (−0.79 to 0.14)	0.06 (−0.20 to 0.32)	0.52 (0.10 to 0.94)[b]	.02

[a]AQLQ: Asthma Quality of Life Questionnaire.

[b]Statistically significant.

Asthma Control

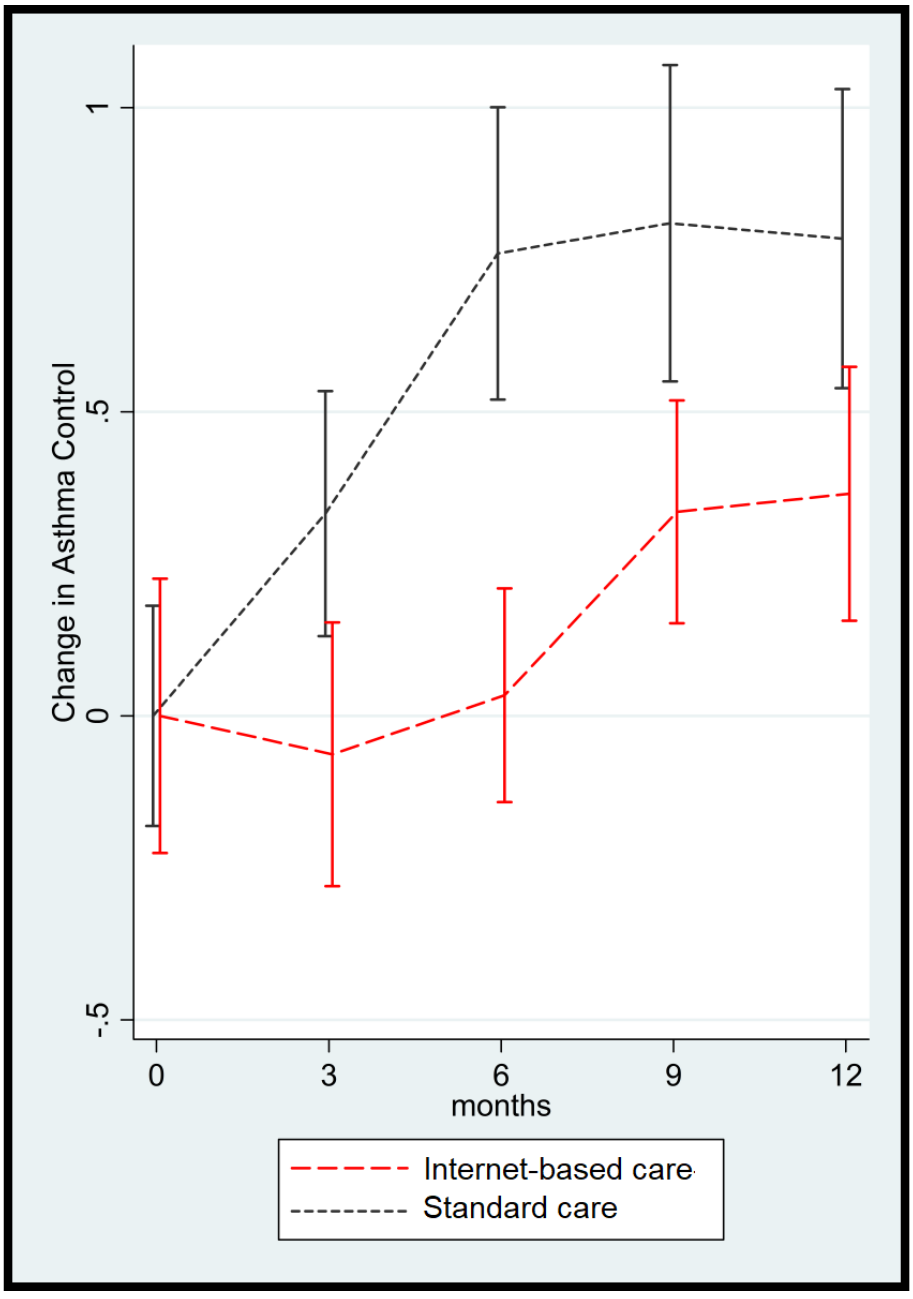
The favorable findings in quality of life for the internet-based strategy were confirmed by changes in asthma control. A similar mixed model showed a favorable and statistically significant change in asthma control (a lower ACQ6 score is better) in the internet-based strategy group during follow-up as compared with the usual care strategy group (mean ACQ6 score difference −0.50, 95% CI −0.86 to −0.15; *P*=.006) (Figure 4).

A similar graph showing asthma control during follow-up using absolute values instead of changes is shown in Multimedia Appendix 3. With respect to the stratified analysis, a similar pattern was observed for asthma control. Among controlled patients, both groups showed a comparable trend of declining

XSL•FO
RenderX

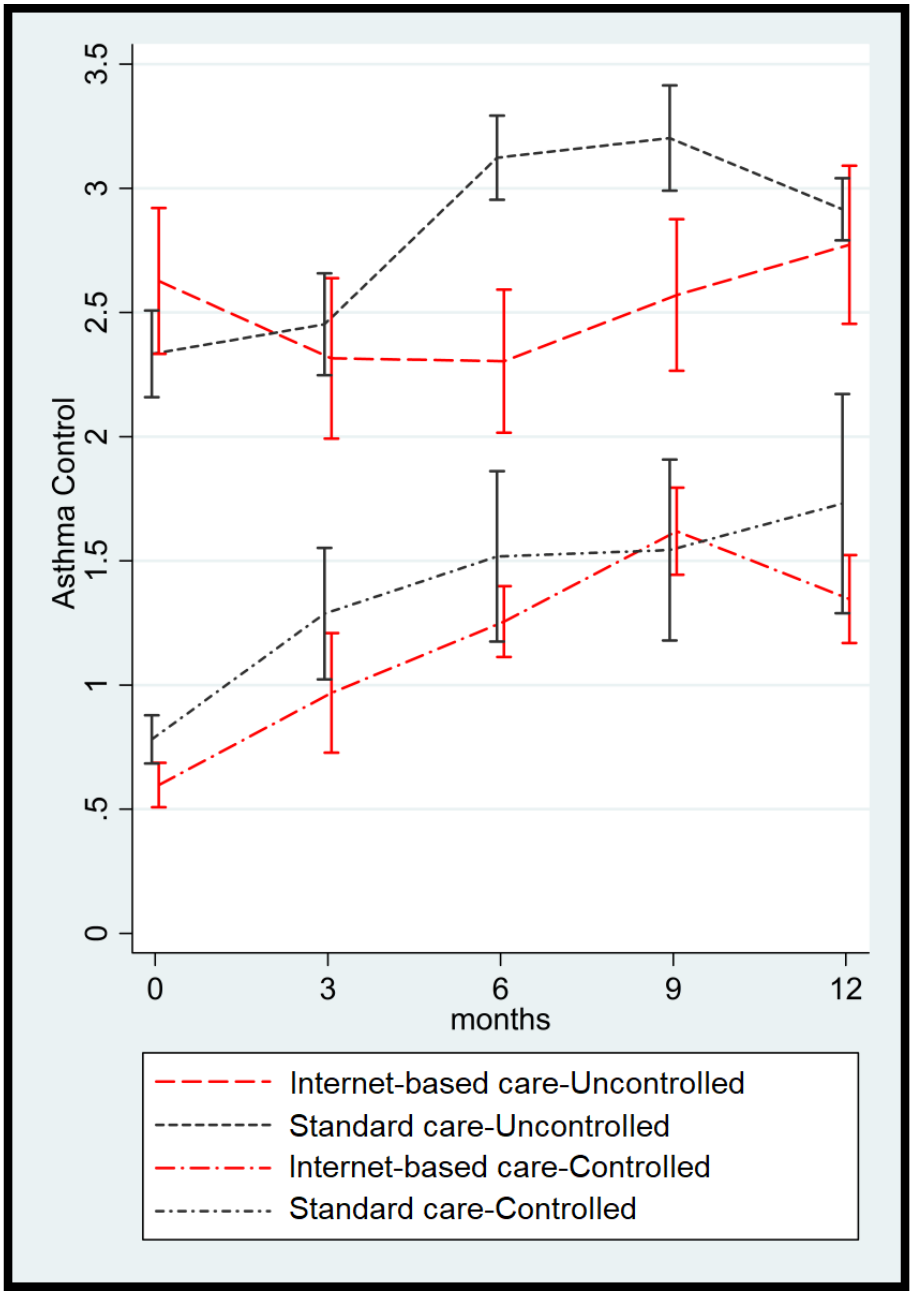
asthma control during follow-up (mean ACQ6 score difference 0.04, 95% CI −0.60 to 0.53, *P*=.89) (Figure 5). However, among uncontrolled patients, a clinically and significantly better asthma control level was sustained with the internet-based strategy (mean ACQ score difference −0.73, 95% CI −1.176 to −0.277; *P*=.002), and this was most noticeable at 6 and 9 months (Figure 5).

Figure 4. Asthma control according to the Asthma Control Questionnaire (a lower score is better) during follow-up.



XSL•FO
RenderX

Figure 5. Asthma control according to the Asthma Control Questionnaire (a lower score is better) during follow-up in controlled and uncontrolled asthma patients.



eHealth User Engagement

The results with respect to eHealth user engagement are depicted in Figures 6 and 7. There was a decline in login frequency in both the low and high user engagement groups. During the last 6 months of follow-up, asthma-related quality of life was significantly better maintained in the high user engagement group than in the low user engagement group (mean AQLQ score difference −0.39, 95% CI 0.049-0.73; *P*=.03).

XSL•FO
RenderX

Figure 6. Login frequency in low (left panel; n=18) and high (right panel; n=13) user engagement groups.

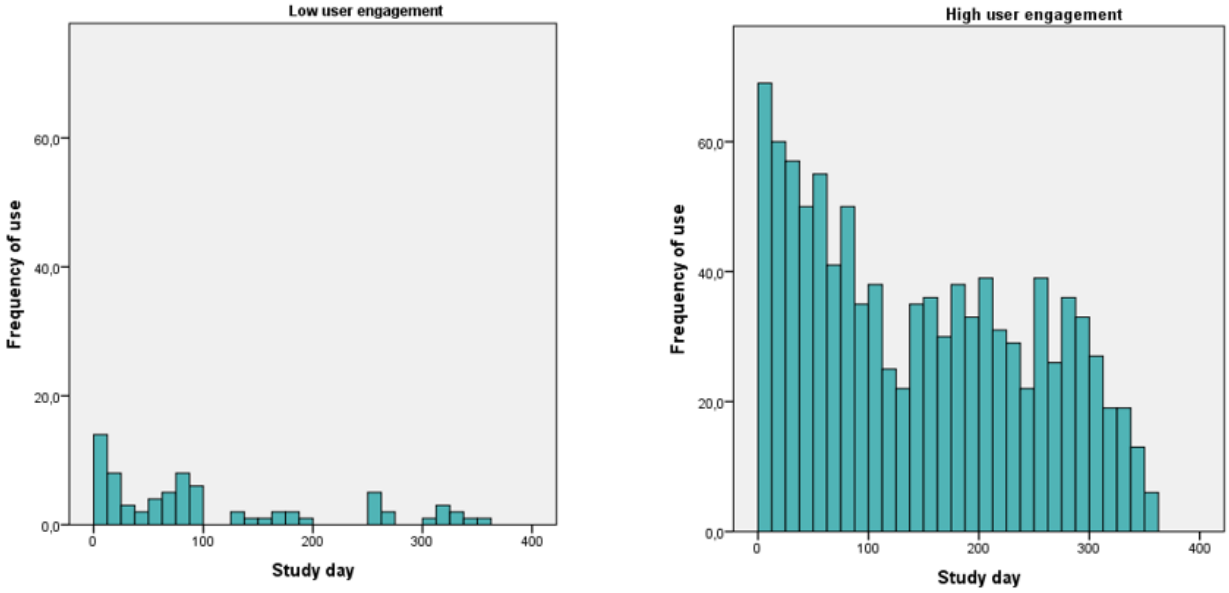
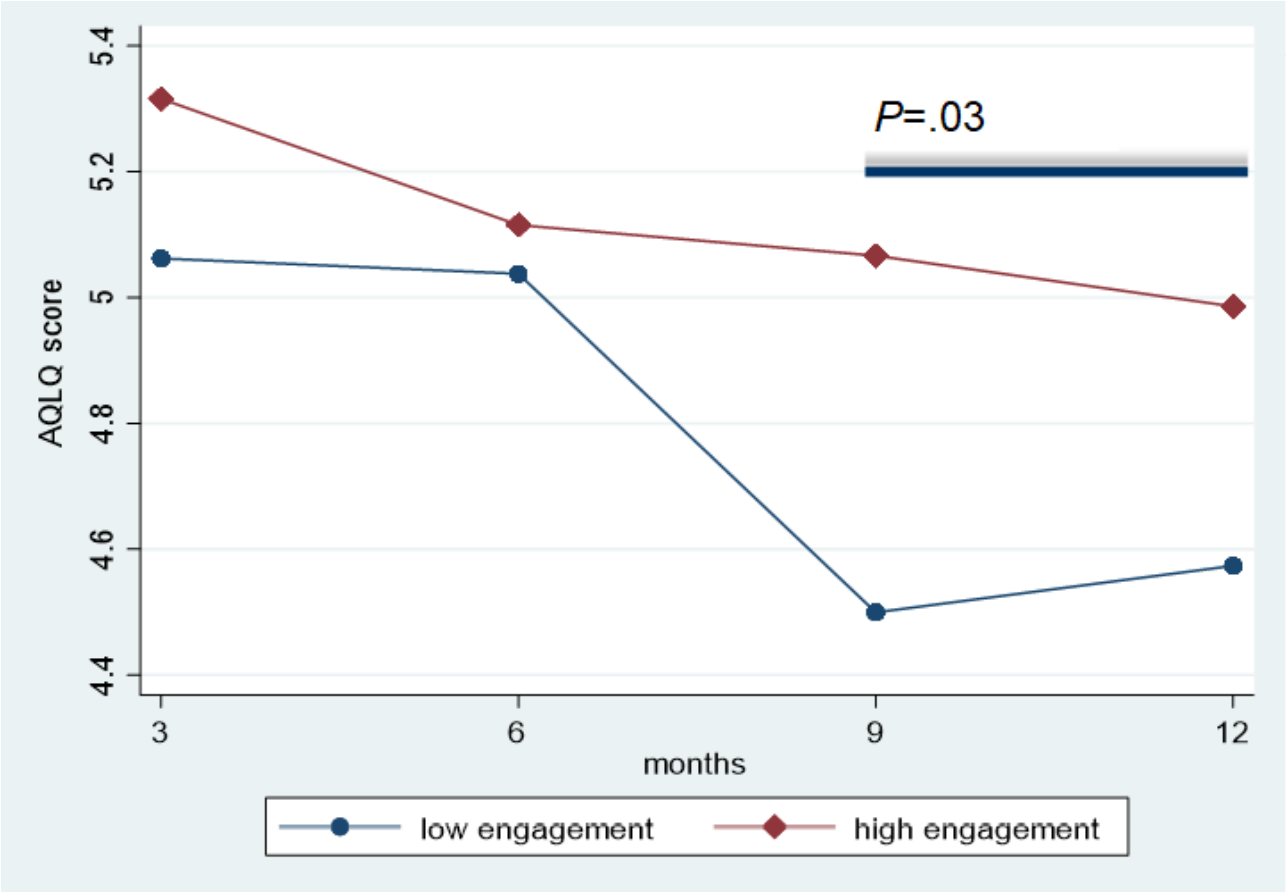


Figure 7. Asthma-related quality of life in low and high user engagement groups. AQLQ: Asthma Quality of Life Questionnaire.



Health Education Impact

The results of the mixed models for health education impact are shown in Table 4. We assessed whether the improvement in asthma-related quality of life could be explained by the changes in the separate heiQ domains. Improvement in the domain of health care navigation during follow-up was positively related to improvement in quality of life (AQLQ score difference 0.55, 95% CI 0.06-1.04; *P*=.03) in patients with access to the intervention. Although the domains of health-directed behavior, positive engagement, and monitoring and insight suggested a positive relationship with asthma-related quality of life (effect estimates of 0.28 [*P*=.19], 0.29 [*P*=.19],

XSL•FO
RenderX

and 0.40 [*P*=.18] points, respectively), none of these domains reached statistical significance.

Table 4. Relation between improvement in health education impact domains and improvement in quality of life.

heiQ[a] domains	Mixed model effect sizes[b] (β), mean (95% CI)	*P* value
Health-directed behavior	0.28 (−0.14 to 0.70)	.19
Positive engagement	0.29 (−0.15 to 0.73)	.19
Negative affect	−0.20 (−0.60 to 0.19)	.31
Monitoring and insight	0.40 (−0.18 to 0.97)	.18
Constructive attitude	0.15 (−0.26 to 0.57)	.47
Skill acquisition	0.03 (−0.58 to 0.64)	.92
Social support	0.13 (−0.24 to 0.49)	.49
Health care navigation	0.55 (0.06 to 1.04)	.03

[a]heiQ: Health Education and Impact Questionnaire.

[b]In the domain of negative affect, a negative effect size reflects a favorable outcome, whereas in the other domains, a positive effect size reflects favorable outcomes.

Discussion

In this study, we assessed the effectiveness of internet-based self-management support in addition to usual care in patients with severe asthma after completing high-altitude climate treatment. Both the internet-based strategy and usual care strategy groups showed declines in asthma-related quality of life and asthma control during follow-up. This trend was expected as patients were enrolled upon completing a pulmonary rehabilitation just before returning to their home situation at sea level. However, the declines in both endpoints in the internet-based strategy group were significantly smaller than those in the usual care strategy group. In particular, in the internet-based strategy group, patients who still exhibited uncontrolled asthma at discharge did not show a further decline in quality of life in contrast to their counterparts in the usual care strategy group.

The individual domains of the AQLQ (activity limitation, symptoms, emotional function, and environmental stimuli) showed results similar to those of the AQLQ aggregated score described above. The most prominent results were found within the symptoms domain, where patients with controlled asthma reported an increase in symptoms during follow-up, with comparable effect sizes between the study arms. However, among patients with uncontrolled asthma at baseline, those who received the usual care strategy reported clinically relevant increases in symptoms, whereas those who received the internet-based strategy reported no increase in symptoms.

Asthma control did not decline greatly during follow-up in the internet-based strategy group, whereas asthma control declined in the usual care strategy group. Similar to the findings for quality of life, no significant difference in the decline in asthma control was found between the study groups among patients with controlled asthma at discharge. However, among patients with uncontrolled asthma, no decline in asthma control was observed with the internet-based strategy.

There are sparse data regarding the effectiveness of self-management support strategies in patients with severe asthma after completion of high-altitude climate treatment. Our results are consistent with the findings of previously published articles regarding interventions in patients with mild asthma [10-12,15]. In addition, our results substantiate the findings of a recently published uncontrolled observational study reporting sustained improvements in asthma control and quality of life after high-altitude climate treatment in the years before this study started [6,7]. To our knowledge, this is the first study to assess the effectiveness of internet-based self-management support in patients with severe asthma after completing high-altitude climate treatment.

Several possible limitations in this study need to be addressed, including possible selection bias, the nonresponse rate, and statistical power. As this was a self-management study and a pragmatic trial, we did not spend large amounts of time in motivating people during follow-up to fill out questionnaires, as this would be an intervention by itself and was not possible in the clinical setting after implementation. Therefore, in our data set, 27% (91/310) of possible entries were missing. However, as the rates of missing data in both groups were comparable, we assumed that there was no impact on the internal validity of the study. In addition, the relatively high drop-out rate is consistent with an earlier study in patients with severe asthma [6]. It is not clear whether this is due to omissions within the studies or the requirement of an alternative approach for the population to adhere to the study protocol. Another explanation might involve patients' initial willingness to cooperate with a study conducted by the rehabilitation center where they stayed for many weeks, resulting in the inclusion of patients without a high intrinsic motivation, who are usually not included in a typical 12-month trial. We imputed missing data using linear regression modelling with 100 sets of imputations, thereby minimizing the impact of missing data.

As we calculated the differences between baseline and follow-up, it could be argued that our results might be explained by selection bias as the usual care population had better, although not statistically significant, asthma-related quality of life and asthma control at baseline. Therefore, there might be more room for a decline in the follow-up period. To correct for

XSL•FO
RenderX

this possible bias, we stratified for asthma control at baseline, showing that quality of life and asthma control did not decrease in patients with uncontrolled asthma. We therefore conclude that this bias does not explain the current results.

Not all differences found in this study reached statistical significance, possibly due to the limited number of patients enrolled. Based on our power analysis, we aimed to enroll 72 patients (36 patients per arm) to have sufficient power for the detection of a minimally important difference. Of the 92 eligible patients, 62 contributed to the study. However, we had sufficient power to detect a minimally important and consistent difference of 0.5 points in the AQLQ and ACQ. Therefore, despite the fact that the number of patients could still be regarded as a limitation, we feel that the number of patients included was sufficient for our conclusions.

A further limitation of this study is that the results were based upon self-reported outcome measures, as we had no access to FEV_1 measurements or other biomarkers. Although such markers would definitely be of additional value, we feel that by using well-validated instruments our self-reported outcomes are sufficiently robust to draw conclusions.

One of the strengths of this study is the pragmatic approach we adopted. Patients were included by members of the rehabilitation center, indicating the feasibility of imbedding a self-management program on completing the rehabilitation program. The patients had access to a helpdesk service for technical support, but no follow-up from the rehabilitation center or the patient's own physician was required for support during the intervention. Certainly, regular visits to their own physicians continued to be applicable. Although this study does not amount to a formal feasibility study, the usability, accessibility, and more than monthly use of the intervention in half of the intervention group suggest that this self-management support intervention might be feasible without investments other than the introduction to the program during follow-up and the costs for technical maintenance of the system.

These findings confirm the presumption that patients with severe asthma who complete a high-altitude climate treatment deteriorate at a group level in terms of asthma control and quality of life after discharge. Partially, this could be attributed to regression to the mean, as patients leave the rehabilitation in an optimized health status. However, the results suggest that this decline could be reduced by offering an internet-based self-management program specifically in patients who leave rehabilitation with uncontrolled asthma (ACQ score ≥1.5). These patients represent the group of patients in which achieving asthma control is most complicated. Additionally, without self-management support intervention, their quality of life and asthma control declined even further, resulting in lower scores than patients who were able to achieve asthma control during rehabilitation. However, with an internet-based self-management program, this group may be able to stabilize their symptoms, which is an important and clinically relevant improvement.

In this study, the results were achieved without any interference by health care professionals in the internet-based self-management program after discharge, suggesting that the intervention is easily feasible for implementation and the approach might be very cost-effective, although this needs to be confirmed in future research.

Other patients with asthma who have not completed rehabilitation might also benefit from such interventions. For instance, patients discharged after asthma-related hospital admission may benefit from the same self-management support tools. The support tools are easily tailored to patients with various degrees of severity, enabling health care providers to provide this intervention to patients with less severe asthma. Access to the internet has become ever more widespread, implying that internet-based self-management tools are already accessible to most patients in developed countries. The large dropout rate emphasizes the importance of motivating people prior to enrollment in postrehabilitation self-management interventions. Future research is needed as evidence of self-management support in patients with severe asthma remains limited. The assessment of lung function and biomarkers may be of additional value to establish the effectiveness of self-management outcomes.

In conclusion, internet-based self-management support is an effective addition to usual care in patients with severe asthma on completion of a pulmonary rehabilitation program. Implementation of such interventions is feasible in clinical practice and may contribute to the stabilization of symptoms in patients with severe asthma.

Acknowledgments

We thank the involved physicians, nurses, and lung function technicians for their help and active guidance throughout the study. This study was funded by Lung Foundation Netherlands, The Netherlands Organization for Health Research and Development (ZonMW), and Chiesi the Netherlands.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the mobile version of the PatientCoach app.
[PDF File (Adobe PDF File), 835 KB - jmir_v22i7e13145_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Quality of life in absolute values (a higher score is better) during follow-up.
[PNG File , 804 KB - jmir_v22i7e13145_app2.png]

Multimedia Appendix 3
Asthma control in absolute values (a lower score is better) during follow-up.
[PNG File , 774 KB - jmir_v22i7e13145_app3.png]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 12826 KB - jmir_v22i7e13145_app4.pdf]

References

1. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald JM, et al. Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J 2008 Jan;31(1):143-178 [FREE Full text] [doi: 10.1183/09031936.00138707] [Medline: 18166595]
2. Bel EH, Sousa A, Fleming L, Bush A, Chung KF, Versnel J, Unbiased Biomarkers for the Prediction of Respiratory Disease Outcome (U-BIOPRED) Consortium, Consensus Generation. Diagnosis and definition of severe refractory asthma: an international consensus statement from the Innovative Medicine Initiative (IMI). Thorax 2011 Oct;66(10):910-917. [doi: 10.1136/thx.2010.153643] [Medline: 21106547]
3. Rijssenbeek-Nouwens LH, Bel EH. High-altitude treatment: a therapeutic option for patients with severe, refractory asthma? Clin Exp Allergy 2011 Jun;41(6):775-782. [doi: 10.1111/j.1365-2222.2011.03733.x] [Medline: 21518039]
4. Vinnikov D, Khafagy A, Blanc PD, Brimkulov N, Steinmaus C. High-altitude alpine therapy and lung function in asthma: systematic review and meta-analysis. ERJ Open Res 2016 Apr;2(2) [FREE Full text] [doi: 10.1183/23120541.00097-2015] [Medline: 27730196]
5. de Jonge P, Huyse FJ, Stiefel FC. Case and care complexity in the medically ill. Med Clin North Am 2006 Jul;90(4):679-692. [doi: 10.1016/j.mcna.2006.04.005] [Medline: 16843768]
6. Rijssenbeek-Nouwens LH, Fieten KB, Bron AO, Hashimoto S, Bel EH, Weersink EJ. High-altitude treatment in atopic and nonatopic patients with severe asthma. Eur Respir J 2012 Dec;40(6):1374-1380 [FREE Full text] [doi: 10.1183/09031936.00195211] [Medline: 22441741]
7. Fieten KB, Rijssenbeek-Nouwens LH, Hashimoto S, Bel EH, Weersink EJ. Less exacerbations and sustained asthma control 12 months after high altitude climate treatment for severe asthma. Allergy 2019 Mar;74(3):628-630. [doi: 10.1111/all.13664] [Medline: 30428132]
8. van Stel HF, Colland VT, Heins NL, Rijssenbeek-Nouwens LH, Everaerd W. Assessing inpatient pulmonary rehabilitation using the patient's view of outcome. J Cardiopulm Rehabil 2002;22(3):201-210. [doi: 10.1097/00008483-200205000-00013] [Medline: 12042690]
9. Fischer MJ, Scharloo M, Abbink JJ, Thijs-Van A, Rudolphus A, Snoei L, et al. Participation and drop-out in pulmonary rehabilitation: a qualitative analysis of the patient's perspective. Clin Rehabil 2007 Mar;21(3):212-221. [doi: 10.1177/0269215506070783] [Medline: 17329278]
10. Lewis KE, Annandale JA, Warm DL, Rees SE, Hurlin C, Blyth H, et al. Does home telemonitoring after pulmonary rehabilitation reduce healthcare use in optimized COPD? A pilot randomized trial. COPD 2010 Feb;7(1):44-50. [doi: 10.3109/15412550903499555] [Medline: 20214462]
11. Nguyen HQ, Gill DP, Wolpin S, Steele BG, Benditt JO. Pilot study of a cell phone-based exercise persistence intervention post-rehabilitation for COPD. Int J Chron Obstruct Pulmon Dis 2009;4:301-313 [FREE Full text] [doi: 10.2147/copd.s6643] [Medline: 19750190]
12. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010 Aug 04(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
13. Pinnock H, Parke HL, Panagioti M, Daines L, Pearce G, Epiphaniou E, PRISMSRECURSIVE groups. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64 [FREE Full text] [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
14. Xiao Q, Wang J, Chiang V, Choi T, Wang Y, Sun L, et al. Effectiveness of mHealth Interventions for Asthma Self-Management: A Systematic Review and Meta-Analysis. Stud Health Technol Inform 2018;250:144-145. [Medline: 29857410]
15. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, SMASHING (Self-Management in Asthma Supported by Hospitals, ICT, NursesGeneral Practitioners) Study Group. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [doi: 10.7326/0003-4819-151-2-200907210-00008] [Medline: 19620163]

XSL•FO
RenderX

16. van Gaalen JL, Beerthuizen T, van der Meer V, van Reisen P, Redelijkheid GW, Snoeck-Stroband JB, SMASHING Study Group. Long-term outcomes of internet-based self-management support in adults with asthma: randomized controlled trial. J Med Internet Res 2013 Sep 12;15(9):e188 [FREE Full text] [doi: 10.2196/jmir.2640] [Medline: 24028826]
17. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
18. Reddel HK, Jenkins CR, Partridge MR. Self-management support and other alternatives to reduce the burden of asthma and chronic obstructive pulmonary disease. Int J Tuberc Lung Dis 2014 Dec;18(12):1396-1406. [doi: 10.5588/ijtld.14.0371] [Medline: 25517803]
19. Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. Am Rev Respir Dis 1993 Apr;147(4):832-838. [doi: 10.1164/ajrccm/147.4.832] [Medline: 8466117]
20. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. J Clin Epidemiol 1994 Jan;47(1):81-87. [doi: 10.1016/0895-4356(94)90036-1] [Medline: 8283197]
21. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. Eur Respir J 1999 Oct;14(4):902-907 [FREE Full text] [doi: 10.1034/j.1399-3003.1999.14d29.x] [Medline: 10573240]
22. Ammerlaan JW, van Os-Medendorp H, Sont JK, Elsworth GR, Osborne RH. Validation of the dutch version of the health education impact questionnaire (HEIQ) and comparison of the Dutch translation with the English, German and French HEIQ. Health Qual Life Outcomes 2017 Jan 31;15(1):28 [FREE Full text] [doi: 10.1186/s12955-017-0601-4] [Medline: 28143551]
23. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
24. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
25. van der Meer V, van Stel HF, Bakker MJ, Roldaan AC, Assendelft WJ, Sterk PJ, SMASHING (Self-Management of Asthma Supported by Hospitals, ICT, NursesGeneral practitioners) Study Group. Weekly self-monitoring and treatment adjustment benefit patients with partly controlled and uncontrolled asthma: an analysis of the SMASHING study. Respir Res 2010 Jun 10;11:74 [FREE Full text] [doi: 10.1186/1465-9921-11-74] [Medline: 20537124]

Abbreviations

ACQ: Asthma Control Questionnaire
ACQ6: six-question Asthma Control Questionnaire
AQLQ: Asthma Quality of Life Questionnaire
eHealth: electronic health
FEV1: forced expiratory volume in one second
heiQ: Health Education and Impact Questionnaire

Edited by G Eysenbach; submitted 21.05.19; peer-reviewed by A Nguyen, D Goyal, S Hashimoto, L de Witte, R Ramsey, JB Park; comments to author 03.09.19; revised version received 29.10.19; accepted 29.03.20; published 22.07.20.

Please cite as:
Beerthuizen T, Rijssenbeek-Nouwens LH, van Koppen SM, Khusial RJ, Snoeck-Stroband JB, Sont JK
Internet-Based Self-Management Support After High-Altitude Climate Treatment for Severe Asthma: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e13145
URL: https://www.jmir.org/2020/7/e13145
doi:10.2196/13145
PMID:32706692

©Thijs Beerthuizen, Lucia H Rijssenbeek-Nouwens, Sophia M van Koppen, Rishi J Khusial, Jiska B Snoeck-Stroband, Jacob K Sont. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Novel Stepped Care Approach to Provide Education and Exercise Therapy for Patellofemoral Pain: Feasibility Study

Danilo De Oliveira Silva[1*], PhD; Marcella F Pazzinatto[1*], PhD; Kay M Crossley[1*], PhD; Fabio M Azevedo[2*], PhD; Christian J Barton[1*], PhD

[1]La Trobe Sport and Exercise Medicine Research Centre, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Australia
[2]Laboratory of Biomechanics and Motor Control, Sao Paulo State University, Presidente Prudente, Brazil
[*]all authors contributed equally

Corresponding Author:
Danilo De Oliveira Silva, PhD
La Trobe Sport and Exercise Medicine Research Centre
School of Allied Health, Human Services and Sport
La Trobe University
Plenty Road & Kingsbury Drive, Bundoora VIC
Melbourne, 3086
Australia
Phone: 61 0426188766
Email: danilo110190@hotmail.com

Abstract

Background: Patellofemoral pain (PFP) impairs joint- and health-related quality of life and may be associated with knee osteoarthritis. We developed a novel, 2-phase, stepped-care approach for PFP, combining (1) self-directed web-based education and exercise therapy with (2) physiotherapist-supported education and exercise therapy. Physiotherapy sessions can be provided using 2 different modalities: face-to-face and telerehabilitation.

Objective: This study aims to (1) determine the feasibility of our stepped-care approach, (2) explore patient-reported outcomes following self-directed web-based education and exercise therapy in people with PFP (phase 1), and (3) estimate the differences in treatment effects between face-to-face and telerehabilitation to support further education and exercise therapy (phase 2) in those who had not completely recovered following self-directed care.

Methods: Phase 1 involved 6 weeks of self-directed web-based education and exercise therapy. Phase 2 involved random allocation to a further 12 weeks of physiotherapist-led (up to 8 sessions) education and exercise therapy delivered face-to-face or via telerehabilitation to participants who did not rate themselves as completely recovered following phase 1. Feasibility indicators of process, adherence, and participant retention were collected as primary outcomes alongside patient-reported outcomes on Global Rating of Change and knee pain, disability, knee-related quality of life, pain catastrophism, kinesiophobia, and knee self-efficacy. All participants were assessed at baseline, 6 weeks, and 18 weeks.

Results: A total of 71 participants were screened to identify 35 participants with PFP to enter the study. Overall, 100% (35/35) and 88% (31/35) of the participants were followed up with at 6 and 18 weeks, respectively. In phase 1 of the study, participants accessed the My Knee Cap website for an average of 6 (7.5) days and performed the exercises for an average of 2.5 (3.6) times per week. A total of 20% (7/35) of the participants reported that they had completely recovered at 6 weeks. Furthermore, 93% (26/28) of the participants who were followed up and had not completely recovered at 6 weeks agreed to be enrolled in phase 2. No statistically significant differences were found between the face-to-face and telerehabilitation groups for any outcome. The novel stepped-care approach was associated with marked improvement or complete recovery in 40% (14/35) of the participants following phase 1 and 71% (25/35) of the participants following phase 2.

Conclusions: Self-directed web-based education and exercise therapy for people with PFP is feasible, as noted by the high rate of participant retention and home exercise adherence achieved in this study. Furthermore, 20% (7/35) of people reported complete recovery at 6 weeks. Both face-to-face and telerehabilitation physiotherapy should be considered for those continuing to seek care, as there is no difference in outcomes between these delivery modes. Determining the efficacy of the stepped-care model may help guide more efficient health care for PFP.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e18584) doi:10.2196/18584

KEYWORDS

pain; internet; knee; rehabilitation; eHealth

Introduction

Background

Patellofemoral pain (PFP) is the most common knee condition experienced by young adults, affecting approximately one in four people worldwide [1]. PFP is characterized by diffuse anterior knee pain during activities that load the patellofemoral joint, including stair negotiation, running, and squatting [2,3]. PFP also impairs joint- and health-related quality of life [4]; reduces physical function [2]; results in fear of movement [5]; alters movement patterns [6,7]; and is associated with manifestations of peripheral and central sensitization, including lower pressure pain threshold and facilitated temporal summation of pain [8].

Exercise therapy is considered the cornerstone treatment for PFP [9-11]. When provided with or without other treatments (manual therapy, taping, and bracing), exercise therapy reduces pain, improves function, and leads to greater recovery rates in the short- and long-term compared with a placebo or *wait and see* approach (when participants do not receive any intervention over the study period) [9,12,13]. However, 57% of people with PFP report unfavorable outcomes 5 to 8 years following treatment [14]. Patient education for PFP is considered a key treatment by international experts to optimize self-management and long-term outcomes [9,11,15]. Our recent systematic review indicates that patient education may lead to similar short-term outcomes for pain and function as exercise therapy in PFP, but it is under researched [16]. During the development of an education leaflet for people with PFP, patients consistently requested a dedicated website to facilitate patient education [17]. Web-based interventions might facilitate self-directed education and exercise therapy to reduce health care–related costs at the system and personal levels [18]. Self-directed web-based interventions for low back pain [19,20] and osteoarthritis [21] can lead to improved patient-reported outcomes, including disability and perceived benefit compared with the *wait and see* approach. The impact of self-directed web-based interventions for people with PFP is unknown.

Implementation of a stepped-care model for PFP could benefit both patients and health systems. Stepped care can be defined as a staged evidence-based system comprising hierarchically delivered interventions, from the least to the most intensive, linked to patients' needs [22,23]. The goal is to provide effective care with the least intensive treatment. Self-directed web-based care may reduce the burden of health professional consultations, but it is unlikely to benefit or satisfy all people with PFP. Additional supported education and exercise therapy may provide better outcomes in those continuing to seek care because of persistent pain. Education and exercise therapy can be provided via face-to-face or telerehabilitation delivery. Telerehabilitation produces similar benefits as the face-to-face delivery across several chronic conditions [19,24,25], including reports of no difference in pain and function when compared with face-to-face rehabilitation following a total knee replacement [26]. No study has compared telerehabilitation with face-to-face delivery of physiotherapy care in people with PFP.

Objectives

This study was designed to (1) evaluate the feasibility of a stepped-care approach for people with PFP, (2) explore the effect of self-directed web-based education and exercise therapy on the perceived recovery and clinical outcomes of people with PFP (phase 1), and (3) estimate the differences in treatment effects between face-to-face and telerehabilitation delivery of physiotherapy to support further education and exercise therapy (phase 2) in those who did not completely recover following initial self-directed care.

Methods

Reporting, Registry, and Ethics Approval

This study was reported in accordance with the Consolidated Standards of Reporting Trials statement [27]. The study was approved by the La Trobe University Human Ethics Committee (process number: HEC17-102), and all participants provided written and verbal informed consent. We used 2 separate consent forms for phases 1 and 2 of our study. The protocol was a priori registered and approved by the Australian New Zealand Clinical Trials Registry (ACTRN12618000224224).

Participants

Participants with PFP aged between 18 and 40 years were recruited using advertisements at La Trobe University and gyms of Melbourne (Australia) and on social media (Facebook, blogs, and Twitter) between February 26 and July 1, 2018.

Eligibility criteria were based on a consensus statement on terminology, definitions, and clinical examination of people with PFP [2]. The following eligibility criteria were assessed by an experienced (>6 years) physiotherapist from the research group. Participants were included if they presented anterior or retropatellar pain (1) corresponding to at least thirty on a 100-mm visual analog scale (VAS) in the previous week; (2) for at least 3 months; or (3) during at least two or more activities from prolonged sitting, squatting, kneeling, running, ascending and descending stairs, jumping, and landing. Exclusion criteria included a history of any lower limb surgery, history of patellar subluxation or dislocation, ligament or meniscus tears assessed clinically, presence of neurological diseases, or individuals who had received oral steroids and opiate treatment in the last month [28].

Procedures

This study was designed to evaluate the feasibility of a stepped-care approach, including a randomized, clinical, single-center trial, comparing physiotherapy delivery modes. The novel stepped-care approach evaluated consists of 2 phases.

XSL•FO
RenderX

Phase 1 involved a pre-post design where all participants received 6 weeks of self-directed web-based education and exercise therapy.

Phase 2 involved a parallel-group randomized clinical trial (RCT). Participants were not informed of the existence of phase 2 when entering phase 1. Participants who were offered inclusion in phase 2 were those who did not rate themselves as *completely recovered* on the Global Rating of Change (GROC) 6-item Likert scale following phase 1. Each patient was offered the option to receive further 12 weeks of physiotherapy, which involved either face-to-face or telerehabilitation delivery, allocated randomly. All participants were reassessed after phase 2 (ie, at 18 weeks). Participants not receiving treatment during phase 2 (ie, those who *completely recovered* after 6 weeks) were also followed up at 18 weeks.

Randomization and Blinding in Phase 2

A member of the research team not involved in data collection generated randomization lists (block randomization; block size of 4-6) using a random number generator on the website (sealedenvelope.com). Group allocations were concealed using sequentially numbered, sealed opaque envelopes, which were opened by one member of the research team not involved in data collection or randomization following baseline assessment.

Assessor

The assessor was blinded to participant group allocation.

Participants

Participants were told that they would be randomly allocated to one of the two different education and exercise therapy treatments. Once randomized, care was taken to ensure that the participants were unaware of the details of the alternative program (delivery mode). Participants were also instructed not to reveal the details of their allocated intervention to the blinded outcome assessor at the final follow-up (18 weeks).

Therapist

A total of 5 registered physiotherapists (>5 years of clinical experience) delivered the physiotherapy treatment—3 physiotherapists delivered the intervention in face-to-face mode and 2 physiotherapists delivered telerehabilitation via Skype. Precautions were taken to ensure that the physiotherapists were unaware of the existence of another intervention group, with physiotherapists believing that they were participating in a case series design to evaluate the feasibility of the respective intervention they delivered.

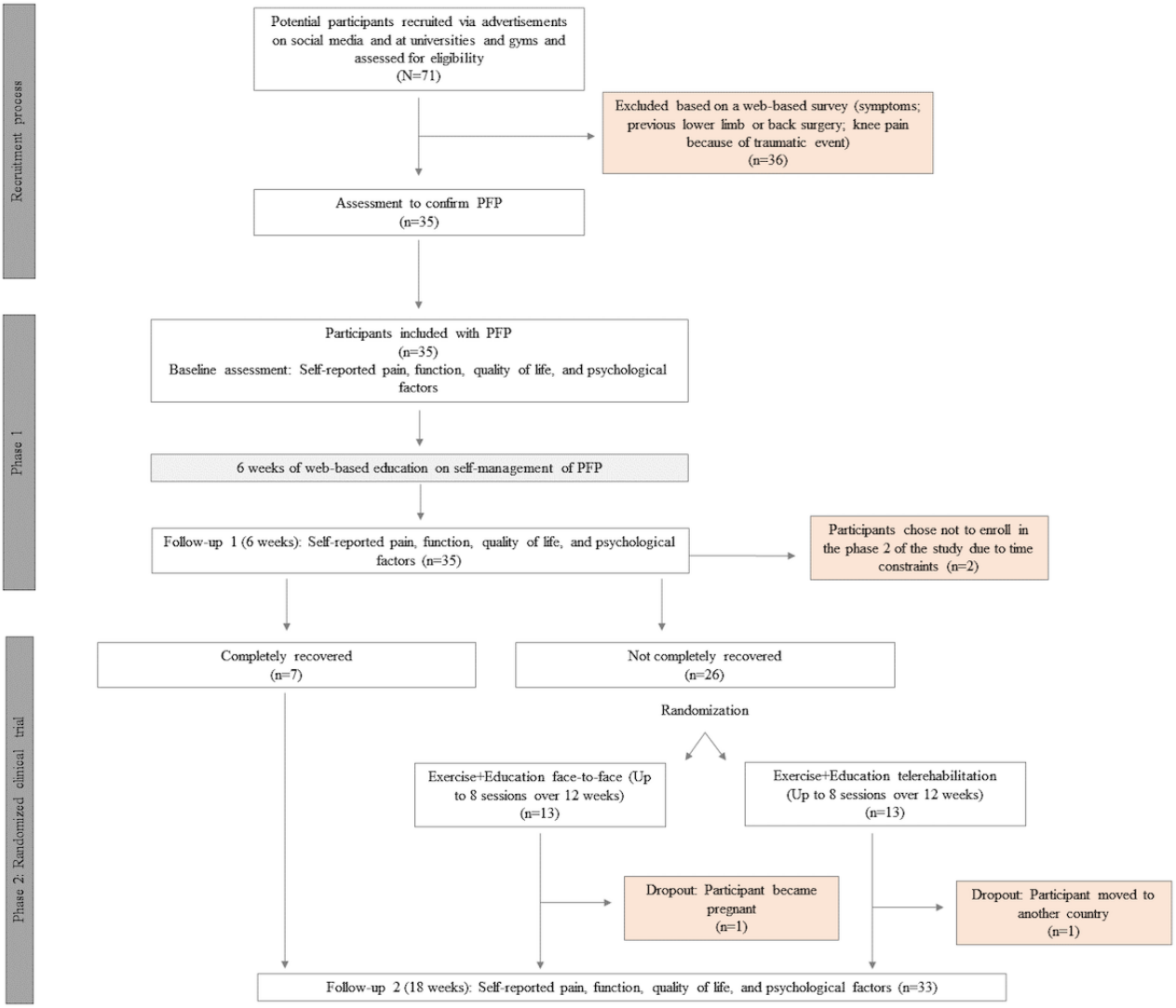
Outcomes

The baseline, 6-week, and 18-week follow-up face-to-face assessments (Figure 1) were performed by the same blinded assessor at La Trobe Sports and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia.

All participants were asked to attend one 60-min initial screening and assessment at the university's physiotherapy department. All questionnaires were self-administered, and participants completed the questionnaires independently on paper. Before starting the questionnaires, participants were asked to report their age in years. Body mass and height were measured using a scale and a measuring tape fixed on the wall, respectively. BMI was then calculated as weight in kilograms divided by height in meters squared. Participants also reported their symptom duration (months).

XSL•FO
RenderX

Figure 1. Flow chart of the study. PFP: patellofemoral pain.



Primary Outcome

Recruitment rates were recorded and defined as the number of participants recruited each week. The consent rate was calculated by dividing the number of people who met the inclusion criteria by the number of people who consented to participate in both phases 1 and 2.

Adherence to phase 1 was monitored by assessing the number of accesses to the *My Knee Cap* website, monitored through an algorithm embedded in the website. Participants also completed an exercise log daily for 6 weeks to indicate the days they performed the exercises available on the website. Adherence to phase 2 was monitored through the number of appointments scheduled with the physiotherapist for each group. In addition, all participants received a logbook to add information regarding how many times they accessed the website and performed the exercises each week and to record any adverse events (defined as negative consequences of care that result in unintended injury or illness that may or may not have been preventable).

Primary Estimates of the Treatment Effect

GROC

Participants rated their perceived recovery at 6 weeks (phase 1) and 18 weeks (phase 2) following trial commencement on a 6-item Likert scale: (1) completely recovered, (2) markedly better, (3) moderately better, (4) same, (5) moderately worse, and (6) markedly worse. GROC is highly reliable, easily understandable, sensitive to change, and the preferred measure of pain and disability among people with PFP [29]. It also moderately correlates with pain and disability and strongly correlates with health-related quality of life changes [30].

Pain (VAS)

The worst knee pain intensity during the previous week was assessed on a 100-mm VAS, with 0 indicating *no pain* and 100 indicating *the worst pain possible* [31]. The minimal clinically important difference for people with PFP is 20 mm [32]. The VAS has been validated for people with PFP, and it has been reported to demonstrate high test-retest reliability [32].

XSL•FO
RenderX

Secondary Estimates of the Treatment Effect

Disability

The Anterior Knee Pain Scale (AKPS) is a self-reported 13-item questionnaire, with scores ranging from 0 to 100 and higher values indicating better knee function [33]. The minimal clinically important difference for people with PFP is 10 points [32]. AKPS has been validated for people with PFP and has been reported to demonstrate high test-retest reliability [32].

Knee-Related Quality of Life

The Knee Injury and Osteoarthritis Outcome Score-Quality of Life subscale is a self-reported 4-item subscale with responses ranging from 0 to 4. Responses are transformed to a scale of 0 to 100, with 0 representing extreme knee problems and 100 representing no knee problems at all [34].

Pain Catastrophizing

The Pain Catastrophizing Scale (PCS) is a self-reported 13-item questionnaire to describe the various thoughts and feelings a person might experience when they are in pain. Scores range from 0 to 52, with higher values indicating more severe catastrophic thoughts about pain [35]. PCS has been previously validated for people with musculoskeletal conditions and has high test-retest reliability [36].

Kinesiophobia

The Tampa Scale for Kinesiophobia (TSK) is a self-reported 17-item questionnaire to quantify fear of movement. Scores range from 17 to 68, with higher values indicating greater fear [37,38]. TSK has been previously validated for people with musculoskeletal conditions and has moderate test-retest reliability [37].

Knee Self-Efficacy

The Knee Self-Efficacy Scale (K-SES) is a self-reported 22-item questionnaire where the participant rates the certainty about the capability of performing an activity, despite pain/discomfort, on an 11-point Likert scale, ranging from 0 (not at all certain) to 10 (very certain). The sum of the item scores is calculated and then divided by the number of items, yielding a total K-SES score ranging from 0 to 10. Higher scores indicate greater knee self-efficacy.

Adverse Events

Participants were given a specific form to record any adverse symptoms or events they experienced during the study (eg, pain flares, muscle soreness, and injury unrelated to the trial). No instructions were provided regarding the intensity of pain that would characterize an adverse event.

Intervention

Phase 1 (0 to 6 Weeks)

During this phase, all participants received the same intervention: access to the *My Knee Cap* website for 6 weeks. The website provides education to participants with PFP on self-management of their knee pain based on current evidence [10,11,15], along with a self-directed exercise therapy program. The website was designed by the authors DS and CB, with the assistance of a web designer, to be both didactic and engaging, with a simple user interface and navigation appropriate for patients with low health literacy. We asked for feedback on the content, language, and functionality of the website from one layperson diagnosed with PFP (end user), one physiotherapist experienced in managing PFP, and one researcher with a strong track record of publications in PFP. Refinements were made to the website based on this feedback before the commencement of the study.

The website provides public access to 4 main sections: (1) understanding your pain, (2) treatment options, (3) exercise program, and (4) patient stories.

The *understanding your pain* section provides information covering a variety of subjects, including diagnosis, prognosis, incidence and prevalence, noisy knees, fear of movement, self-management of exercise load, and self-management of pain.

The *treatment options* section includes treatment options that can be beneficial (taping/bracing, foot orthosis, and exercises) and information about common treatments with inconsistent evidence (knee surgery, ultrasound, and platelet-rich plasma).

The *exercise program* section is focused on 4 types of exercises targeting the trunk, hip, and knee muscles, based primarily on a previously published exercise therapy trial. The participants had access to videos demonstrating the exercises proposed in the *exercise program* section. Explanations about how to progress the exercises were also provided (Multimedia Appendix 1).

The *patient stories* section presents the stories of 2 patients (1 woman and 1 man), from a private physiotherapy clinic in Melbourne, with knee pain who had successful outcomes after engaging in education and exercise therapy.

All information on the website was provided in plain language, with multimedia resources such as infographics, animated videos, and podcasts, to facilitate engagement and understanding. After the baseline assessment, a physiotherapist introduced the content of the website and gave the password of the exercise videos that were included on the website during a 30-min orientation session. In addition, participants were asked to complete one exercise of each type 3 times a week for a period of 6 weeks and access the website as often as needed.

Phase 2 (6 to 18 Weeks)

Participants who reported themselves to be *completely recovered* did not receive further treatment during phase 2. Participants who reported not being completely recovered were randomized to 12 weeks of face-to-face– or telerehabilitation-delivered education and exercise therapy with a physiotherapist.

Participants were permitted up to 8 sessions with their physiotherapists to receive guidance on education and exercise therapy. The upper limit of physiotherapy sessions was determined through a previous feasibility study [39], high-quality RCTs [13], and discussion with physiotherapists providing our intervention. To reflect clinical practice, we did not specify a minimum number of sessions and at what exact time points physiotherapy sessions should occur. Typically, physiotherapy sessions were closer together in the early stages

XSL•FO
RenderX

(eg, every 1-2 weeks) and spread out toward the end of the 12 weeks (eg, monthly), as participants became more confident with their home exercise programs and knowledgeable about the condition. Education and exercise therapy that were provided were similar to what was available on the website, and the website was used to reinforce educational messages and exercise where relevant. The exercise therapy program and education content can be found on the *My Knee Cap* website and Multimedia Appendix 1. Physiotherapists were permitted to provide additional exercise input, including gym-focused progression, if deemed appropriate. When required, gym memberships were provided.

Statistical Analyses

Statistical analyses were performed using SPSS version 23 (IBM, SPSS Inc). The level of significance was set a priori at $P<.05$. All statistical analyses were conducted by an investigator blinded to the group allocation.

Phase 1

Global perceived recovery after 6 weeks of self-directed web-based education and exercise therapy was reported descriptively (percentages). Within-group changes in primary and secondary estimates of effects were evaluated using paired *t* tests. Effect sizes (ESs) were calculated to guide the interpretation of the power of the comparison. ES values were defined as small (0.2-0.5), medium (0.51-0.80), or large (≥0.81) [40]. Correlations of exercise adherence with changes in patient-reported outcomes were performed using Pearson coefficient correlations.

Phase 2

Demographic data from the 2 groups (face-to-face and telerehabilitation) at 6 weeks were compared using independent *t* tests. Intention-to-treat analyses were used for all outcomes. We performed a 2×5 chi-square test for independence to investigate whether there was an association between GROC outcomes and the groups to which the participants were allocated at the end of 18 weeks (face-to-face and telerehabilitation groups). Independent *t* test and ES were used to compare patient-reported outcomes between the face-to-face and telerehabilitation groups. Using the same criteria as phase 1, ES values were defined as small (0.2-0.5), medium (0.51-0.80), or large (≥0.81) [40].

Results

Recruitment and Feasibility

Between February 2018 and July 2018, 35 participants (27 women and 8 men) with PFP were recruited from 71 potential candidates. The recruitment rate was 2 participants per week over an 18-week period, with a consenting rate of 100% (35/35). Overall, 77% (27/35) of the participants included were recruited via advertisements on social media, 11% (4/35) via advertisements at La Trobe University, and 11% (4/35) via advertisements at gyms in the neighborhood of La Trobe University. The trial was completed in December 2018, with 100% (35/35) of the participants followed up at 6 weeks and 88% (31/35) followed up at 18 weeks. For phase 2, the consenting rate was 93% (26/28), with 2 participants choosing not to enter because of time constraints, despite ongoing symptoms (Figure 1).

Adherence

Phase 1

All participants accessed the website at least once, with web usage data indicating that the participants accessed the website for an average of 6 (7.5) days over the 6 weeks (Multimedia Appendix 2). Exercises were reported to be performed on an average of 15 (12.2) days over 6 weeks, equating to an average of 2.5 (3.6) times per week (18/35, 52% of the participants completed the exercises on an average of at least three times per week). Low to moderate correlations were found for exercise adherence (total number of days on which exercise was performed) with self-reported pain (r=0.33; P=.03), knee-related quality of life (r=−0.52; P=.001), self-reported function (r=−0.37; P=.01), pain catastrophizing (r=0.52; P=.001), kinesiophobia (r=0.39; P=.009), and knee self-efficacy (r=−0.51; P=.001).

Phase 2

Participants enrolled in phase 2 (Multimedia Appendix 3) performed an average of 4.5 (1.5) physiotherapy sessions in the face-to-face group and an average of 5.2 (2.1) physiotherapy sessions in the telerehabilitation group.

Adverse Events

Phase 1

One participant reported an adverse event unrelated to the trial. Furthermore, 11 participants reported knee pain flares or muscle soreness of low intensity while performing the exercises, with most occurring in the first week (Multimedia Appendix 2) and no impact on participation in future exercise sessions.

Phase 2

Two adverse events unrelated to the trial were reported, one from each group. One participant fell on their knees while running to catch a bus, but this event did not alter the treatment. The other participant fell on a staircase, which led to a few scratches on her knee, delaying physiotherapy treatment for 2 weeks. No adverse event related to the trial was reported because of physiotherapy consultations or exercise therapy during phase 2.

Secondary Outcomes

Phase 1

After 6 weeks of self-directed web-based education and exercise therapy, 20% (7/35) of the participants starting the trial reported that they had *completely recovered*, 20% (7/35) were markedly better, 40% (14/35) were moderately better, 17% (6/35) were same, and 3% (1/35) were moderately worse. Patient-reported outcomes at baseline and after 6 weeks are described in Table 1. Large improvements across the cohort occurred for worst knee pain in the previous week (ES=1.04), knee-related quality of life (ES=−1.13), disability (ES=−0.89), pain catastrophizing (ES=1.21), and knee self-efficacy (ES=−1.01), along with a medium improvement in kinesiophobia (ES=0.78).

XSL•FO
RenderX

Table 1. Comparisons between patient-reported outcomes at baseline and after the 6-week self-directed web-based education and exercise therapy.

Outcomes	At baseline, mean (SD)	At 6 weeks, mean (SD)	Mean difference (95% CI)
Worst knee pain (visual analog scale[a], range: 0-100)	58.03 (17.54)	29.39 (25.52)	28.64 (18.88-38.39)
Knee-related quality of life (Knee Injury and Osteoarthritis Outcome Score-Quality of Life subscale[b], range: 0-100)	41.86 (20.28)	66.67 (21.69)	−24.81 (−32.60 to −17.02)
Disability (Anterior Knee Pain Scale[b], range: 0-100)	69.27 (13.41)	82.36 (13.87)	−13.09 (−18.34 to −7.85)
Pain catastrophizing (Pain Catastrophizing Scale[a], range: 0-52)	21.73 (11.91)	10.09 (9.14)	11.64 (8.22-15.05)
Kinesiophobia (Tampa Scale for Kinesiophobia[a], range: 17-68)	38.48 (6.14)	33.03 (6.83)	5.46 (2.99-7.92)
Knee self-efficacy (Knee Self-Efficacy Scale KSE-S[b], range: 0-10)	5.57 (1.71)	7.28 (1.53)	−1.71 (−2.31 to −1.11)

[a]Higher scores indicate worse condition.

[b]Lower values indicate worse condition.

Phase 2

A total of 26 participants entered phase 2 and were randomly allocated to face-to-face– (n=13) or telerehabilitation-delivered education and exercise therapy (n=13). No between-group differences were found for demographics between these 2 groups (Table 2), and no between-group differences were found for any secondary estimates of treatment effect (Table 3). Both groups had one dropout during the intervention period, with one participant becoming pregnant and another moving to a different country.

The participants' GROC outcomes after 18 weeks are shown in Figure 2. There was no significant difference in the proportion of GROC outcomes between the face-to-face and telerehabilitation groups (χ^2_5=2.03; *P*=.73).

No significant difference between face-to-face– and telerehabilitation-delivered education and exercise therapy was found for any patient-reported outcomes (Table 3). The between-group differences for the patient-reported outcomes had a moderate effect on worst knee pain in the previous week (ES=0.72) and a small effect on knee-related quality of life (ES=0.22), disability (ES=0.16), pain catastrophizing (ES=0.33), kinesiophobia (ES=0.22), and knee self-efficacy (ES=0.16). Overall, 71% (25/35) of all participants entering the study reported having completely recovered or markedly improved following phase 2.

Table 2. Demographic characteristics of face-to-face and telerehabilitation groups.

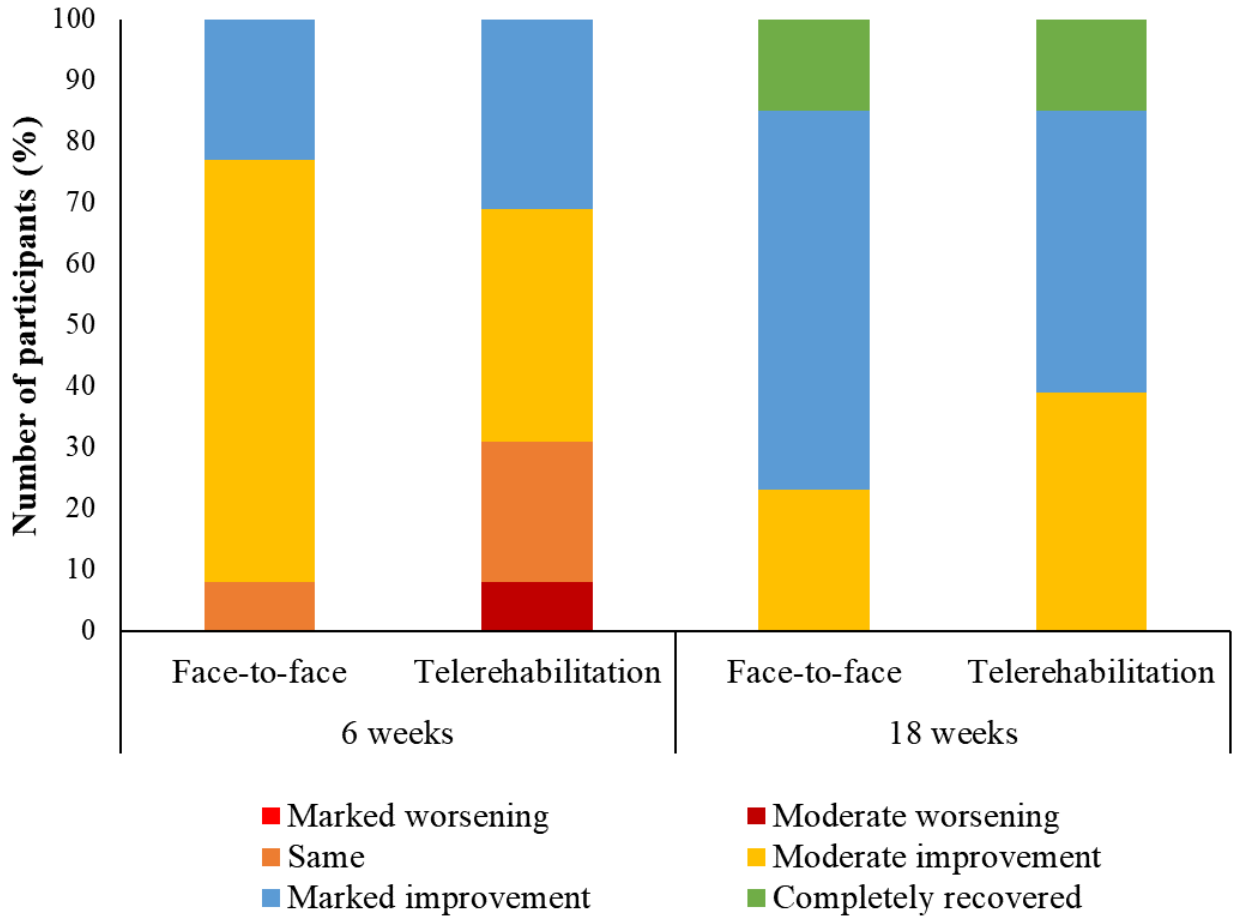
Variables	Face-to-face group, mean (SD)	Telerehabilitation group, mean (SD)	*P* value
Age (years)	32 (6)	31 (6)	.62
BMI (kg/m^2)	27.58 (6.71)	26.46 (8.97)	.59
Symptoms duration (months)	53 (55)	32 (29)	.14

XSL•FO
RenderX

Table 3. Between-group comparisons for patient-reported outcomes at 6 and 18 weeks.

Outcomes	Face-to-face group, mean (SD)	Telerehabilitation group, mean (SD)	Mean difference (95% CI)
Worst knee pain (visual analog scale, range: 0-100)			
6 weeks	32.31 (22.51)	42.31 (23.15)	−10.00 (−26.36 to 6.36)
18 weeks	18.46 (18.46)	33.08 (22.13)	−14.62 (−31.25 to 2.02)
Knee-related quality of life (Knee Injury and Osteoarthritis Outcome Score-Quality of Life subscale, range: 0-100)			
6 weeks	63.94 (16.76)	56.73 (21.87)	7.21 (−7.27 to 21.69)
18 weeks	68.27 (21.57)	73.08 (21.71)	−4.81 (−20.56 to 10.95)
Disability (Anterior Knee Pain Scale, range: 0-100)			
6 weeks	78.54 (12.61)	77.69 (12.94)	0.85 (−8.34 to 10.03)
18 weeks	86.54 (10.59)	84.62 (13.49)	1.92 (−6.99 to 10.84)
Pain catastrophizing (Pain Catastrophizing Scale, range: 0-52)			
6 weeks	8.23 (6.92)	14.46 (10.26)	−6.23 (−13.70 to 0.76)
18 weeks	5.92 (6.02)	8.54 (9.44)	−2.62 (−8.29 to 3.06)
Kinesiophobia (Tampa Scale for Kinesiophobia, range: 17-68)			
6 weeks	34.15 (7.72)	35.31 (5.01)	−1.15 (−6.07 to 3.76)
18 weeks	30.38 (7.12)	31.77 (5.43)	−1.39 (−6.07 to 3.30)
Knee self-efficacy (Knee Self-Efficacy Scale, range: 0-10)			
6 weeks	6.95 (0.93)	6.74 (1.75)	0.21 (−0.84 to 1.25)
18 weeks	7.64 (1.45)	7.36 (1.98)	0.28 (−1.08 to 1.65)

Figure 2. Global Rating of Change outcomes of the face-to-face and telerehabilitation groups.



XSL•FO
RenderX

Discussion

Summary of the Findings

This study confirms the feasibility of evaluating the efficacy of a stepped-care approach, including initial self-directed web-based education and exercise therapy, followed by physiotherapist-led education and exercise therapy. Overall, 35 participants with PFP were recruited from 71 potential candidates, with a recruitment rate of 2 participants per week and all eligible participants consenting to enter the trial. The follow-up rate was high for both phases of the trial, including 100% at 6 weeks and 88% at 18 weeks.

Six weeks of self-directed web-based education and exercise therapy was associated with 1 in 5 participants with PFP rating themselves as *completely recovered*. Self-directed care was associated with moderate to large improvements in pain, disability, kinesiophobia, pain catastrophizing, knee self-efficacy, and knee-related quality of life. Self-reported adherence to exercise (number of days completed) during phase 1 was related to clinical improvements in pain, function, and psychological outcomes. Although not significant, findings reported by Rathleff et al [41] in adolescents with PFP also indicate a link between greater exercise adherence and improved outcomes. In addition, a systematic review evaluating the influence of exercise dose on outcomes in knee osteoarthritis has reported that a greater number of exercise sessions per week is related to greater improvements in pain and function [42]. An interesting finding from our study is that participants reduced the number of exercise sessions they completed and the frequency in which they accessed the website during phase 1. Including consumers in the further development of the website may help to improve website engagement and exercise adherence.

Beneficial outcomes for pain and disability following self-directed web-based education and exercise therapy in this study are consistent with previous RCTs reporting improvements in the same outcomes following web-based interventions for low back pain [19] and hip and knee osteoarthritis [21,43]. Clinical improvements for the 20% (7/35) of participants who reported complete recovery at 6 weeks were sustained at 18 weeks, highlighting the potential for sustained recovery through self-directed care in people with PFP. Further research comparing self-directed web-based education and exercise therapy with a *wait and see* approach will help to determine efficacy and potential to reduce the current burden of overtreatment in health care systems around the world [44].

Overall, 4 out of 5 participants did not report complete recovery following self-directed web-based education and exercise therapy, indicating that some people may require or benefit from additional care. Face-to-face and telerehabilitation delivery of physiotherapy following initial self-directed care produced only small differences for all but one patient-reported outcome. The exception was for pain reduction, which was moderately greater (ES=0.72) in the face-to-face group, possibly because of differences exceeding 10 mm between groups at 6 weeks. Similar outcomes between face-to-face and telerehabilitation delivery of physiotherapy in this study are consistent with a large (n=201) high-quality noninferiority RCT in people following total knee joint replacement [26]. Thus, the physiotherapy delivery mode among people with knee pathologies may have minimal influence on patient-reported outcomes. Increasing the availability of physiotherapy care delivered via telerehabilitation is recommended to improve access to people with limited mobility, distressing symptoms, and/or inability to access center-based programs [18]. In addition, telerehabilitation approaches to physiotherapy practice could reduce the need for physiotherapy facilities requiring consultations and work-related travel for physiotherapists themselves.

Our findings indicate that 31% (11/35) of the participants reported low-intensity pain flares or muscle soreness while performing the exercises, with most occurring in the first week. This finding is not surprising, considering that participants need to initially learn new exercise skills, including how to manage and increase exercise loads without increasing knee pain. Reductions in the frequency of pain flares over the 6 weeks of self-directed care found in this study may be the result of participants improving their ability to independently manage exercise loads or ongoing active management leading to improved knee self-efficacy, which is typically impaired in people with PFP [45,46].

The overall proportion of participants reporting marked improvement or complete recovery across the cohort increased from 40% (14/35) at 6 weeks to 71% (25/35) at 18 weeks. High-quality RCTs, including long-term follow-up, are now needed to test this stepped-care approach against usual care or current best care [10] to evaluate efficacy, effectiveness, and cost-effectiveness. Findings from our study can be used to inform larger studies comparing face-to-face mode with telerehabilitation mode or investigating our stepped-care approach for people with PFP, which will allow more definitive conclusions.

Limitations and Future Directions

Before making definitive recommendations regarding its potential value, our novel stepped-care approach requires comparison with a *wait and see* approach in both phases. Specifically, it is unclear whether self-directed care facilitated by our platform is superior to *wait and see* approach or whether any mode of physiotherapy following self-directed care can improve outcomes. Nonetheless, our findings suggest that if a self-directed web-based education and exercise therapy is provided as a first-line treatment, 1 in 5 patients may be able to independently manage their condition without the need for additional health professional referral.

Adherence seems to be related to better patient-reported outcomes. Future trials should develop strategies to optimize adherence to interventions (eg, digital support). In addition, the web-based platform was created with limited co-design processes [47]. Further modifications and improvements are currently being guided by additional qualitative research with consumers and physiotherapists and by using tools such as Health on the Net Foundation Code and DISCERN. Low computer and health literacy have been reported as barriers to improvement following other web-based interventions for

XSL•FO
RenderX

chronic musculoskeletal conditions [48,49]. Our findings may not be applicable for people with low computer literacy skills. Further evaluation of efficacy, barriers, and cost-effectiveness of self-directed use of the website evaluated in this study, compared with usual care, is needed to inform potential implementation.

Although we measured the number of physiotherapy sessions attended during phase 2, we did not measure adherence to exercise, subjectively or objectively. Considering conflicting findings related to the importance of exercise adherence among people with knee pain [42,50], further evaluation of the importance of exercise adherence in PFP is strongly encouraged. In addition, approaches to treatment may have varied across the 5 physiotherapists providing care in our trial. However, all physiotherapists followed the same structure and content for education and exercise, guided by the *My Knee Cap* web-based platform. Each had a minimum of 5 years of clinical experience and received training and ongoing support as required, facilitated by an experienced physiotherapist (15 years) researcher involved in the study. Finally, we only included individuals with PFP aged between 18 and 40 years because of the high prevalence of PFP in this population [1], limiting the extrapolation of findings to adolescents and older adults with PFP.

Conclusions

This study confirms the feasibility of evaluating the efficacy of a stepped-care approach, including initial self-directed web-based education and exercise therapy, followed by physiotherapist-led education and exercise therapy. Self-directed web-based education and exercise therapy were associated with 1 in 5 participants with PFP rating themselves as completely recovered at 6 weeks and having large improvements in pain. An additional 12 weeks of physiotherapy provided face-to-face or via telerehabilitation to support education and exercise therapy was associated with 71% (25/35) of participants reporting to be completely recovered or markedly improved. The absence of differences in outcomes between face-to-face and telerehabilitation delivery modes indicates that either mode could be considered depending on patient preference and need. Evaluating the efficacy of self-directed web-based education and exercise therapy as a stand-alone intervention and as part of a stepped-care model that includes additional physiotherapy care may help guide more efficient health care for people with PFP.

Acknowledgments

DS received a scholarship from São Paulo Research Foundation, process number 2016/11665-1. CB is supported by a Medical Research Future Fund, Translating Research into Practice Fellowship (APP1150439). The authors would like to acknowledge the support of the La Trobe University Social Research Platform for funding contributions to help refine the resources on the *My Knee Cap* website used in this study. The financial sponsors played no role in the design, execution, analysis, and interpretation of data or writing of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exercise-therapy program.
[PDF File (Adobe PDF File), 2007 KB - jmir_v22i7e18584_app1.pdf]

Multimedia Appendix 2
Website usage and exercise adherence (phase 1).
[DOCX File , 43 KB - jmir_v22i7e18584_app2.docx]

Multimedia Appendix 3
Exercise adherence (phase 2).
[DOCX File , 15 KB - jmir_v22i7e18584_app3.docx]

References

1. Smith BE, Selfe J, Thacker D, Hendrick P, Bateman M, Moffatt F, et al. Incidence and prevalence of patellofemoral pain: a systematic review and meta-analysis. PLoS One 2018;13(1):e0190892 [FREE Full text] [doi: 10.1371/journal.pone.0190892] [Medline: 29324820]
2. Crossley KM, Stefanik JJ, Selfe J, Collins NJ, Davis IS, Powers CM, et al. 2016 patellofemoral pain consensus statement from the 4th international patellofemoral pain research retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br J Sports Med 2016 Jul;50(14):839-843 [FREE Full text] [doi: 10.1136/bjsports-2016-096384] [Medline: 27343241]

XSL•FO
RenderX

3. Briani RV, Pazzinatto MF, De Oliveira Silva D, Azevedo FM. Different pain responses to distinct levels of physical activity in women with patellofemoral pain. Braz J Phys Ther 2017;21(2):138-143 [FREE Full text] [doi: 10.1016/j.bjpt.2017.03.009] [Medline: 28460712]
4. Coburn SL, Barton CJ, Filbay SR, Hart HF, Rathleff MS, Crossley KM. Quality of life in individuals with patellofemoral pain: a systematic review including meta-analysis. Phys Ther Sport 2018 Sep;33:96-108. [doi: 10.1016/j.ptsp.2018.06.006] [Medline: 30059951]
5. Priore LB, Azevedo FM, Pazzinatto MF, Ferreira AS, Hart HF, Barton C, et al. Influence of kinesiophobia and pain catastrophism on objective function in women with patellofemoral pain. Phys Ther Sport 2019 Jan;35:116-121. [doi: 10.1016/j.ptsp.2018.11.013] [Medline: 30529861]
6. Powers CM, Witvrouw E, Davis IS, Crossley KM. Evidence-based framework for a pathomechanical model of patellofemoral pain: 2017 patellofemoral pain consensus statement from the 4th international Patellofemoral pain research retreat, Manchester, UK: part 3. Br J Sports Med 2017 Dec;51(24):1713-1723. [doi: 10.1136/bjsports-2017-098717] [Medline: 29109118]
7. de Oliveira SD, Barton CJ, Briani RV, Taborda B, Ferreira AS, Pazzinatto MF, et al. Kinesiophobia, but not strength is associated with altered movement in women with patellofemoral pain. Gait Posture 2019 Feb;68:1-5. [doi: 10.1016/j.gaitpost.2018.10.033] [Medline: 30408709]
8. De Oliveira SD, Rathleff MS, Petersen K, Azevedo FM, Barton CJ. Manifestations of pain sensitization across different painful knee disorders: a systematic review including meta-analysis and metaregression. Pain Med 2019 Feb 1;20(2):335-358. [doi: 10.1093/pm/pny177] [Medline: 30423181]
9. Willy RW, Hoglund LT, Barton CJ, Bolgla LA, Scalzitti DA, Logerstedt DS, et al. Patellofemoral pain. J Orthop Sports Phys Ther 2019 Sep;49(9):CPG1-CP95. [doi: 10.2519/jospt.2019.0302] [Medline: 31475628]
10. Collins NJ, Barton CJ, van Middelkoop M, Callaghan MJ, Rathleff MS, Vicenzino BT, et al. 2018 consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th international patellofemoral pain research retreat, gold coast, Australia, 2017. Br J Sports Med 2018 Sep;52(18):1170-1178. [doi: 10.1136/bjsports-2018-099397] [Medline: 29925502]
11. Lack S, Neal B, de Oliveira Silva SD, Barton C. How to manage patellofemoral pain - understanding the multifactorial nature and treatment options. Phys Ther Sport 2018 Jul;32:155-166. [doi: 10.1016/j.ptsp.2018.04.010] [Medline: 29793124]
12. Crossley K, Bennell K, Green S, Cowan S, McConnell J. Physical therapy for patellofemoral pain: a randomized, double-blinded, placebo-controlled trial. Am J Sports Med 2002;30(6):857-865. [doi: 10.1177/03635465020300061701] [Medline: 12435653]
13. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev 2015 Jan 20;1:CD010387. [doi: 10.1002/14651858.CD010387.pub2] [Medline: 25603546]
14. Lankhorst NE, van Middelkoop M, Crossley KM, Bierma-Zeinstra SM, Oei EH, Vicenzino B, et al. Factors that predict a poor outcome 5-8 years after the diagnosis of patellofemoral pain: a multicentre observational analysis. Br J Sports Med 2016 Jul;50(14):881-886. [doi: 10.1136/bjsports-2015-094664] [Medline: 26463119]
15. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'best practice guide to conservative management of patellofemoral pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med 2015 Jul;49(14):923-934. [doi: 10.1136/bjsports-2014-093637] [Medline: 25716151]
16. de Oliveira Silva D, Pazzinatto MF, Rathleff MS, Holden S, Bell E, Azevedo F, et al. Patient education for patellofemoral pain: a systematic review. J Orthop Sports Phys Ther 2020 Apr 29;50(7):388-396. [doi: 10.2519/jospt.2020.9400] [Medline: 32349640]
17. Barton CJ, Rathleff MS. 'Managing my patellofemoral pain': the creation of an education leaflet for patients. BMJ Open Sport Exerc Med 2016;2(1):e000086 [FREE Full text] [doi: 10.1136/bmjsem-2015-000086] [Medline: 27900163]
18. Holland AE. Telephysiotherapy: time to get online. J Physiother 2017 Oct;63(4):193-195 [FREE Full text] [doi: 10.1016/j.jphys.2017.08.001] [Medline: 28939309]
19. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
20. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]
21. Bossen D, Veenhof C, van Beek KE, Spreeuwenberg PM, Dekker J, de Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
22. Ho FY, Yeung W, Ng TH, Chan CS. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Sci Rep 2016 Jul 5;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]
23. March S, Donovan CL, Baldwin S, Ford M, Spence SH. Using stepped-care approaches within internet-based interventions for youth anxiety: three case studies. Internet Interv 2019 Dec;18:100281 [FREE Full text] [doi: 10.1016/j.invent.2019.100281] [Medline: 31890629]

XSL•FO
RenderX

24. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
25. Tsai LL, McNamara RJ, Moddel C, Alison JA, McKenzie DK, McKeough ZJ. Home-based telerehabilitation via real-time videoconferencing improves endurance exercise capacity in patients with COPD: the randomized controlled TeleR study. Respirology 2017 May;22(4):699-707 [FREE Full text] [doi: 10.1111/resp.12966] [Medline: 27992099]
26. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: a noninferiority randomized controlled trial. J Bone Joint Surg Am 2015 Jul 15;97(14):1129-1141. [doi: 10.2106/JBJS.N.01066] [Medline: 26178888]
27. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
28. de Oliveira Silva D, Barton C, Crossley K, Waiteman M, Taborda B, Ferreira AS, et al. Implications of knee crepitus to the overall clinical presentation of women with and without patellofemoral pain. Phys Ther Sport 2018 Sep;33:89-95. [doi: 10.1016/j.ptsp.2018.07.007] [Medline: 30059950]
29. Winters M, Holden S, Vicenzino B, Welton NJ, Caldwell DM, Lura CB, et al. Which treatment is most effective for patients with patellofemoral pain? A protocol for a living systematic review including network meta-analysis. BMJ Open 2018 Nov 18;8(11):e022920 [FREE Full text] [doi: 10.1136/bmjopen-2018-022920] [Medline: 30455387]
30. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163-170 [FREE Full text] [doi: 10.1179/jmt.2009.17.3.163] [Medline: 20046623]
31. Crossley K, Vicenzino B, Lentzos J, Schache A, Pandy M, Ozturk H, et al. Exercise, education, manual-therapy and taping compared to education for patellofemoral osteoarthritis: a blinded, randomised clinical trial. Osteoarthritis Cartilage 2015 Sep;23(9):1457-1464 [FREE Full text] [doi: 10.1016/j.joca.2015.04.024] [Medline: 25960116]
32. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? Arch Phys Med Rehabil 2004 May;85(5):815-822. [doi: 10.1016/s0003-9993(03)00613-0] [Medline: 15129407]
33. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy 1993;9(2):159-163. [doi: 10.1016/s0749-8063(05)80366-4] [Medline: 8461073]
34. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
35. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
36. Wheeler CH, Williams AC, Morley SJ. Meta-analysis of the psychometric properties of the pain catastrophizing scale and associations with participant characteristics. Pain 2019 Sep;160(9):1946-1953. [doi: 10.1097/j.pain.0000000000001494] [Medline: 30694929]
37. French DJ, France CR, Vigneau F, French JA, Evans TR. Fear of movement/(re)injury in chronic pain: a psychometric assessment of the original English version of the Tampa scale for kinesiophobia (TSK). Pain 2007 Jan;127(1-2):42-51. [doi: 10.1016/j.pain.2006.07.016] [Medline: 16962238]
38. de Souza FS, Marinho CS, Siqueira FB, Maher CG, Costa LO. Psychometric testing confirms that the Brazilian-Portuguese adaptations, the original versions of the Fear-Avoidance Beliefs Questionnaire, and the Tampa Scale of Kinesiophobia have similar measurement properties. Spine (Phila Pa 1976) 2008 Apr 20;33(9):1028-1033. [doi: 10.1097/BRS.0b013e31816c8329] [Medline: 18427325]
39. Barton CJ, de Oliveira Silva D, Patterson BE, Crossley KM, Pizzari T, Nunes GS. A proximal progressive resistance training program targeting strength and power is feasible in people with patellofemoral pain. Phys Ther Sport 2019 Jul;38:59-65. [doi: 10.1016/j.ptsp.2019.04.010] [Medline: 31055059]
40. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, USA: Routledge; 1988.
41. Rathleff MS, Graven-Nielsen T, Hölmich P, Winiarski L, Krommes K, Holden S, et al. Activity modification and load management of adolescents with patellofemoral pain: a prospective intervention study including 151 adolescents. Am J Sports Med 2019 Jun;47(7):1629-1637. [doi: 10.1177/0363546519843915] [Medline: 31095417]
42. Juhl C, Christensen R, Roos EM, Zhang W, Lund H. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. Arthritis Rheumatol 2014 Mar;66(3):622-636 [FREE Full text] [doi: 10.1002/art.38290] [Medline: 24574223]
43. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, et al. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain 2018 Sep;159(9):1833-1842. [doi: 10.1097/j.pain.0000000000001281] [Medline: 29794609]
44. Friedman DJ, Khan KM. Preventing overdiagnosis and the harms of too much sport and exercise medicine. Br J Sports Med 2019 Oct;53(20):1314-1318 [FREE Full text] [doi: 10.1136/bjsports-2018-100039] [Medline: 30518520]

XSL•FO
RenderX

45. Smith BE, Moffatt F, Hendrick P, Bateman M, Rathleff MS, Selfe J, et al. The experience of living with patellofemoral pain-loss, confusion and fear-avoidance: a UK qualitative study. BMJ Open 2018 Jan 23;8(1):e018624 [FREE Full text] [doi: 10.1136/bmjopen-2017-018624] [Medline: 29362256]
46. Priore LB, Lack S, Garcia C, Azevedo FM, de Oliveira Silva D. Two weeks of wearing a knee brace compared with minimal intervention on kinesiophobia at 2 and 6 weeks in people with patellofemoral pain: a randomized controlled trial. Arch Phys Med Rehabil 2020 Apr;101(4):613-623. [doi: 10.1016/j.apmr.2019.10.190] [Medline: 31830433]
47. Ward ME, de Brún A, Beirne D, Conway C, Cunningham U, English A, et al. Using co-design to develop a collective leadership intervention for healthcare teams to improve safety culture. Int J Environ Res Public Health 2018 Jun 5;15(6):1182 [FREE Full text] [doi: 10.3390/ijerph15061182] [Medline: 29874883]
48. Jeon Y, Flaherty I, Urban H, Wortley S, Dickson C, Salkeld G, et al. Qualitative evaluation of evidence-based online decision aid and resources for osteoarthritis management: understanding patient perspectives. Arthritis Care Res (Hoboken) 2019 Jan;71(1):46-55 [FREE Full text] [doi: 10.1002/acr.23572] [Medline: 29609208]
49. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
50. Nicolson PJ, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Effects of covertly measured home exercise adherence on patient outcomes among older adults with chronic knee pain. J Orthop Sports Phys Ther 2019 Jul;49(7):548-556. [doi: 10.2519/jospt.2019.8843] [Medline: 31258045]

Abbreviations

AKPS: Anterior Knee Pain Scale
ES: effect size
GROC: Global Rating of Change
K-SES: Knee Self-Efficacy Scale
PCS: Pain Catastrophizing Scale
PFP: patellofemoral pain
RCT: randomized clinical trial
TSK: Tampa Scale for Kinesiophobia
VAS: visual analog scale

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by N Glaviano, R Dodd; comments to author 26.03.20; revised version received 09.04.20; accepted 04.06.20; published 22.07.20.

Please cite as:
De Oliveira Silva D, Pazzinatto MF, Crossley KM, Azevedo FM, Barton CJ
Novel Stepped Care Approach to Provide Education and Exercise Therapy for Patellofemoral Pain: Feasibility Study
J Med Internet Res 2020;22(7):e18584
URL: http://www.jmir.org/2020/7/e18584/
doi:10.2196/18584
PMID:32706674

©Danilo De Oliveira Silva, Marcella F Pazzinatto, Kay M Crossley, Fabio M Azevedo, Christian J Barton. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Scheduled Telephone Support for Internet Cognitive Behavioral Therapy for Depression in Patients at Risk for Dropout: Pragmatic Randomized Controlled Trial

Satu Pihlaja[1*], LicPs; Jari Lahti[2*], PhD; Jari Olavi Lipsanen[2*], MA; Ville Ritola[1*], MA; Eero-Matti Gummerus[1*], MA; Jan-Henry Stenberg[1*], PhD; Grigori Joffe[1*], MD, PhD

[1]Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Hospital District of Helsinki and Uusimaa, Helsinki, Finland
[2]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[*]all authors contributed equally

Corresponding Author:
Grigori Joffe, MD, PhD
Department of Psychiatry
University of Helsinki and Helsinki University Hospital
Hospital District of Helsinki and Uusimaa
PO Box 590
Helsinki, 00029 HUS
Finland
Phone: 358 40 513 6500
Fax: 358 9 471 63815
Email: Grigori.joffe@hus.fi

Abstract

Background: Therapist-supported, internet-delivered cognitive behavioral therapy (iCBT) is efficient in the treatment of depression. However, the optimal mode and intensity of therapist support remain to be identified. Scheduled telephone support (STS) may improve adherence and outcomes but, as it is time- and resource-consuming, should be reserved for patients for whom the usual support may be insufficient.

Objective: This paper aims to reveal whether add-on STS for patients at risk of dropping out improves treatment adherence and symptoms in iCBT for depression.

Methods: Among patients participating in an ongoing large observational routine clinical practice study of iCBT for depression delivered nationwide by Helsinki University Hospital (HUS-iCBT), those demonstrating a ≥14-day delay in initiation of treatment received invitations to this subsidiary STS study. A total of 100 consenting patients were randomly allocated to either HUS-iCBT as usual (control group, n=50) or HUS-iCBT plus add-on STS (intervention group, n=50). Proportions of those reaching midtreatment and treatment end point served as the primary outcome; secondary outcomes were change in Beck Depression Inventory (BDI)–measured depressive symptoms and time spent in treatment.

Results: Add-on STS raised the proportion of patients reaching midtreatment compared with HUS-iCBT as usual (29/50, 58% vs 18/50, 36%; *P*=.045) and treatment end point (12/50, 24% vs 3/50, 6%; *P*=.02). Change in BDI score also favored add-on STS (3.63 points vs 1.1 points; *P*=.049), whereas duration of treatment did not differ.

Conclusions: Add-on STS enhances adherence and symptom improvement of patients at risk of dropping out of iCBT for depression in routine clinical practice.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) 55123131; http://www.isrctn.com/ISRCTN55123131.

(J Med Internet Res 2020;22(7):e15732) doi:10.2196/15732

KEYWORDS

internet CBT; depression; scheduled telephone support; adherence; routine clinical practice

XSL•FO
RenderX

Introduction

Despite the growing burden of mental health disorders, treatment is available for fewer than half of those in need [1]. Among mental health conditions, depressive disorders are some of the most prevalent [2]. As a leading cause of disability worldwide [3], they represent a major burden for health care systems [4].

For depressive disorders, recommended first-line treatments include psychological interventions, specifically psychotherapies [5]. Although psychotherapies are highly acceptable among clients [6,7], their availability is limited [8]. Some of the main challenges are shortage of professionals, perceived stigma, cost, and long distance to services [9].

Computer-based delivery of psychotherapy improves access to psychological treatments of depression [10]. Internet-delivered cognitive behavioral therapy (iCBT) entails modules or lessons delivering cognitive behavioral therapy (CBT) concepts via the internet [11]. iCBT for depression offers solutions to such challenges as poor treatment availability, fidelity and affordability, and constraints of location and time [10,12]. In randomized controlled trials (RCTs), iCBT for depression demonstrates efficacy equal to that of traditional face-to-face CBT but only if accompanied by therapist support [11,13,14], wherein the program involves therapist guidance, typically via email [15]. Fully automated iCBT treatments, supported by only computerized reminders, tend to be less efficacious than the ones with therapist guidance [16,17]**.**

High variation in attrition rates indicates that in some iCBT studies, a considerable number of patients fail to complete the treatment even in therapist-supported programs [18]. In a review comprising 19 studies on iCBT with therapist support for various mental disorders, the dropout (defined as termination at any point between registering for treatment and completing follow-up questionnaires) ranged from 2% to 83%, with a weighted average of 31% [19]. Two reviews of RCTs for therapist-supported iCBTs have shown a dropout rate of approximately 20% [10,20] and one meta-analysis on iCBT for depression showed a mean dropout rate of 28% [21]. However, in a large routine clinical practice study of a primary care population, 82% of participants failed to finish the program [22].

A meta-analysis of iCBT studies [19] identified several potential predictors of attrition, including age, gender, socioeconomic and relationship status, duration and target of the psychological problem, comorbidity, client-related psychological variables, treatment credibility, computer experience, motivation, the type of support medium, and quality and duration of clinician contact. Research evidence on these variables is, so far, limited. Adding fees, choice, reminders, and clinician support seems to strengthen adherence [23]. The essential elements of the contact are, however, debatable. Eysenbach [18] hypothesized that dropout is more likely the more “virtual” the support is. Synchronous telephone contact may be less virtual than is support in writing. Thus far, to our knowledge, no meta-analysis has yet directly investigated the impact of the contact medium used in iCBT for depression [24].

Telephone calls offering technical assistance facilitated self-monitoring during a web-based intervention versus only automated assistance [25]. Automated but personalized telephone prompts were useful for maintaining adherence in a study on weight control intervention [26]. However, in iCBT for depression, weekly telephone calls by a lay telephone counsellor did not confer additional advantage for callers to a helpline service [27]. The researchers proposed that more benefits might be achieved with support provided by a clinician. Nevertheless, therapist-delivered telephone and email support in iCBT for depression did not yield improvement in outcomes or in dropout rates [24]. Neither did technician-delivered telephone support demonstrate advantages over a clinician-moderated online discussion forum in iCBT for social phobia [28], although in the same program, telephone calls facilitated symptomatic improvement and adherence versus automatic reminders only [29]. In an iCBT aftercare program for bulimia nervosa, telephone prompts by a research assistant improved adherence [30], as did telephone support by a technician in iCBT for depression in primary care compared with text-based support [31]. Furthermore, in a trial of patients with treatment-resistant obsessive-compulsive disorder (OCD), scheduled telephone support (STS) reduced dropout rate compared with optional telephone support [32].

Even though STS has improved adherence, it consumes more resources than iCBT as usual and the target group therefore needs careful defining. Several potential factors predict dropout [19], but to the best of our knowledge, no clear-cut criteria for identification of increased dropout risk exist, and no trials have directly studied the effects of extra support in iCBT for depression to aid patients at such risk.

This RCT investigated whether STS added to iCBT as usual enhanced adherence to and effectiveness of iCBT for depression in a sample of patients at high risk for dropping out, (ie, those demonstrating delayed start of iCBT).

We expected that the add-on STS would enhance adherence to treatment, reduce symptoms of depression, and shorten the duration of treatment in patients at risk for dropping out.

Methods

HUS-iCBT as Usual

The department of psychiatry of Helsinki University Hospital (HUS) has developed and is providing nationwide a range of original Finnish-language iCBT programs (further referred to as HUS-iCBTs) for common psychiatric disorders to which all physicians licensed in Finland can refer patients. To receive HUS-iCBT for depression, patients must be diagnosed with a depressive episode (code F32-F33 in International Classification of Diseases 10) and be aged ≥18 years. Exclusion criteria are current alcohol misuse as judged by the referring physician; known diagnosis of schizophrenia or other psychotic disorder, bipolar disorder, serious personality disorder, or neurological or neuropsychiatric disorder that adversely affects the patient’s cognitive performance; or demonstrated, reported, or observed suicidal intentions. However, each referring physician holds primary judicial responsibility for overall treatment and, prior

XSL•FO
RenderX

to referral, verifies the diagnosis and checks all of these criteria. No pretreatment interview by the therapy provider is thus necessary.

The HUS-iCBT for depression consists of 7 consecutive modules and is 109 pages in total, including texts, videos, illustrations, and assignments. Contents include information on depression and CBT, goal definition, behavioral activation, cognitive restructuring, advice on a balanced life, relapse prevention, and homework. Patients use a secure online identity system to enter the program. They are required to report possible suicidal thoughts. The time schedule of the treatment is flexible, although patients are prompted to progress at a pace of 1 module per week. Active engagement is required, since progress to the next module is prevented if any of the previous module assignments is not completed.

The program sends email prompts for new messages and login reminders. The same internet therapist follows each patient throughout the treatment, and they communicate asynchronously via text-based message board within the therapy program. The therapist sends messages in the beginning, at midtreatment, at the sixth module, and at the end point; patients receive encouragement to write to the therapist any time with their questions or concerns. The therapist comments on the completed tasks, offering praise and support for the patient. The patients receive automatic messages recapitulating the contents of the modules, which are distinguished from the messages sent by therapists. In addition, the patients receive an email prompt if no login occurs for 2 weeks and when they receive a new message. Automatic prompts notify therapists of new messages or in case of any sign of suicidality. When expected progress fails, the therapist tries to contact the patient by a message within 2 to 4 weeks. If the patient still does not appear in the program, the therapist tries to reach the patient by telephone (therapist-initiated telephone calls are not used for other purposes). If the patient remains unreachable by telephone, a letter on paper is sent to uncover the reason for no show.

Study Patients and Design

Large Simple Observational Study

All patients referred to the HUS-iCBT for depression are invited to participate in an ongoing nationwide, low-threshold, flexible–time schedule, observational, routine clinical care study on the effectiveness of HUS-iCBT in depressive disorders (original report under preparation). The only additional criterion for that study is a signed informed consent. All consenting patients (currently 79% of those accepted for the HUS-iCBT) are eligible to participate and are enrolled into that study.

The STS Study

The current study was a subsidiary RCT branch of the observational study described above. Patients participating in the observational study received invitations to participate in the STS study if they had not proceeded in the HUS-iCBT for ≥14 days after their first entry, a delay interpreted as increased risk for dropping out. The patients were enrolled subsequently from September 2015 to October 2016 until the number of participants reached 100. These patients were randomly allocated to either HUS-iCBT as usual plus add-on STS (add-on STS group, n=50) or to HUS-iCBT as usual (control group, n=50) when a 14-day delay was detected, either during or after completing the first module of the treatment. The maximum time span allowed for this study participation since initiation was 6 months.

Add-on STS Intervention

In addition to the standard HUS-iCBT, the add-on STS intervention group received 8 weekly 15-minute telephone calls, the first at the beginning of iCBT and the subsequent calls during each of the 7 modules, without any further calls regardless of patient progress. If necessary (eg, if a patient was reached in a bad moment or asked the therapist to call again), new calls were allowed, amounting to more than the intended 8 calls. During the first call, selection of individual goals took place. The tasks and themes of each module were discussed in the calls that followed each module. Support was individually tailored and followed the principles of the model of supportive accountability [33], which combines elements of motivational theory, organizational psychology, and computer-mediated communication to create a framework for supported computerized treatment. According to this model, combining elements of support and accountability with a legitimized and trustworthy relationship increases adherence. If delays occurred, the prescheduled calls were generally not replaced, although therapists had final judgement over this.

Therapists

All 5 therapists involved in this trial were clinical psychologists employed by HUS and had at least 2 years of work experience with depressed patients. They attended a 1-day training session in HUS-iCBT on the study protocol of the above-mentioned observational study and of the current STS study, the STS methodology, and the International Council for Harmonisation Good Clinical Practice (ICH-GCP) guidelines. Their role included providing feedback on assignments and support with any patient issues. Every 2 months, they participated in a group supervision session, discussing any concerns or issues arising. At any time, the therapists could consult a clinical psychologist with advanced iCBT experience.

All 5 STS therapists were at the same time also providing regular HUS-iCBT for other patients, and in this study, each treated both add-on STS intervention and control groups. Study patients were assigned to each therapist randomly, depending on that therapist's current workload, resulting in the 5 therapists having 35, 30, 18, 14, and 3 patients.

Outcome Measures

Primary outcome measures were proportion of patients reaching midtreatment (third module) and end point (seventh module). Secondary outcome measures were change (from start to completion or to dropout, with last observation carried forward) in Beck Depression Inventory (BDI) [34] scores and treatment time (in days) from start to end point (or premature discontinuation). The internet-administered BDI has good psychometric properties [35]. The patients completed it online at the beginning, midtreatment, and end point.

At enrollment, all patients completed 6 demographic questions. The rest of the demographic data were collected from the

XSL•FO
RenderX

referrals. The therapists recorded the number and duration of telephone calls. The experienced usefulness was recorded at the end of each module by visual analog scale, where "not at all useful" equaled 0 and "very useful" equaled 10.

Ethics and Legislation

The study followed the ICH-GCP and Finnish national regulations. The study protocol was approved by the Ethics Committee of HUS and by pertinent institutional authorities. After reading a complete description of the study, the patients provided informed consent electronically. The trial was registered at the International Standard Randomised Controlled Trial Number (ISRCTN) registry (ISRCTN55123131).

Statistical Analyses

Primary analyses, based on the intention-to-treat principle, included all patients. Analysis of variance and chi-square tests served to reveal group differences in baseline characteristics and in dropout rates.

Group comparisons of patients reaching midtreatment and end point employed chi-square tests. Cox regression survival analysis served to determine group differences in survival probabilities (ie, time spent in treatment before dropping out during the 6-month time period). This analysis was adjusted for experienced session usefulness at the end of the first session, measured by visual analog scale. The Mann-Whitney test served for group comparisons of change in depression (BDI scores) from baseline to last observation carried forward and treatment length (in days).

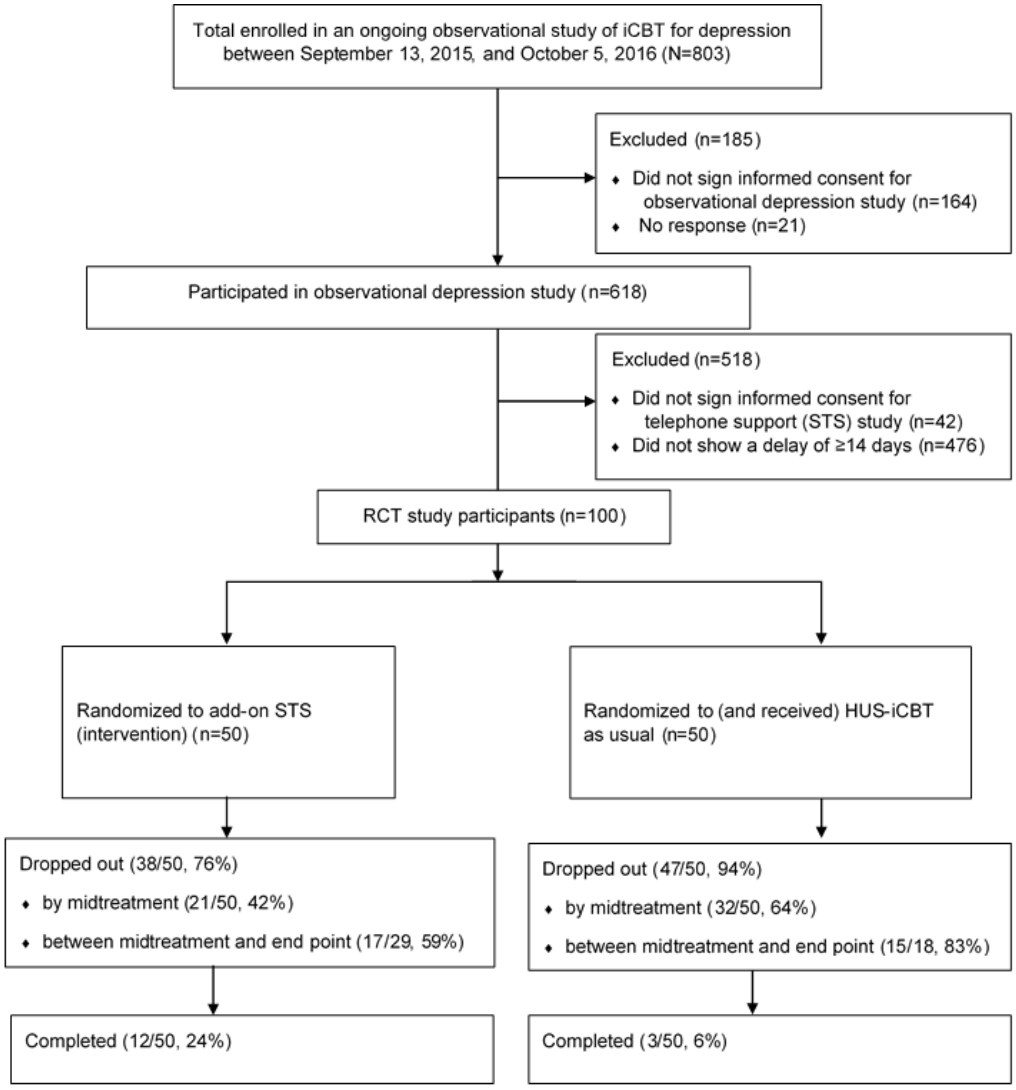
Statistical analyses employed SPSS Statistics (version 25; IBM Corp) for Windows (Microsoft Corp).

Results

Baseline Characteristics

Analyses comprised 50 patients in the add-on STS group and 50 patients in the control group (see Figure 1).

Figure 1. Scheduled telephone support in iCBT for depression: patient flow diagram. HUS-iCBT: Helsinki University Hospital internet-based cognitive behavioral therapy programs; iCBT: internet-delivered cognitive behavioral therapy; RCT: randomized controlled trial; STS: scheduled telephone support.



Of 618 patients of the observational depression study, all 618 (100%) were invited to participate in the STS study, which amounts to 77.0% of the 803 patients enrolled during the time period between September 13, 2015, and October 5, 2016. Of

XSL•FO
RenderX

the 618 invited, 576 patients (93.2%) agreed to participate. Of these 576 patients, 476 (82.6%) did not show a 14-day delay in beginning the treatment and were excluded on that criterion. A total of 6 patients of the add-on STS group never received any telephone calls (ie, did not receive the STS intervention).

No significant differences between groups emerged at baseline (see Table 1). Some patients did not provide several baseline demographic data (marital, educational, and employment statuses and use of medication or sick leave within last 6 months). However, there was no association between missing data and group membership. Interpretations of results concerning primary and secondary outcomes remained the same after controlling for an indicator of nonresponse to these demographic questions.

Table 1. Baseline group characteristics.

	Add-on STS[a] (n=50)	Control (n=50)	Total (n=100)	Chi-square (*df*)	*t* test[b] (*df*)	*P* value
Female, n (%)	34 (68.0)	32 (64.0)	66 (66.0)	0.18 (1)	N/A[c]	.67
Referred from, n (%)				6.50 (5)	N/A	.26
Primary health care	28 (56.0)	23 (46.0)	51 (51.0)			
Private health care	11 (22.0)	15 (30.0)	26 (26.0)			
Occupational health care	7 (14.0)	2 (4.0)	9 (9.0)			
Student health care	2 (4.0)	5 (10.0)	7 (7.0)			
Specialized psychiatry	1 (2.0)	3 (6.0)	4 (4.0)			
Unspecified	1 (2.0)	2 (4.0)	3 (3.0)			
Marital status, n (%)[d]				0.80 (3)	N/A	.85
Married	6 (42.9)	4 (28.6)	10 (35.7)			
Living together	2 (14.3)	2 (14.3)	4 (14.3)			
Not married	4 (28.6)	6 (42.9)	10 (35.7)			
Divorced	2 (14.3)	2 (14.3)	4 (14.3)			
Educational level, n (%)[d]				0.80 (3)	N/A	.85
Elementary school	1 (7.1)	2 (14.3)	3 (10.7)			
Secondary/vocational	8 (57.1)	7 (50.0)	15 (53.6)			
College/university Bachelor	3 (21.4)	2 (14.3)	5 (17.9)			
College/university Master	2 (14.3)	3 (21.4)	5 (17.9)			
Employment status, n (%)[d]				4.88 (3)	N/A	.18
Full time	11 (78.6)	7 (50.0)	18 (64.3)			
Part time	1 (7.1)	0 (0)	1 (3.6)			
Unemployed	2 (14.3)	6 (42.9)	8 (28.6)			
Retired	0 (0)	1 (7.1)	1 (3.6)			
Medication, n (%)[d]				0.16 (1)	N/A	.69
None	5 (35.7)	4 (28.6)	9 (32.1)			
Present[e]	9 (64.3)	10 (71.4)	19 (67.9)			
Sick leave within 6 months, n (%)[d]	9 (64.3)	6 (42.9)	15 (53.6)	129 (1)	N/A	.26
Age, mean (SD)	37.40 (12.16)	34.82 (10.99)	36.11 (11.10)	N/A	1.11 (98)	.27
BDI[f] at baseline, mean (SD)	21.49 (7.15)	23.33 (10.04)	22.41 (8.75)	N/A	1.04 (96)	.30

[a]STS: scheduled telephone support.

[b]2-tailed *t* test.

[c]N/A: not applicable.

[d]Information available for 14 patients in add-on STS intervention group and for 14 control group patients.

[e]Anxiolytic or antidepressant.

[f]BDI: Beck Depression Inventory.

XSL•FO
RenderX

Since there was a considerable amount of missing data in some of the baseline characteristics, we performed comparisons (using a 2-tailed *t* test) of depression at baseline (BDI score) between patients that filled in (n=28) and did not fill in (n=72) a separate questionnaire of demographics. There was no difference in depression at baseline between the groups (t_{96}=0.445; *P*=.66).

No adverse events, such as hospitalization or serious illness, occurred during the study. Altogether, 44 of the 50 (88%) patients in the add-on STS intervention group received the previously scheduled telephone calls (versus a total of 5 patients received optional calls in the control group), as seen in Table 2.

Patients in the add-on STS group reached an average of 3.54 modules, while patients in the control group reached an average of 2.46 modules (see Table 3).

Table 2. Mean number and duration of telephone calls received.

Telephone call characteristics	Add-on STS[a] (n=50)	Control (n=50)
Proportion of patients who received calls, n (%)[b]	44 (88)	5 (10)
Number of calls per patient, mean (SD), range[b]	4.63 (3.58), 0-11	0.32 (0.99), 0-5
Average duration of calls (minutes), mean (SD)[b]	13.09 (5.25)	6.20 (2.17)
Cumulative duration of calls (minutes), mean (SD), range[b]	73.37 (48.95), 5-165	1.95 (6.11), 0-30
Assessed average therapist time (minutes), n[c]	132	55

[a]STS: scheduled telephone support.

[b]Missing values for 9 patients in each group.

[c]In addition to contact time itself, STS takes approximately 4 more minutes for preparation of the call and additional documentation, resulting in a total of 77 additional minutes for STS intervention vs HUS-iCBT as usual.

Table 3. Module reached by patients in each group during 6 months.

	Module reached, n (%)								Last module reached, mean (SD)
	0	1	2	3	4	5	6	7	
Add-on STS[a]	50 (100)	45 (90)	33 (66)	28 (56)	26 (52)	19 (38)	13 (26)	13 (26)	3.54 (2.57)
Control	50 (100)	48 (96)	27 (54)	19 (38)	15 (30)	7 (14)	4 (8)	3 (6)	2.46 (1.88)

[a]STS: scheduled telephone support.

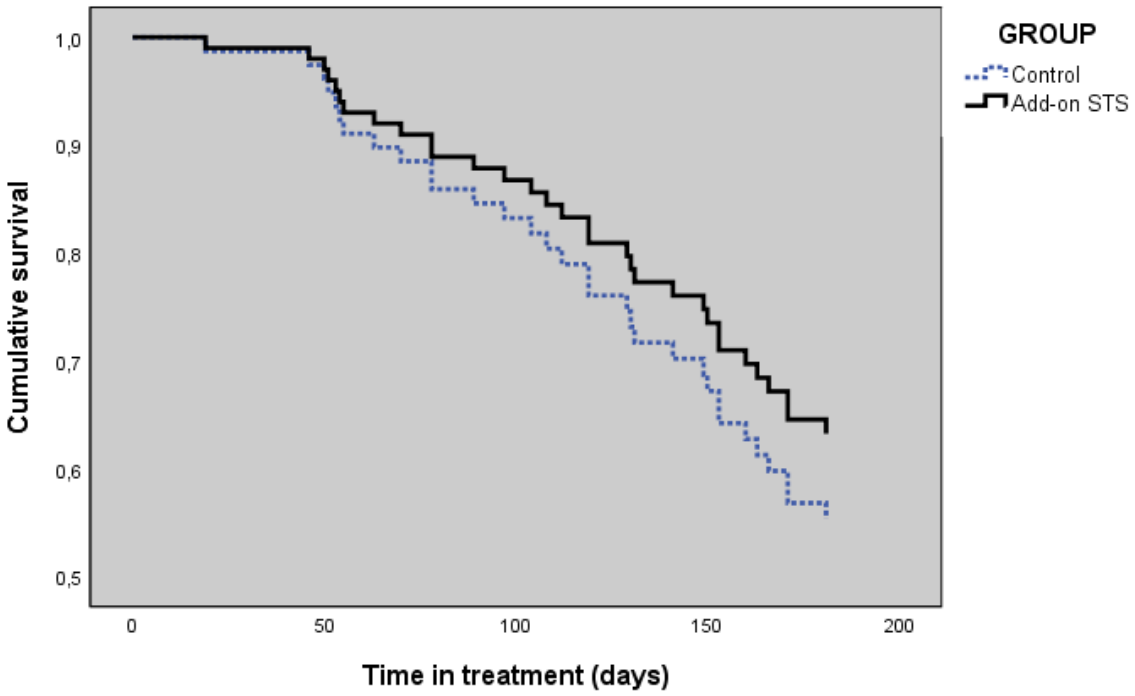
Outcomes

Adherence (Primary Outcome)

Of the 100 randomized patients, 21 of the 50 patients in the add-on STS group (42%) and 32 of the 50 patients in the control group (64%) dropped out by midtreatment. Corresponding figures for the end point were 38 of 50 patients (76%) and 47 of 50 patients (94%). The proportion who reached midtreatment and end point favored the add-on STS group significantly (Table 4). Cox regression survival analysis revealed no difference in the timing of dropout between groups (survival curves χ^2_2=2.5, *P*=.48; hazard ratio=1.29, 95% CI 0.64-2.61, *P*=0.48), as shown in Figure 2.

XSL•FO
RenderX

Figure 2. Cox regression survival curves for staying in treatment for the add-on STS and control groups at 6 months. STS: scheduled telephone support.



Secondary Outcomes

The BDI-measured change from baseline showed a statistically significant difference also favoring the add-on STS group. This result also remained when analyses were carried out using linear mixed models, which handles missing values more efficiently and without imputing the missing values ($F_{2,30}$=3.77; P=.04). No statistically significant differences emerged for days spent on therapy (P=.67).

Table 4. Effects of add-on STS on adherence, depression, and time in treatment.

Outcome	Add-on STS[a] (n=50)	Control (n=50)	Chi-square (*df*)	Mann-Whitney test	*P* value
Reached midtreatment, n (%)[b]	29 (58)	18 (36)	4.9 (1)	N/A	.045[c]
Completed the program, n (%)[b]	12 (24)	3 (6)	6.4 (1)	N/A	.02[c]
Change in BDI[d] from baseline[c,e]			N/A[f]	1455.5	.049[c]
Change in BDI from baseline, mean (SD)	3.63 (5.94)	1.06 (4.82)			
Change in BDI from baseline, median (P25;P75)[g]	0 (0.0;8.5)	0 (0.0;1.0)			
Treatment time in days[e]			N/A	1190.5	.67
Treatment time (days), mean (SD)	136.61 (52.18)	141.36 (48.08)			
Treatment time (days), median (P25;P75)	166.95 (96.27;183.00)	161.5 (100.25;183.00)			

[a]STS: scheduled telephone support.

[b]Primary outcome.

[c]Last observation carried forward.

[d]BDI: Beck Depression Inventory.

[e]Secondary outcome.

[f]N/A: not applicable.

[g]25th and 75th percentiles.

XSL•FO
RenderX

Discussion

Principal Results

In patients at risk of dropout during iCBT for depression, add-on STS yielded an increased proportion of patients reaching midtreatment (29/50, 58% vs 18/50, 36%; *P*=.045) and proportion of patients who completed treatment (12/50, 24% vs 3/50, 6%; *P*=.02) compared with the control group. Moreover, depressive symptoms decreased more for those with add-on STS than with HUS-iCBT alone (mean change 3.63 vs 1.06; *P*=.049), but the survival analysis did not reveal any statistically significant differences in survival probabilities, that is, time spent in treatment before dropping out (*P*=.67).

Comparison With Prior Work

Two other RCTs on STS have reported positive results. Notably, both studies used iCBT with no text-based support, meaning that comparison of our results with these previous reports should be undertaken with caution.

Gilbody and coauthors [31] compared STS to iCBT with no STS in primary care patients with depression, reporting improvement in symptom change and a decrease in dropout rate from 71% to 54% at midtreatment and from 90% to 81% at end point in their STS group patients compared with their control group patients with access only to a support helpline. Despite the at-risk population in the current study, our figures seem comparable (drop from 32/50, 64% to 21/50, 42% at midtreatment and from 48/50, 96% to 38/50, 76% at end point). Kenwright and colleagues [32], in their iCBT trial comparing STS to patient-initiated calls, reported an STS-induced change from 59% to 14% for early dropout rate and from 64% to 50% for late dropout rate. Their patients, however, suffered not from depression but from OCD. OCD is known to often be difficult to treat and, as a group, patients in that study could thus also be counted as being at risk of dropping out. Nevertheless, due to the differences in population and methodology, the results cannot be directly compared with ours.

Of note, completion of the program does not always indicate treatment success in iCBTs. For instance, in iCBT for depression, there seems to exist a subgroup of patients who benefit rapidly and therefore prematurely discontinue the treatment because it is subjectively perceived as unneeded [36]. In our study, in-depth analysis of reasons and exact time point of dropout could not be performed due to having only 3 measurement points of depression.

Not all RCTs on STS showed the desired results, however. In one depression study [24], patients received either email or STS support added to iCBT, but with no differences in dropout rates. Similarly, adherence did not improve with STS added to email in a headache study by Andersson et al [37], nor did STS demonstrate advantages over a clinician-moderated online discussion forum in terms of adherence in an RCT on social phobia from Titov and colleagues [38]. Interestingly enough, all these negative studies included only unselected, self-referred participants. In contrast, we focused only on physician-referred patients (51/100, 51% of them from primary care) whose initiation of iCBT was delayed by ≥2 weeks, which was considered a risk of dropping out. These patients may have experienced a special challenge in terms of adherence, since unlike their counterparts in the majority of previous RCTs, they may not always have originally been motivated to use iCBT; they had no contact with a therapist or other research personnel at screening (a contact that might have improved adherence), and they showed a possible lack of engagement at the beginning of iCBT.

The education level of supporting personnel may also matter, but data on this issue are conflicting. In the negative-result depression study by Lindner and colleagues [24], support providers were psychology students, whereas our positive results were achieved with support personnel being certified, experienced clinical psychologists. Nevertheless, positive results also emerged in the depression study of Gilbody and colleagues [31], in which contact personnel comprised telephone support workers. Moreover, STS personnel involved in other (nondepression) conditions differed. Therapists were the support providers in one positive-result OCD trial [32] and in one negative headache trial [37], whereas technicians took this role in the negative-result social phobia trial of Titov and colleagues [38].

In psychotherapy research, researcher allegiance, a tendency to favor certain preferred treatments, is widely discussed [39]. Treatment developers' trials tend to yield better outcomes than the later research performed by others, probably due to researcher allegiance. One alternative explanation is that developers themselves gain an advantage through their intense clinical and research involvement with the treatment; they become superior in treatment delivery [40]. Gilbody and colleagues [31] attributed their poor outcomes and adherence—results inferior to those demonstrated by treatment developers—to differing means of service provision. Gilbody and his group performed their trial within the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy trial, a large National Health Service–funded program in which treatment provision was decentralized. In contrast, in our setting, the developer of the HUS-iCBT program, which was the Department of Psychiatry of Helsinki University Hospital, was the only nationwide centralized provider of both iCBT and the add-on STS. This may explain our somewhat better-than-expected results.

The mechanism of the desirable effects of STS is uncertain. These effects may result from increased therapist contact time, from the telephone as the medium versus written contact only, or from the proactivity of therapist-initiated scheduled contact.

What has been maintained is that increased therapist contact time in iCBT for mood and anxiety disorders correlates strongly with treatment outcome [41]. This association has not always been detectable in reviews, however [24], and has been considered true only up to a certain threshold [42]. Gilbody and colleagues [31] did not describe their average STS time. The STS time in the iCBT study on OCD [32] was 232 minutes versus 178 minutes for the optional contact, which was markedly more than the typical 90 to 150 minutes per patient in iCBT trials [43]. In our study, average added telephone contact time for the add-on STS intervention group was 73 minutes per

XSL•FO
RenderX

patient (versus 2 minutes for the control group). Given the average 55-minute therapist time per patient for HUS-iCBT, SPS means a 2.4-fold increase in contact time. Neither the study by Kenwright and colleagues [32] or by Gilbody and colleagues [31] nor the present one has been a dose-finding trial. We were unable to locate any dose-finding trials on the effects of contact time in added STS in any iCBT program. It therefore remains unclear whether those 54 additional minutes of STS time for OCD or our 73 additional minutes for depression are optimal in terms of adherence or symptomatic improvement in iCBT. Nevertheless, the 2.4-fold increase in resource allocation when weighted against health benefits achieved may be of interest to administrative decision makers.

The telephone as a medium is one of the synchronous contact modes, in contrast to asynchronous contact via written message (usually referred to as minimal contact) in the majority of iCBT studies. Furthermore, unlike text message, a telephone conversation conveys nonverbal voice signals (such as intonation and pausation). Since in these 2 studies [31,32] the control groups received no text-based support, no conclusions can be drawn on the specific role of the telephone as a medium. Moreover, other RCTs comparing telephone support with email support found no differences in treatment outcomes or dropout rates [24,37,38]. The role of the voice contact in iCBTs and the specific mechanisms of action therein remain poorly understood.

Limitations and Strengths

Our sample size was relatively small. Although between-group differences were statistically significant, statistical power was insufficient for subgroup analyses and hence for identification of a subpopulation for optimal therapist resource allocation.

Due to selection of population (patients at risk for dropout in everyday clinical practice), the dropout rate in our study was expectedly high, with 76% (38/50) in the STS group and 94% (47/50) in the control group dropping out, when dropout is defined as those who did not finish the treatment in 6 months. In addition, the patients did not have a strict deadline of 6 months, even though the measurements were conducted at that point. Studies comparing iCBT programs with unselected populations reported a 62% dropout rate in everyday clinical practice and 15% to 25% in RCTs [29]. Our findings underline the difficulties of maintaining adherence in routine clinical care for digital health products that demonstrate impressive results in clinical trials [44]. Nevertheless, the STS in this study showed results comparable to an 82% dropout rate in earlier observational studies on iCBT for depression with unselected population [22].

Therapist effects in our study could not be ruled out, since 2 of altogether 5 therapists treated 63 of the 100 (63%) patients. This small number of therapists is, however, too small for reliable assessment of possible therapist effects [45].

The effect of support may depend on the patient group [46]. We defined the ≥14-day delay prior to the beginning of iCBT treatment as a criterion of dropout risk based on clinical experience, since, to the best of our knowledge, no clear-cut criteria for such risk have yet been defined. This criterion might fit only a certain subgroup of our patients, meaning that the add-on STS intervention was applied for no reason also to a not-at-risk subgroup. Understanding the reasons for patient delay could help to sharpen this criterion, but in our study, data as to the reasons were unavailable.

The HUS-iCBT with add-on STS required 77 minutes more therapist time than did the HUS-iCBT as usual. Cost-benefit analysis of STS was not a subject of this study, so our results are insufficient to inform decision makers on whether to use STS for patients at risk for dropout in routine clinical practice, not to mention for patients at no such risk.

Future Research

Future research should explore in depth possible criteria for those at risk for dropping out to discover optimal contact time with larger populations. Add-on STS in other populations, such as depressed patients who demonstrate no delay in initiation of iCBT or patients with disorders other than depression, is worth studying. Reasons for the delayed initiation of iCBT as well as other predictors of poor adherence call for exploration. Cost-benefit analysis could optimize the use of add-on STS. Add-on STS provided by professional groups of a lower educational level than that of our clinical psychologists also demands exploration.

Our study employed STS added to text-based support, but STS alone should also be a focus of research in populations at risk of dropout to reveal the unconfounded effects of STS as such. Comparison of STS with other support media, such as voice messages (asynchronous contact mode but enriched with nonverbal voice modulations) or chat (synchronous contact but devoid of nonverbal voice modulations), can yield new insights into both the mechanism of action of STS and the practical means to optimize cost benefits of support.

Possible therapist effects should be explored in further iCBT studies with a greater number of therapists with equal caseloads providing STS.

Conclusion

STS added to the usual iCBT appears to improve both adherence and clinical symptomatology in patients with depression at risk for dropout from iCBT, but more research is required to optimize its use.

Acknowledgments

We would like to thank Dr Carolyn Brimley Norris for the English language edit of this paper.

This work was supported by the Finnish Cultural Foundation, the Government of Finland, and the Hospital Region of Helsinki and Uusimaa.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V. 1. 6. 1).
[PDF File (Adobe PDF File), 2288 KB - jmir_v22i7e15732_app1.pdf]

References

1. Emmelkamp PMG, David D, Beckers T, Muris P, Cuijpers P, Lutz W, et al. Advancing psychotherapy and evidence-based psychological interventions. Int J Methods Psychiatr Res 2014 Jan;23 Suppl 1:58-91 [FREE Full text] [doi: 10.1002/mpr.1411] [Medline: 24375536]
2. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS One 2015;10(2):e0116820 [FREE Full text] [doi: 10.1371/journal.pone.0116820] [Medline: 25658103]
3. Depression. World Health Organization. 2020 Jan. URL: http://www.who.int/mediacentre/factsheets/fs369/en/ [accessed 2020-06-26]
4. Eaton WW, Martins SS, Nestadt G, Bienvenu OJ, Clarke D, Alexandre P. The burden of mental disorders. Epidemiol Rev 2008;30:1-14 [FREE Full text] [doi: 10.1093/epirev/mxn011] [Medline: 18806255]
5. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602 [FREE Full text] [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
6. van Schaik DJF, Klijn AFJ, van Hout HPJ, van Marwijk HWJ, Beekman ATF, de Haan M, et al. Patients' preferences in the treatment of depressive disorder in primary care. Gen Hosp Psychiatry 2004;26(3):184-189. [doi: 10.1016/j.genhosppsych.2003.12.001] [Medline: 15121346]
7. Leykin Y, Derubeis RJ, Gallop R, Amsterdam JD, Shelton RC, Hollon SD. The relation of patients' treatment preferences to outcome in a randomized clinical trial. Behav Ther 2007 Sep;38(3):209-217. [doi: 10.1016/j.beth.2006.08.002] [Medline: 17697846]
8. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [doi: 10.1590/S0042-96862004001100011] [Medline: 15640922]
9. Mechanic D. Barriers to help-seeking, detection, and adequate treatment for anxiety and mood disorders: implications for health care policy. J Clin Psychiatry 2007;68 Suppl 2:20-26 [FREE Full text] [Medline: 17288503]
10. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
11. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
12. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided Internet-Based CBT for Common Mental Disorders. J Contemp Psychother 2013 May 21;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
13. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
14. Andersson G, Wagner B, Cuijpers P. ICBT for depression. In: Lindefors N, Andersson G, editors. Guided Internet-Based Treatments in Psychiatry. London, England: Springer; 2016:17-32.
15. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
16. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
17. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes Internet therapy work? Cogn Behav Ther 2009;38 Suppl 1:55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
20. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and

XSL•FO
RenderX

methodological validity. BMC Psychiatry 2013 Apr 15;13:113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
21. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
22. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
23. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
24. Lindner P, Olsson E, Johnsson A, Dahlin M, Andersson G, Carlbring P. The impact of telephone versus e-mail therapist guidance on treatment outcomes, therapeutic alliance and treatment engagement in Internet-delivered CBT for depression: A randomised pilot trial. Internet Interv;1(4) 2014 Oct:A [FREE Full text] [doi: 10.1016/j.invent.2014.09.001]
25. Greaney ML, Sprunck-Harrild K, Bennett GG, Puleo E, Haines J, Viswanath KV, et al. Use of email and telephone prompts to increase self-monitoring in a Web-based intervention: randomized controlled trial. J Med Internet Res 2012 Jul 27;14(4):e96 [FREE Full text] [doi: 10.2196/jmir.1981] [Medline: 22842775]
26. Stevens VJ, Funk KL, Brantley PJ, Erlinger TP, Myers VH, Champagne CM, et al. Design and implementation of an interactive website to support long-term maintenance of weight loss. J Med Internet Res 2008 Jan 25;10(1):e1 [FREE Full text] [doi: 10.2196/jmir.931] [Medline: 18244892]
27. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
28. Titov N, Andrews G, Schwencke G. Shyness 2: treating social phobia online: replication and extension. Aust N Z J Psychiatry 2008 Jul;42(7):595-605. [doi: 10.1080/00048670802119820] [Medline: 18612863]
29. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
30. Beintner I, Jacobi C. Impact of telephone prompts on the adherence to an Internet-based aftercare program for women with bulimia nervosa: A secondary analysis of data from a randomized controlled trial. Internet Interv 2019 Mar;15:100-104 [FREE Full text] [doi: 10.1016/j.invent.2017.11.001] [Medline: 30792960]
31. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, REEACT collaborative. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. Br J Psychiatry 2017 May;210(5):362-367. [doi: 10.1192/bjp.bp.116.192435] [Medline: 28254959]
32. Kenwright M, Marks I, Graham C, Franses A, Mataix-Cols D. Brief scheduled phone support from a clinician to enhance computer-aided self-help for obsessive-compulsive disorder: randomized controlled trial. J Clin Psychol 2005 Dec;61(12):1499-1508. [doi: 10.1002/jclp.20204] [Medline: 16173084]
33. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
34. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
35. Carlbring P, Brunt S, Bohman S, Austin D, Richards J, Öst L, et al. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Computers in Human Behavior 2007 May;23(3):1421-1434. [doi: 10.1016/j.chb.2005.05.002]
36. Lutz W, Arndt A, Rubel J, Berger T, Schröder J, Späth C, et al. Defining and Predicting Patterns of Early Response in a Web-Based Intervention for Depression. J Med Internet Res 2017 Jun 09;19(6):e206 [FREE Full text] [doi: 10.2196/jmir.7367] [Medline: 28600278]
37. Andersson G, Lundström P, Ström L. Internet-based treatment of headache: does telephone contact add anything? Headache 2003 Apr;43(4):353-361. [doi: 10.1046/j.1526-4610.2003.03070.x] [Medline: 12656706]
38. Titov N, Andrews G, Schwencke G, Solley K, Johnston L, Robinson E. An RCT Comparing Effect of Two Types of Support on Severity of Symptoms for People Completing Internet-Based Cognitive Behaviour Therapy for Social Phobia. Aust N Z J Psychiatry 2009 Jan;43(10):920-926. [doi: 10.1080/00048670903179228]
39. Dragioti E, Dimoliatis I, Evangelou E. Disclosure of researcher allegiance in meta-analyses and randomised controlled trials of psychotherapy: a systematic appraisal. BMJ Open 2015 Jun 01;5(6):e007206 [FREE Full text] [doi: 10.1136/bmjopen-2014-007206] [Medline: 26033943]
40. Munder T, Brütsch O, Leonhart R, Gerger H, Barth J. Researcher allegiance in psychotherapy outcome research: an overview of reviews. Clin Psychol Rev 2013 Jun;33(4):501-511. [doi: 10.1016/j.cpr.2013.02.002] [Medline: 23500154]

XSL•FO
RenderX

41. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
42. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
43. Cavanagh K, Millings A. (Inter)personal Computing: The Role of the Therapeutic Relationship in E-mental Health. J Contemp Psychother 2013 Jul 17;43(4):197-206. [doi: 10.1007/s10879-013-9242-z]
44. Kvedar JC, Fogel L. Why real-world results are so challenging for digital health. NEJM Catalyst. URL: https://catalyst.nejm.org/real-world-results-digital-health-products/ [accessed 2019-12-11]
45. Bryan ML, Jenkins SP. Multilevel Modelling of Country Effects: A Cautionary Tale. Eur Sociol Rev 2015 May 08;32(1):3-22. [doi: 10.1093/esr/jcv059]
46. Bergman Nordgren L, Carlbring P, Linna E, Andersson G. Role of the working alliance on treatment outcome in tailored internet-based cognitive behavioural therapy for anxiety disorders: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 18;2(1):e4 [FREE Full text] [doi: 10.2196/resprot.2292] [Medline: 23612437]

Abbreviations

BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
HUS: Helsinki University Hospital
HUS-iCBT: Helsinki University Hospital internet-based cognitive behavioral therapy programs
iCBT: internet-based cognitive behavioral therapy
ICH-GCP: International Council for Harmonisation Good Clinical Practice
ISRCTN: International Standard Randomised Controlled Trial Number
OCD: obsessive-compulsive disorder
RCT: randomized controlled trial
STS: scheduled telephone support

Edited by G Eysenbach; submitted 01.08.19; peer-reviewed by E Littlewood, I Beintner; comments to author 28.10.19; revised version received 20.12.19; accepted 21.04.20; published 23.07.20.

Please cite as:
Pihlaja S, Lahti J, Lipsanen JO, Ritola V, Gummerus EM, Stenberg JH, Joffe G
Scheduled Telephone Support for Internet Cognitive Behavioral Therapy for Depression in Patients at Risk for Dropout: Pragmatic Randomized Controlled Trial
J Med Internet Res 2020;22(7):e15732
URL: http://www.jmir.org/2020/7/e15732/
doi:10.2196/15732
PMID:32706658

©Satu Pihlaja, Jari Lahti, Jari Olavi Lipsanen, Ville Ritola, Eero-Matti Gummerus, Jan-Henry Stenberg, Grigori Joffe. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of a Blended Physical Activity Intervention for Older Adults: Mixed Methods Study

Sumit Mehra[1,2,3], MSc; Jantine van den Helder[4,5], MSc; Bart Visser[4], PhD; Raoul H H Engelbert[4,6], PhD; Peter J M Weijs[5,7,8], PhD; Ben J A Kröse[2,3], PhD

[1]Applied Psychology, Faculty of Applied Social Sciences and Law, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[2]CREATE-IT Applied Research, Faculty of Digital Media and Creative Industries, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[3]Informatics Institute, Faculty of Science, University of Amsterdam, Amsterdam, Netherlands
[4]Center of Expertise Urban Vitality, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[5]Center of Expertise Urban Vitality, Faculty of Sports and Nutrition, Amsterdam University of Applied Sciences, Amsterdam, Netherlands
[6]Department of Rehabilitation, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[7]Department of Nutrition and Dietetics, Amsterdam University Medical Centers, VU University, Amsterdam, Netherlands
[8]Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, VU University, Amsterdam, Netherlands

Corresponding Author:
Sumit Mehra, MSc
Applied Psychology
Faculty of Applied Social Sciences and Law
Amsterdam University of Applied Sciences
Wibautstraat 3b
Amsterdam, 1091 GH
Netherlands
Phone: 31 0621156956
Email: s.mehra@hva.nl

Abstract

Background: Physical activity can prolong the ability of older adults to live independently. Home-based exercises can help achieve the recommended physical activity levels. A blended intervention was developed to support older adults in performing home-based exercises. A tablet and a personal coach were provided to facilitate the self-regulation of exercise behavior.

Objective: In line with the Medical Research Council framework, this study aimed to carry out process evaluation of a blended intervention. The objectives were (1) to assess the long-term usability of the tablet adopted in the blended intervention and (2) to explore how the tablet, in conjunction with a personal coach, supported older adults in performing home-based exercises.

Methods: The process evaluation was conducted with a mixed-methods approach. At baseline, older adults participating in the blended intervention were asked to fill out a questionnaire about their general experience with information and communication technology (ICT) devices and rate their own skill level. After 6 months, participants filled out the Usefulness, Satisfaction, and Ease of use (USE) questionnaire to assess the usefulness, satisfaction, and ease of use of the tablet. With a random selection of participants, in-depth interviews were held to explore how the tablet and coach supported the self-regulation. The interviews were double coded and analyzed with the directed content analysis method.

Results: At baseline, 29% (65/224) of participants who started the intervention (mean age 72 years) filled out the ICT survey and 36% (37/103) of participants who used the tablet for 6 months (mean age 71 years) filled out the USE questionnaire. Furthermore, with 17% (18/103) of participants (mean age 73 years), follow-up interviews were held. The results of the baseline questionnaire showed that the large majority of participants already had experience with a tablet, used it regularly, and reported being skillful in operating ICT devices. After 6 months of use, the participants rated the usefulness, satisfaction, and ease of use of the tablet on average as 3.8, 4.2, and 4.1, respectively, on a 5-point scale. The analysis of the interviews showed that the participants felt that the tablet supported action planning, behavior execution, and self-monitoring. On the other hand, especially during the first few months, the personal coach added value during the goal setting, behavior execution, and evaluation phases of self-regulation.

Conclusions: The results of the process evaluation showed that older adults who participated in the study were positive about the blended intervention that was designed to support them in performing home-based exercises. Participants reported that the

XSL•FO
RenderX

tablet helped them to perform the exercises better, more frequently, and safely. It supported them in various phases of self-regulation. The availability of a personal coach was nevertheless crucial. To support physical activity in older adults, a blended approach is promising.

(J Med Internet Res 2020;22(7):e16380) doi:10.2196/16380

KEYWORDS

frail elderly; aged; activities of daily living; exercise; health behavior; telemedicine; mobile devices; tablet computers; usability testing; evaluation

Introduction

Background

As people age, they face a decline in daily functioning and mobility [1,2]. Physical activity can delay the onset and slow the decline associated with aging [3,4]. Older adults who exercise on a regular basis can prevent impairments and remain self-reliant for a longer period of time [5,6]. Accordingly, various community centers around the world offer senior citizens the opportunity to participate in group-based exercise classes under the guidance of an instructor [7-9]. For instance, in the Netherlands, over 400,000 older adults participate in the weekly activities of "More Exercise for Seniors" ("*Meer Bewegen voor Ouderen*," which is abbreviated as MBvO in Dutch). Despite the popularity of this program, its effects on physical health are limited. A previous study has shown that older adults who participate once a week in the exercise classes, do not achieve a higher health-related quality of life or an increased ability to perform daily tasks [10]. In order to capitalize on the health benefits of physical activity, the frequency, intensity, and duration of exercises have to be sufficient [9,11].

Older adults can increase the level of physical activity by doing exercises at home, either as an independent program or in conjunction with group-based classes [12-16]. The latter approach combines the motivational aspects of exercising along with peers with the flexibility of a home-based exercise program that is tailored to individual needs. However, in the absence of an instructor, older adults may have adherence and safety concerns about home-based exercises [17]. The use of mobile technology (mobile health [mHealth]) can help overcome these issues by providing detailed instructions, offering tailored programs, and tracking progress [18-23] for individuals, including older adults [24,25].

Development of a Blended Intervention

In order to enhance community-based exercise programs like MBvO, a blended intervention was developed as part of the MOTO-B (Motivating Technology for Older Adults' Behavior) and VITAMIN (VITal AMsterdam older adults IN the city) research projects. The aim of the intervention was to support older adults in performing home-based exercises. In line with the self-determination theory [26,27], the intervention was conceived to increase competence and stimulate the autonomy of older adults, but at the same time, to maintain relatedness with peers [17]. The intervention consisted of a home-based exercise program that was supported by a tablet and a personal coach, and could be followed alongside community-based exercise programs or other sport activities.

Objective

According to the UK Medical Research Council (MRC), complex interventions need to be evaluated systematically [28]. Three different types of evaluations can be distinguished as follows: (1) assessing the feasibility, (2) assessing the effectiveness, and (3) understanding the underlying change process.

First, prior to assessing effectiveness, feasibility should be investigated thoroughly. For the blended intervention described here, a previous usability study that was conducted in a laboratory showed that older adults (age ranging from 69 to 99 years) who used the app for the first time during a 45-minute session could operate it without any relevant problems [29], suggesting that the blended intervention is feasible. However, a more thorough evaluation is needed to account for the long-term use in a real-world setting. The usability of mHealth apps is often not tested sufficiently, thereby limiting their effectiveness [30,31].

Second, the effectiveness of a complex intervention can be assessed with randomized controlled trials (RCTs). To assess the effectiveness of the blended intervention in terms of health outcomes, a trial study is currently ongoing and will be reported in the future elsewhere [32].

Third, an explorative process evaluation can provide insights into the underlying change process. By exploring the mechanisms of action, a process evaluation is a valuable extension of effectiveness studies. The aim of this study was to conduct such an evaluation. The objectives were as follows: (1) to assess the long-term usability of the tablet in a real-world setting and (2) to explore how the tablet, in conjunction with a personal coach, supported older adults in performing home-based exercises.

Methods

Intervention

The intervention consisted of two components to support older adults in performing exercises at home. The first component was the tablet containing a custom-developed app that was designed to ensure behavior change by facilitating self-regulation. Self-regulation is the process of consciously guiding one's own behavior in order to achieve goals. In particular, behavior change techniques that support goal setting, action planning, behavior execution, self-monitoring, and evaluation appear to be important for the self-regulation of behavior [33-35]. Figure 1 presents a schematic representation. The app supported self-regulation by allowing older adults to

XSL•FO
RenderX

set goals, tailor a weekly schedule to their individual needs, and watch video instructions. It also tracked their progress and facilitated remote guidance by a personal coach. An elaborate description of the app and its theoretical underpinning have been presented previously [36].

The second component of the intervention was counselling. Each participant was appointed a personal coach. The coach paid house visits, helped the participants to get acquainted with the tablet, and counselled, either remotely or face-to-face, the participants in setting up and following the tailored exercise schedule. The coaches were third- and fourth-year physical therapy bachelor students. Prior to taking on their responsibilities, the coaches received a 2-week training on functional exercises, good clinical practice, and e-coaching by faculty staff members. Furthermore, during their 6-month internship, they received weekly supervision from faculty staff members. When the responsibilities of a coach ended at the end of a teaching semester, ongoing cases were transferred to a new coach. As a result, participants received in sequence counselling by two personal coaches during a 6-month period. The complete details of the intervention have been reported previously [32].

Figure 1. Behavior change through self-regulation.



Study Design and Participants

An RCT was conducted to assess the effectiveness of the blended intervention in terms of health outcomes. Older adults were recruited from the surroundings of Amsterdam, the Netherlands, through postal mailing and local community-based centers offering weekly exercise programs. Applicants were included in the trial if they met the following criteria: (1) age 55 years or older, (2) ability to understand the Dutch language, and (3) absence of specific cognitive or physical impairments. The protocol that describes the RCT has been published, including detailed methods, inclusion criteria, measurement procedures, and interventions [32].

To increase the fidelity of the trial, an additional nested mixed-methods study was set up (described in this paper) by administering questionnaires to the participants who received a tablet and coaching, as well as conducting follow-up interviews among a random selection of those participants [37].

Measurements

Before commencing the intervention, at baseline, the trial participants filled out a three-item questionnaire about their general experience with information and communication technology (ICT) devices, such as computers, smartphones, and tablets. To assess the long-term usability of a tablet (objective 1), after 6 months, participants who received a tablet were asked to fill out a usability questionnaire that was based on the Usefulness, Satisfaction, and Ease of use (USE) questionnaire [38]. The first part of the questionnaire consists of 23 items (Likert) that measure the following three components of usability: usefulness, satisfaction, and ease of use. Each item consists of a statement with the following five response options: strongly disagree, disagree, neither agree or disagree, agree, and strongly agree. The second part of the questionnaire contains three general questions about tablet use in the past 6 months, whether participants would recommend the tablet to friends, and an item participants could use for general remarks. All questionnaires were administered by paper and pencil in the Dutch language.

Furthermore, to explore how the tablet, in conjunction with the coach, supported the self-regulation of exercise behavior (objective 2), 18 participants were randomly selected for an in-depth interview. The interview questions were previously piloted among two participants. The interviews were conducted in a home setting, were held in Dutch, and lasted for about 45 minutes. All interviews were recorded.

Analysis

The questionnaires were processed with double-entry verification. For usefulness, satisfaction, and ease of use, separate mean scores were calculated. The mean scores could range from 1 (very low) to 5 (very high). The interviews were transcribed verbatim and subsequently double coded by two researchers. The directed content analysis method was used to explore how the participant's experience related to the five key constructs of self-regulation that the intervention was based upon (goal setting, action planning, behavior execution, self-monitoring, and evaluation). Directed content analysis was deemed more appropriate than conventional content analysis, because of the focus on existing theoretical constructs [39-41]. To minimize differences in interpretation, first calibration sessions were held. Subsequently, both researchers coded all

XSL•FO
RenderX

the transcripts independently with the key constructs of self-regulation and then compared the results. Differences were resolved via discussion. In the rare case no interrater consensus was reached, the first author settled the dispute.

Results

Questionnaire About ICT Experience and Skills

In total, 224 older adults with a mean age of 72 years (SD 7 years; 71% female) participated in the RCT at baseline. The questionnaire about the prior use of ICT devices and self-reported skill level was filled out by 29% (65/224) of the participants, of which 72% (47/65) were female. Their mean age was 71 years (SD 5.8 years). The tablet was one of the most popular devices among the participants. A large majority of the participants used this device several times a week. See Table 1 for the results.

Most participants rated themselves as somewhat skilled with ICT devices. Specifically, 6% (4/65) of the participants rated themselves as very unskilled, 2% (1/65) as unskilled, 38% (25/65) as somewhat skilled, 40% (26/65) as skilled, and 6% (4/65) as very skilled.

Table 1. Prior use of information and communication technology devices (N=65).

Device[a]	Use, n (%)				
	Never	Rarely[b]	Sometimes[c]	Regularly[d]	Often[e]
Personal computer	17 (26%)	1 (2%)	3 (5%)	8 (12%)	24 (37%)
Laptop	17 (26%)	3 (5%)	3 (5%)	7 (11%)	25 (39%)
Tablet	16 (25%)	2 (3%)	3 (5%)	3 (5%)	35 (59%)
Smartphone	13 (20%)	0 (0%)	2 (3%)	3 (5%)	41 (63%)
Mobile phone[f]	25 (39%)	1 (2%)	2 (3%)	5 (8%)	14 (22%)

[a]The values of individual items are less as some items were skipped by the participants.

[b]Once a year or less.

[c]Few times a year.

[d]Few times a month.

[e]Few times a week.

[f]Device without touchscreen.

Usability Questionnaire

The usability questionnaire was filled out by 36% (37/103) of the participants who had used the tablet for 6 months, of which 60% (22/37) were female. The mean age was 71 years (SD 5.1 years). The questionnaire had excellent internal consistency with Cronbach α of .89. The internal consistencies for the subscales were as follows: usefulness, .82; ease of use, .89; and satisfaction, .71.

Participants indicated that they found the tablet very useful (item 1) and it helped them to perform their exercises better (item 3) and safely (item 4). They were, however, neutral about how this affected their daily lives (item 6 and item 7). Overall, they were satisfied with the tablet and found it easy to use. Table 2 presents the results of the USE items. Finally, 68% (25/37) of the participants indicated that they would recommend the tablet to friends. Participants who indicated that they would not recommend the tablet provided varying reasons like “it didn’t work properly,” “I don’t need it to be active,” or “it’s too noncommittal.”

XSL•FO
RenderX

Table 2. Scores of usefulness, ease of use, and satisfaction items (5-point Likert scale) (N=37).

Questionnaire item[a]	Score[b], mean (SD)
Usefulness	3.8 (0.6)
1. The tablet is useful.	4.5 (0.6)
2. With the tablet, I can follow an individual exercise program that suits me.	4.0 (0.7)
3. The tablet helps me to perform my exercises better.	4.2 (0.7)
4. With the tablet, I can perform exercises safely.	3.9 (0.9)
5. The tablet helps me to perform exercises more often.	3.8 (1.1)
6. Since using the tablet, I have a more active life.	2.9 (1.0)
7. The tablet supports my daily activities.	3.0 (1.1)
8. The tablet has everything I need to be physically active.	3.6 (1.1)
9. The information about the exercises is understandable.	4.4 (0.6)
Ease of use	4.2 (0.6)
10. I learned to use the tablet quickly.	3.9 (1.1)
11. I easily remember how to use the tablet.	4.3 (0.9)
12. I am capable of using the tablet.	4.2 (0.8)
13. I can use the tablet without any help.	4.3 (0.7)
14. I understand how the tablet operates.	4.3 (0.7)
15. I can easily find what I am looking for on the tablet.	4.2 (0.9)
16. The tablet is easy to use.	4.4 (0.7)
17. Using the tablet is effortless.	4.4 (0.5)
Satisfaction	4.1 (0.6)
18. I am satisfied with the tablet.	4.2 (0.7)
19. The tablet is pleasant to use.	4.1 (0.8)
20. The tablet is fun to use.	4.0 (0.7)
21. I am going to keep on using the tablet.	4.0 (1.1)
Miscellaneous	
22. Family and friends believe I should use the tablet.	1.9 (0.9)
23. The trainer/coach believes that I should use the tablet.	3.3 (1.3)

[a]Translated from Dutch.

[b]The minimum score is 1, and the maximum score is 5.

Interviews

In total, 17% (18/103) of the participants who used the tablet for 6 months were approached for a follow-up interview. One participant declined without giving a specific reason. The interviews were conducted with the remaining 17 participants, of which 53% (9/17) were female. Their mean age was 73 years (SD 7.0 years). The results of the interviews are described below according to the following five phases of self-regulation: goal setting, action planning, behavior execution, self-monitoring, and evaluation.

Goal Setting

Goal setting involves the process of determining the objective a person aspires. Setting goals was the departure point of the blended intervention. The tablet was designed to support the participants by letting them rank a set of daily activities and subsequently formulating their goals. It was also the main topic of the first two meetings with the coach.

The participants' goals varied greatly (ie, from decreasing backache to improving balance). For some participants, the goal was not to improve physical health but to maintain it. Occasionally, participants formulated the goals with only the tablet, but most participants first consulted with the coach to explore related issues and translate top-level goals to specific and challenging, but realistic and measurable, goals. The attention the coach paid to the individual situation of the participant was appreciated. For instance, participants commented as follows:

> *…then we looked what is useful for me, what will help me to improve?... well, this was decided in consultation.* [Participant #14]

XSL•FO
RenderX

> *I believe it is important that the coach kept in mind: what does this person want to achieve?* [Participant #6]
>
> *well… they asked me about everything… which problems do I face?* [Participant #1]

In summary, the tablet, in conjunction with the coach, supported the participants in setting goals. The sensitivity of the coach for the personal circumstances was valued by the participants.

Action Planning

Action planning involves the process of making a plan regarding how the goals will be achieved. After determining the goals, participants could draw up a personal exercise schedule on the tablet. They could select functional exercises that would increase balance, strength, flexibility, and endurance. Each exercise was available in three variations that differed in difficulty.

The choice of different exercises was valued. For instance, participants commented as follows:

> *... then you always can choose your own exercises. I think it is great you have a lot of choice.* [Participant #4]
>
> *That's good. Then I can adjust it entirely to my own needs.* [Participant #6]

Some participants commented that customizing the exercise schedule was not easy to do, either because of technical limitations of the tablet or because of limited knowledge about the benefits of each exercise. In those cases, the coach was available to help. For instance, participants commented as follows:

> *…well, which exercise should you choose? … that I could do this together with my coach was very effective.* [Participant #4]
>
> *with his help I had in no time an entire exercise program.* [Participant #3]

The weekly overview of planned exercises helped the participants to be physically active. The majority of participants exercised daily. They commented that this was due to the intervention as follows:

> *I am chaotic and have no discipline, this helped me a lot!* [Participant #17]
>
> *I do the exercises every day at home. I did not do that before.* [Participant #13]
>
> *now I am consistently doing exercises, every day. Actually, because of this [tablet].* [Participant #10]

When asked about the underlying reason for this, they mentioned different aspects. Several participants indicated that the tablet provided them structure to build a routine. For many, this was doing the exercises at a fixed time of day, generally in the morning. Participants commented as follows:

> *before taking a shower and getting dressed, first those exercises. A fixed structure, that helped.* [Participant #2]
>
> *…well, that rhythm is a good feeling.* [Participant #15]

Others commented that the exercises were more integrated in their daily activities as follows:

> *…sometimes I also do the exercises as I go; then I walk step by step back into the living room after a visit to the bathroom.* [Participant #11]
>
> *…I do the exercises in between times. I stand on one leg when I am brushing my teeth for instance. Well, I kind of integrate it.* [Participant #14]

Besides providing structure, some participants mentioned that the tablet also acted as a cue to action as follows:

> *...when I sit down and see it [tablet] I think 'ah, a reminder!* [Participant #16]

In summary, participants felt that the blended intervention supported them in action planning. It provided them with structure to develop a routine. Several participants indicated that it helped them to do exercises daily, a frequency they previously did not achieve. They valued the possibility to personalize the exercise schedule to their own needs. The help of the coach was essential for some participants.

Behavior Execution

Behavior execution involves performing the actual behavior that should lead to achieving the goals. The tablet was designed to support this by various features like giving an overview of today's exercises, providing background information about each exercise along with video demonstrations, and providing a countdown timer or the ability to modify each exercise with three parameters (duration, number of repetitions, and intensity level).

Participants found the daily overview of exercises to be useful. It provided them in a brief glance which exercise had to be performed today and with what duration, repetition, and intensity level. The countdown timer was used especially in the beginning when participants had to familiarize themselves with the exercise routine. The same applied for the video demonstrations. It helped them to see how the exercises could be performed correctly. For instance, participants remarked as follows:

> *…but I did need it [video demonstration] to do it [the exercise] in the correct manner.* [Participant #6]
>
> *…that was nice, I could perform the exercises better this way.* [Participant #7]

Additionally, attention to safety was valued, with the following statements:

> *the exercises are safe. Well, at least a lot safer than riding a bike. Biking is dangerous.* [Participant #12]
>
> *…yes, attention was paid to this [safety]. That you had to hold on to something, when you stand on one leg, for instance.* [Participant #6]

One of the participants stressed the benefit of using a tablet for the instructions as follows:

> *I can write it down, but it's nice to have visual image of what is meant…. instructions written down are always subject to different interpretations. I think, as*

XSL•FO
RenderX

> *it has been done now, is very instructive.* [Participant #4]

Nevertheless, numerous participants mentioned that the additional instructions of the coach were also valuable as follows:

> *I also asked the coach, "am I doing it right?" He said "yes, that's right" or "you have to do it like this and that".* [Participant #11]

> *…that was nice. Sometimes he would demonstrate the exercise, or I would demonstrate it and ask him if I was doing it correct.* [Participant #6]

The coach also helped participants modify exercises if they were struggling with limitations or wanted more of a challenge. The latter was often needed. Many participants stressed that the exercises were too easy, despite the possibility to increase the difficulty level with the tablet. Apparently, this was not sufficient for numerous participants. An illustrative remark was as follows:

> *…yes, I can say that I wished they were a bit more challenging.* [Participant #10]

Some felt very strongly about this. For instance, a participant remarked as follows:

> *Look, I believe these exercises are meant for people who are in a retirement home and, more or less, don't do anything the entire day.* [Participant #15]

Two participants indicated that they stopped doing the exercises because of this reason. Others found creative ways, together with their coach, to increase the intensity level, for instance, by increasing the repetitions, skipping breaks, or adding weight. For instance, a participant commented as follows:

> *…such as the exercise with shopping bags… I added dumb-bells to it, now it's really challenging.* [Participant #10]

Finally, as participants developed a routine, they relied less on the tablet and on the coach for performing the exercises. Some participants kept on having the tablet in sight during the performance of the exercises, while others merely glanced at which exercises had to performed today and then executed them without the tablet. Watching the video demonstrations or using the countdown timer was not needed anymore. In some cases, participants even did all the exercises by heart, and one participant mentioned the following:

> *... I can do the exercises when I am at work, in between times. I just count the exercises myself.* [Participant #8]

When asked about the necessity of a coach, most of the participants felt that after 2 or 3 months, the coach's help was not needed anymore.

In summary, the video demonstrations and countdown timer helped the participants to perform the exercises safely and correctly, especially during the early stages of the intervention. The coach played an important role in adapting the exercises to meet the capacity of the participants, as many of them sought a bigger challenge. In time, the participants developed a routine and performed the exercises more autonomously.

Self-Monitoring

Self-monitoring involves the process of keeping track of one's progress. The tablet was designed to support this by letting users tick off exercises that had been completed. In a weekly overview, users could see which exercises had been done and which had not been done. Additionally, a progress bar indicated how many exercises still had to be done today and for the current week. Furthermore, the coach could remotely monitor the progress of the participants.

The moment at which participants ticked off exercises varied. Some did this directly after completing the exercises, whereas others did it at the end of the day. The majority of participants felt that keeping track in this manner gave them insights into their own behavior and was motivating. For instance, one participant made the following statement:

> *…for me it's very easy…it gives insight and lets me follow what I have done.* [Participant #9]

Remarkably, various participants expressed that the mere action of ticking off exercises was not only easy but also rewarding. It left them with a feeling of accomplishment. For instance, one participant made the following statement:

> *…look, in the end you want to finish off your list.* [Participant #8]

However, the progress bar, which indicated how many exercises were completed, was hardly used. Many participants did not seem to have noticed this feature, indicating a usability issue. A couple of participants also expressed the desire for more advance features to investigate their progress, like graphs and tables.

Several participants mentioned that remote monitoring by the coach was an important factor for them to keep doing the exercises. For instance, some participants remarked as follows:

> *…the tablet motivates me, but…I must say. I think this is also because…that it is being monitored.* [Participant #7]

> *…I think it helps… there is someone keeping an eye on you.* [Participant #14]

> *you are participating in study, you want to show that you are cooperating.* [Participant #10]

On the other hand, other participants indicated that this was not the case for them. They would keep doing the exercises if there was no coach involved.

In summary, keeping track of progress with the tablet was easy and motivating. Ticking off completed exercises was experienced as rewarding and gave participants insights into their progress. For some participants, the fact that they were remotely being monitored was motivating, while for others, the social presence of a coach was not important.

Evaluation

Evaluation involves the process of reflecting on the effort and the progress that has been made in relation to the goals that

XSL•FO
RenderX

were set out to be achieved. First, the tablet was designed to support the evaluation process by letting participants rate each exercise on three aspects (effort, complexity, and enjoyment). Second, either via video calls on the tablet or with face-to-face meetings, the participants had the opportunity to reflect on the progress together with their personal coach.

The ability to rate exercises with the tablet was superfluous according to several participants. The need to evaluate each exercise after completion seemed tedious. One participant made the following statement:

> *Well, look. This bothers me. I think 'come on guys. Everything is so easy and simple. For me there is no difference in it [the effort, complexity or enjoyment of the various exercises].* [Participant #2]

Some participants suggested that it would have been better if they could rate exercises on a weekly basis instead of on a daily basis or only when they felt the need to do so. In contrast, the participants were more positive about the evaluation with the coach. They felt that it helped them to identify issues. Several participants mentioned, however, that toward the end of the 6-month intervention, the coaching was not needed anymore.

Finally, some participants reported that they experienced an improvement in vitality. They found themselves to be in a better shape than before and attributed this to the blended intervention. One participant made the following comment:

> *…yes, I now really get up without any backache, although this was previously the case. The pain returns in the evening when I am tired, but in the morning it's different. That is a huge benefit.* [Participant #5]

Others did not notice an improvement, despite performing exercises, but also expressed more modest expectations. Maintaining their health status was more important than achieving progress for them, as indicated by the following remarks:

> *Do I notice an improvement in the gym? No. But if I don't do my exercises for a week or two…then I can notice the difference.* [Participant #6]

> *I notice, I am 85, that I am declining… my goal is to stay steady.* [Participant #4]

Another participant mentioned the following:

> *…when you are 18 you can expect to keep on getting better, but for me, after one year I am even more old again… Can I perform some exercises that I couldn't do before? Sure. In that sense there is progress. But it isn't so that I am going to keep on improving.* [Participant #16]

When asked if they would like to keep the tablet for exercising, the vast majority of participants expressed the wish to do so, regardless of whether they notice an improvement.

In summary, the blended intervention supported participants in evaluating their progress. Specifically, the conversations with the coach were responsible for this. Overall, the participants evaluated the blended intervention to be useful. Some felt that their health improved, but others did not have this feeling. Nevertheless, almost all participants indicated that they wanted to continue their exercise routine with support of the tablet.

Discussion

The Value of the Blended Approach

The objectives of this study were to assess the usability of the tablet and how it supported older adults in performing home-based exercises, in conjunction with a personal coach. A previous usability study showed that first-time users (age ranging from 69 to 99 years) could successfully complete various predefined tasks on the tablet during a 45-minute session in a laboratory [29]. This study extends those findings by showing that the tablet can be not only successfully operated in a standardized setting for a short period of time, but also useful, satisfying, and easy to use within the context of exercising at home during daily life for an extensive period of time. The participants indicated that the tablet allowed them to follow a tailored exercise program that suited them. It also helped them to perform the exercises more often, better, and safely. From the perspective of older adults, it can be concluded that the use of the tablet successfully supported them in their exercise behavior.

The interviews revealed a more detailed view on the underlying processes. The tablet was useful in developing an exercise routine. The tablet supported the participants in action planning and behavior execution by providing them with a tailored schedule that gave structure and video instructions demonstrating the appropriate behavior. Furthermore, ticking off exercises as a simple form of self-monitoring appeared to be motivating. On the other hand, the interviews revealed that the personal coach played an essential role. The interactive and social nature of coaching was especially useful during the self-regulation phases of goal setting and evaluation. The ability to interact with users in this manner is yet to be achieved by a virtual coach or avatar [24,42,43]. In addition, although the tablet allowed users to tailor the exercise program to their own needs, the exercises in the app did not sufficiently match the needs of the participants. The expertise of the coach was crucial for adapting the exercises to accommodate preferences. Finally, the presence of the coach in the form of remote monitoring was motivating for some participants. These findings are in line with other research showing that physical activity interventions incorporating access to a remote expert for advice and social support tend to be effective [34,44,45]. This study indicates that the coach might, in particular, be beneficial during the initial period, when participants familiarize themselves with the intervention and develop a routine.

Improvements

Although the participants were overall positive about the blended intervention, the evaluation also revealed several possibilities to improve the intervention. First, fit older adults should be able to add more challenging exercises to their schedule. Taking into account the preferences of some older adults, adding support for outdoor activities to the tablet would be enriching. For instance, a map with walking trails in the vicinity could stimulate older adults in achieving daily physical

XSL•FO
RenderX

activity. Second, the tablet should offer more detailed reports of user progress (eg, graphs that display long-term trends). Some participants requested such a feature. Third, the extent participants relied on a coach varied from person to person. Owing to the protocol of the RCT, coaches contacted the participants with a fixed frequency. When the intervention is implemented in practice, the intensity of counselling should be tuned to the preferences of individuals. Presumably, some older adults will extensively make use of counseling, while others will merely limit it to initial support.

Study Limitations

The aim of the blended intervention was to support older adults in performing home-based exercises. Questionnaires as well as interviews showed that older adults felt that the intervention accomplished this. However, an underlying assumption of the intervention was that regularly performing exercises would support older adults in their daily activities and lead to an active lifestyle and an increase in their vitality. The interviews showed mixed results on this topic, and no support was found for these assumptions from the questionnaires. More challenging exercises or a more comprehensive approach for the vitality of older adults might be needed for such secondary effects; however, the effectiveness of such strategies is also debatable [46-50].

The results of the questionnaires have to be interpreted with caution though. All older adults who participated in the clinical trial were given at the start of the trial a questionnaire about their prior experience with ICT devices. Only 65 of the 224 trial participants completed this baseline questionnaire. Furthermore, the 6-month trial had a 18% drop-out rate. Among the remaining 103 older adults who were part of the group that received a tablet, only 37 filled out the USE questionnaire about the usability of the tablet. The usability results may therefore be biased. Perhaps only participants who had a positive experience with the tablet filled out the usability questionnaire. We do not, however, think this is plausible. First, the baseline questionnaire had a high rate of nonresponses. This fact cannot be explained by a negative experience with tablet use in the blended intervention. The participants were yet to embark on the intervention when filling out the baseline questionnaire. The high nonresponse rate for both the questionnaires might have been caused by the numerous tests that were administered by the researchers as part of the larger clinical trial [32]. The testing procedure, including body measurements, took half a day. This might have led to fatigue, causing participants to skip questionnaires. Second, the positive evaluation based on the questionnaire is in line with the results from the interviews. Although the sample size of the interviews was small, it was not susceptible to selection bias. The interviews were based on random selection of participants. Only one participant declined to be interviewed. Therefore, the sample that was drawn for the interviews can be considered to be representative of the older adults participating in the intervention. The previous usability study among first-time users and the questionnaires and interviews of this study all point in the same general direction of a favorable evaluation.

The extent to which the findings can be generalized to older adults in general is a different issue. The baseline questionnaire showed that prior use of tablets was high among the participants. Studies have shown that among older adults, tablets are easier to operate than smartphones or personal computers owing to the large touchscreen [51-53]. This was one of the reasons to choose a tablet as the delivery device for the blended intervention [36]. The usability of tablets can also explain the increasing popularity of tablets among older adults. In the United States, tablet ownership among adults aged 65 years or older rose from 1% in 2010 to 32% in 2016 [54]. In the Netherlands, a similar trend has taken place, where tablet ownership among those aged 65 to 75 years grew from 28% in 2012 to 60% in 2016 [55]. In this light, the prior use of tablets among the participants of this study is representative of the larger population. Nevertheless, perhaps only older adults with a positive attitude about ICT in general or a tablet in particular signed up to participate in the blended intervention. More research is needed to assess how a wider range of older adults will experience a physical activity intervention that incorporates the use of tablets.

Conclusion

A mixed-methods process evaluation showed that older adults are positive about a blended intervention designed to support them in performing home-based exercises. Participants rated the adoption of a tablet as useful, satisfying, and easy. They indicated that it helped them to perform exercises better, more frequently, and safely. It supported them in various phases of self-regulation. The interactions with a personal coach strengthened this by offering deeper reflection and more fine-grained tailoring during the earlier stages of the intervention. A blended approach appears to be a promising strategy for delivering physical activity interventions in older adults.

Acknowledgments

The authors thank Samir van Rijn and Jordy Schouten for their valuable contributions to this study. This work is part of the research projects MOTO-B (Motivating Technology for Older Adults' Behavior; funded by the Dutch Research Council, grant number 023.006.013) and VITAMIN (VITal AMsterdam older adults IN the city; funded by the Amsterdam University of Applied Sciences' program Urban Vitality).

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. de Vries NM, van Ravensberg CD, Hobbelen JS, Olde Rikkert MG, Staal JB, Nijhuis-van der Sanden MW. Effects of physical exercise therapy on mobility, physical functioning, physical activity and quality of life in community-dwelling older adults with impaired mobility, physical disability and/or multi-morbidity: a meta-analysis. Ageing Res Rev 2012 Jan;11(1):136-149. [doi: 10.1016/j.arr.2011.11.002] [Medline: 22101330]
2. Walston J, Hadley E, Ferrucci L, Guralnik J, Newman A, Studenski S, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. J Am Geriatr Soc 2006 Jun;54(6):991-1001. [doi: 10.1111/j.1532-5415.2006.00745.x] [Medline: 16776798]
3. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
4. Tak E, Kuiper R, Chorus A, Hopman-Rock M. Prevention of onset and progression of basic ADL disability by physical activity in community dwelling older adults: A meta-analysis. Ageing Research Reviews 2013 Jan;12(1):329-338. [doi: 10.1016/j.arr.2012.10.001]
5. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
6. Taylor AH, Cable N, Faulkner G, Hillsdon M, Narici M, Van Der Bij AK. Physical activity and older adults: a review of health benefits and the effectiveness of interventions. J Sports Sci 2004 Aug;22(8):703-725. [doi: 10.1080/02640410410001712421] [Medline: 15370483]
7. King AC, Rejeski W, Buchner DM. Physical activity interventions targeting older adultsaaThis paper was a background paper for the Cooper Institute Conference Series Physical Activity Interventions, an ACSM Specialty Conference. American Journal of Preventive Medicine 1998 Nov;15(4):316-333. [doi: 10.1016/S0749-3797(98)00085-3]
8. King AC. Interventions to promote physical activity by older adults. J Gerontol A Biol Sci Med Sci 2001 Oct;56 Spec No 2:36-46. [doi: 10.1093/gerona/56.suppl_2.36] [Medline: 11730236]
9. Nelson M, Rejeski W, Blair S, Duncan P, Judge J, King A, American College of Sports Medicine, American Heart Association. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1094-1105. [doi: 10.1161/CIRCULATIONAHA.107.185650] [Medline: 17671236]
10. Stiggelbout M, Popkema DY, Hopman-Rock M, de Greef M, van Mechelen W. Once a week is not enough: effects of a widely implemented group based exercise programme for older adults; a randomised controlled trial. J Epidemiol Community Health 2004 Mar;58(2):83-88 [FREE Full text] [doi: 10.1136/jech.58.2.83] [Medline: 14729881]
11. Chodzko-Zajko WJ, Proctor DN, Fiatarone Singh MA, Minson CT, Nigg CR, Salem GJ, et al. Exercise and Physical Activity for Older Adults. Medicine & Science in Sports & Exercise 2009;41(7):1510-1530. [doi: 10.1249/mss.0b013e3181a0c95c]
12. van der Bij AK, Laurant M, Wensing M. Effectiveness of physical activity interventions for older adults: a review. Am J Prev Med 2002 Mar;22(2):120-133. [doi: 10.1016/s0749-3797(01)00413-5] [Medline: 11818183]
13. Lee J, Jung D, Byun J, Lee M. Effects of a Combined Exercise Program Using an iPad for Older Adults. Healthc Inform Res 2016 Apr;22(2):65-72 [FREE Full text] [doi: 10.4258/hir.2016.22.2.65] [Medline: 27200215]
14. Gschwind Y, Kressig R, Lacroix A, Muehlbauer T, Pfenninger B, Granacher U. A best practice fall prevention exercise program to improve balance, strength / power, and psychosocial health in older adults: study protocol for a randomized controlled trial. BMC Geriatr 2013 Oct 09;13:105 [FREE Full text] [doi: 10.1186/1471-2318-13-105] [Medline: 24106864]
15. King A, Haskell WL, Taylor CB, Kraemer HC, DeBusk RF. Group- vs home-based exercise training in healthy older men and women. A community-based clinical trial. JAMA 1991 Sep 18;266(11):1535-1542. [Medline: 1880885]
16. Gill TM, Baker DI, Gottschalk M, Peduzzi PN, Allore H, Byers A. A Program to Prevent Functional Decline in Physically Frail, Elderly Persons Who Live at Home. N Engl J Med 2002 Oct 03;347(14):1068-1074. [doi: 10.1056/nejmoa020423]
17. Mehra S, Dadema T, Kröse BJ, Visser B, Engelbert RH, Van Den Helder J, et al. Attitudes of Older Adults in a Group-Based Exercise Program Toward a Blended Intervention; A Focus-Group Study. Front Psychol 2016;7:1827 [FREE Full text] [doi: 10.3389/fpsyg.2016.01827] [Medline: 27920744]
18. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
19. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
20. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
21. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]

XSL•FO
RenderX

22. Schoeppe S, Alley S, Van Lippevelde W, Bray N, Williams S, Duncan M, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
23. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
24. Dasgupta D, Chaudhry B, Koh E, Chawla NV. A Survey of Tablet Applications for Promoting Successful Aging in Older Adults. IEEE Access 2016;4:9005-9017. [doi: 10.1109/ACCESS.2016.2632818]
25. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke C. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
26. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
27. Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
28. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
29. Mehra S, Visser B, Cila N, van den Helder J, Engelbert RH, Weijs PJ, et al. Supporting Older Adults in Exercising With a Tablet: A Usability Study. JMIR Hum Factors 2019 Mar 01;6(1):e11598 [FREE Full text] [doi: 10.2196/11598] [Medline: 30707106]
30. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
31. Zapata B, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Mar;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
32. van den Helder J, van Dronkelaar C, Tieland M, Mehra S, Dadema T, Visser B, et al. A digitally supported home-based exercise training program and dietary protein intervention for community dwelling older adults: protocol of the cluster randomised controlled VITAMIN trial. BMC Geriatr 2018 Aug 14;18(1):183 [FREE Full text] [doi: 10.1186/s12877-018-0863-7] [Medline: 30107787]
33. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
34. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Mar 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
35. Carver C, Scheier M. Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. Psychological Bulletin 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]
36. Mehra S, Visser B, Dadema T, van den Helder J, Engelbert RH, Weijs PJ, et al. Translating Behavior Change Principles Into a Blended Exercise Intervention for Older Adults: Design Study. JMIR Res Protoc 2018 May 02;7(5):e117 [FREE Full text] [doi: 10.2196/resprot.9244] [Medline: 29720358]
37. Shorten A, Smith J. Mixed methods research: expanding the evidence base. Evid Based Nurs 2017 Jul;20(3):74-75. [doi: 10.1136/eb-2017-102699] [Medline: 28615184]
38. Lund A. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6.
39. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
40. Assarroudi A, Heshmati Nabavi F, Armat MR, Ebadi A, Vaismoradi M. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. Journal of Research in Nursing 2018 Jan 10;23(1):42-55. [doi: 10.1177/1744987117741667]
41. Kirkwood K. Review: Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. Journal of Research in Nursing 2018 Jan 10;23(1):56-57. [doi: 10.1177/1744987117742209]
42. Khaghani-Far I, Nikitina S, Baez M, Taran EA, Casati F. Fitness Applications for Home-Based Training. IEEE Pervasive Comput 2016 Oct;15(4):56-65. [doi: 10.1109/mprv.2016.76]
43. Kwok Chi-Wai R, Hui Sai-Chuen S, Mak So-Ning T, Wu Ka-Shun P, Lee Wing-Kuen K, Wong Choi-Ki C. Can mobile virtual fitness apps replace human fitness trainer? 2011 Presented at: The 5th International Conference on New Trends in Information Science and Service Science; October 24-26, 2011; Macao, China URL: https://ieeexplore.ieee.org/abstract/document/6093392/
44. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Mar 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]

XSL•FO
RenderX

45. Geraedts H, Zijlstra A, Bulstra SK, Stevens M, Zijlstra W. Effects of remote feedback in home-based physical activity interventions for older adults: a systematic review. Patient Educ Couns 2013 Apr;91(1):14-24. [doi: 10.1016/j.pec.2012.10.018] [Medline: 23194823]
46. Steib S, Schoene D, Pfeifer K. Dose-Response Relationship of Resistance Training in Older Adults. Medicine & Science in Sports & Exercise 2010;42(5):902-914 [FREE Full text] [doi: 10.1249/mss.0b013e3181c34465]
47. Peterson M, Rhea M, Sen A, Gordon P. Resistance exercise for muscular strength in older adults: a meta-analysis. Ageing Res Rev 2010 Jul;9(3):226-237 [FREE Full text] [doi: 10.1016/j.arr.2010.03.004] [Medline: 20385254]
48. Spirduso W, Cronin DL. Exercise dose-response effects on quality of life and independent living in older adults. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S598-608; discussion S609. [doi: 10.1097/00005768-200106001-00028] [Medline: 11427784]
49. Taaffe D, Duret C, Wheeler S, Marcus R. Once-weekly resistance exercise improves muscle strength and neuromuscular performance in older adults. J Am Geriatr Soc 1999 Oct;47(10):1208-1214. [doi: 10.1111/j.1532-5415.1999.tb05201.x] [Medline: 10522954]
50. Denison HJ, Cooper C, Sayer AA, Robinson SM. Prevention and optimal management of sarcopenia: a review of combined exercise and nutrition interventions to improve muscle outcomes in older people. Clin Interv Aging 2015;10:859-869 [FREE Full text] [doi: 10.2147/CIA.S55842] [Medline: 25999704]
51. Murata A, Iwase H. Usability of touch-panel interfaces for older adults. Hum Factors 2005;47(4):767-776. [doi: 10.1518/001872005775570952] [Medline: 16553065]
52. Findlater L, Froehlich J, Fattal K, Wobbrock J, Dastyar T. Age-related differences in performance with touchscreens compared to traditional mouse input. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 2013; Paris, France URL: https://dl.acm.org/doi/10.1145/2470654.2470703 [doi: 10.1145/2470654.2470703]
53. Kobayashi M, Hiyama A, Miura T, Asakawa C, Hirose M, Ifukube T. Elderly User Evaluation of Mobile Touchscreen Interactions. 2011 Presented at: 13th IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT 2011; September 5-9, 2011; Lisbon, Portugal p. 83-99 URL: https://link.springer.com/chapter/10.1007%2F978-3-642-23774-4_9 [doi: 10.1007/978-3-642-23774-4_9]
54. Anderson M, Perrin A. Tech Adoption Climbs Among Older Adults. Pew Research Center. 2017. URL: https://www.pewinternet.org/2017/05/17/technology-use-among-seniors/ [accessed 2020-06-29]
55. 75-plussers sterkst groeiende groep internetters. CBS. 2016. URL: https://www.cbs.nl/nl-nl/nieuws/2016/52/75-plussers-sterkst-groeiende-groep-internetters [accessed 2020-06-29]

Abbreviations

MBvO: Meer Bewegen voor Ouderen; More Exercise for Seniors, a Dutch community-based physical activity program for adults aged 55 and above
mHealth: mobile health
MOTO-B: Motivating Technology for Older Adults' Behavior
MRC: Medical Research Council
RCT: randomized controlled trial
USE: Usefulness, Satisfaction, and Ease of use
VITAMIN: VITal AMsterdam older adults IN the city

Edited by G Eysenbach; submitted 24.09.19; peer-reviewed by K Ng, M Macaraig; comments to author 06.01.20; revised version received 11.02.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Mehra S, van den Helder J, Visser B, Engelbert RHH, Weijs PJM, Kröse BJA
Evaluation of a Blended Physical Activity Intervention for Older Adults: Mixed Methods Study
J Med Internet Res 2020;22(7):e16380
URL: http://www.jmir.org/2020/7/e16380/
doi:10.2196/16380
PMID:32459652

©Sumit Mehra, Jantine van den Helder, Bart Visser, Raoul H H Engelbert, Peter J M Weijs, Ben J A Kröse. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet

XSL•FO
RenderX

Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of a Web-Based Self-Management Program for Patients With Cardiovascular Disease: Explorative Randomized Controlled Trial

Marscha M Engelen[1], MSc; Sandra van Dulmen[2,3,4], PhD; Saskia Puijk-Hekman[1], MSc; Hester Vermeulen[1,5], PhD; Maria WG Nijhuis-van der Sanden[1], PhD; Sebastian JH Bredie[6], MD, PhD; Betsie GI van Gaal[1,5], PhD

[1]IQ Healthcare, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands

[2]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands

[3]Netherlands Institute for Health Services Research, Utrecht, Netherlands

[4]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

[5]Institute of Nursing, Faculty of Health and Social Studies, HAN University of Applied Sciences, Nijmegen, Netherlands

[6]Department of Internal Medicine, Radboud university medical center, Nijmegen, Netherlands

Corresponding Author:
Marscha M Engelen, MSc
IQ Healthcare
Radboud Institute for Health Sciences
Radboud university medical center
PO Box 9101
Nijmegen
Netherlands
Phone: 31 654272129
Email: marscha.engelen@radboudumc.nl

Abstract

Background: Web-based self-management programs have the potential to support patients with cardiovascular disease (CVD) in their self-management (eg, by focusing on behavior change and improving physical activity). The intervention mapping framework was used to develop a web-based program called Vascular View. The Vascular View program contained 6 modules (coping with CVD, setting boundaries, lifestyle, healthy nutrition, being physically active, interaction with health professionals) aiming to increase self-management behavior by tailoring to the perceived problems and (support) needs of patients after CVD.

Objective: The aim was to test the effectiveness of Vascular View before embarking on a full-scale randomized clinical trial (RCT) by evaluating the potential effectiveness and effect sizes of the Vascular View program and identifying outcome measures most likely to capture the potential benefits.

Methods: An explorative RCT was performed. Both control and intervention groups received care as usual and, in addition, the intervention group received 12 months of access to a web-based self-management program. Assessment occurred at baseline, 6 months, and 12 months. Outcome measures included general patient-reported outcome measurements: Illness Perception Questionnaire (IPQ), Rand-36, Patient Activation Measure, and patient self-efficacy. Module-specific patient-reported outcome measurements were Beliefs about Medicines Questionnaire, International Physical Activity Questionnaire, Dutch Healthy Diet Index, Fagerström Test for Nicotine Dependence (FTND), Alcohol Use Disorders Identification Test, and Perceived Efficacy in Patient-Physician Interaction. Linear mixed models for repeated measures using intention-to-treat and per-protocol analysis were applied to study differences between the patients in the intervention and control groups. Floor and ceiling effects were explored to give insight into the outcome measures most likely to capture the potential benefits.

Results: A total of 105 patients in the control group and 103 patients in the intervention group participated in the study. A positive direction of change between baseline and 12 months was shown for most outcome measurements in favor of the intervention group, of which 2 out of 10 outcomes showed a significant effect: attribution of cause of the disease to risk factors and immunity factors (IPQ) and dependency of nicotine (FTND). Floor and ceiling effects were seen in the IPQ, Rand-36, and the self-efficacy questionnaire.

XSL•FO
RenderX

Conclusions: No conclusion for the efficacy of the Vascular View program or selection of outcome measurements can be taken yet. A process evaluation will be conducted to gain thorough insight into the working elements of the program, patient needs in eHealth, and the use of the program by patients. This can determine for whom web-based self-management programs will work and help to adapt the program.

Trial Registration: Dutch Trial Register NTR5412; https://www.trialregister.nl/trial/5303

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.6352

(J Med Internet Res 2020;22(7):e17422) doi:10.2196/17422

KEYWORDS

explorative randomized controlled trial; cardiovascular diseases; self-management; eHealth support programs; internet; lifestyle; nursing

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, and survivors of CVD are at high risk for a secondary CVD event [1]. Prevention of secondary CVD events can be influenced by focusing on the interaction of behavioral risk factors (lifestyle components), metabolic risk factors (hypertension, high blood glucose levels, raised blood lipids, and overweight), and other risk factors (eg, advancing age, sex, stress, and depression) [1]. Managing these factors poses high demands on patient self-management skills. Despite support from health care professionals, patients have trouble managing their CVD and its consequences in daily life themselves [2]; they experience disease-related problems such as dealing with the symptoms, treatment, physical and psychological consequences, and recommended lifestyle changes.

Self-management of chronic illness has been widely recognized as a way to support patients in achieving the best possible quality of life with their chronic condition [3,4]. Secondary prevention programs targeting self-management through risk factors and (lifestyle) behavior change have been associated with reduced mortality, reduction of repeated cardiac events, and improved health-related quality of life [5,6]. However, attendance rate in cardiac rehabilitation programs remains less than 50% worldwide [7,8] due to barriers like lack of transportation, embarrassment of participation, a dislike of group environments, and work or domestic commitments [9,10]. The use of web-based self-management programs could be a remedy to these barriers, since these programs have the potential to reach a large group of patients and diminish barriers because of the possibility of accessing programs anytime and anywhere and providing anonymity. Buys and colleagues [11] found that patients with CVD showed a high interest in support through the internet (77%) and mobile phones (68%).

Recent systematic reviews have shown the high potential of electronic health (eHealth) interventions for CVD prevention [12-15]. Widmer and colleagues [14] demonstrated significant reduction on CVD outcomes due to digital health interventions (relative risk –0.61, 95% CI 0.46-0.80; P<.001; I^2=22%). Another systematic review studied behavioral change by using mobile health interventions: studies aiming to improve physical activity (n=2) or medication adherence (n=3) showed positive effects, but no effects were seen in studies aiming to decrease smoking (n=1) or change diets (n=1) [12]. Duff and colleagues [15] investigated the use of behavior change techniques (BCTs) in eHealth interventions for improving physical activity for patients with CVD: 8 of the 15 interventions showed significant improvements, while in 5 of the 15 studies the intervention group scored equal to the control group.

Not all studies recognize the potential of eHealth interventions, and demonstrate more questionable results. Hamine and colleagues [16] evaluated the effectiveness of mobile health (mHealth) interventions in supporting adherence of patients to chronic disease management (including CVD). Out of 41 randomized controlled trials (RCTs) that measured the effects of mHealth interventions on disease-specific clinical outcomes, significant improvements between groups were reported in only 16 studies (39%). Specific for cardiovascular risk factors, another review showed that the use of multiple modifiable internet lifestyle interventions in primary or secondary care is not superior to usual care [17]. Although 4 out of 9 studies demonstrated significant improvement in various risk factors, clinical relevance of these differences is questionable [17]. Inconsistent evidence is often due to limitations in the intervention and research design and a lack of power in the studies [18]. Limitations like small sample sizes, unclear description of intervention characteristics, short duration, and selection bias were noted in review studies [12,13].

To reduce the risk of limitations, we used the intervention mapping (IM) framework to develop a comprehensive, multicomponent web-based self-management program for patients with CVD called Vascular View (Vaat in Zicht in Dutch). IM contains 6 steps to design, implement, and evaluate an intervention based on the foundation of theoretical, empirical, and practical information [19]. The program Vascular View aims to increase self-management behavior tailored to the perceived problems and (support) needs of patients after CVD diagnosis [20]. Each of the 6 modules has a specific goal based on determinants of the I-change model [21]. For each of these selected determinants, BCTs were selected [22]. A detailed description of the development and content of Vascular View has been described elsewhere [20].

The last step of IM (step 6) is evaluation. This paper describes the testing of the effectiveness of Vascular View before embarking on a full-scale RCT. This explorative RCT study in patients with CVD aims to evaluate the potential effectiveness and effect sizes of the Vascular View program on 10 patient outcome measures and identify the outcome measures most

XSL•FO
RenderX

likely to capture the potential benefits of the Vascular View program.

Methods

Design

An explorative RCT was conducted at four outpatient clinics (cardiology, internal medicine, neurology, and vascular surgery) at Radboud University Medical Center in the Netherlands. To explore the efficacy of the web-based self-management program and identify suitable outcome measures, questionnaire data of patients allocated to the intervention and control groups were compared at 6 and 12 months after baseline on 10 outcome measurements related to the performance objectives of the intervention. Table 1 shows the modules, performance objectives, and related outcomes of the intervention. Since all outcomes are targeted through the intervention, positive effects are expected on all outcomes. The trial is registered in the Dutch Trial Register [NTR5412]. The medical ethics committee of Arnhem-Nijmegen in the Netherlands approved this study (registration number: 2015/1908).

Table 1. Modules, performance objectives, determinants, and related outcomes of the intervention.

Module	Performance objectives	Determinants	Module-specific outcomes
Coping with CVD[a] (3 sessions)	Patients have insight into CVD and accompanying symptoms and consequences. Patients cope with CVD and accompanying symptoms and consequences. Patients cope with (changed) sexuality and intimacy. Patients cope with stress in daily life. Patients cope with fear and emotions related to CVD. Patients cope with pain related to CVD. Patients adhere to medication instructions.	Knowledge, awareness, risk perception, attitude, self-efficacy, subjective norm, intention, action plans	BMQ[b], Self-efficacy (subscale: acceptation)
Setting boundaries in daily life (4 sessions)	Patients set boundaries. Patients adapt to changed circumstances. Patients ask for support from partner, relatives, and social environment. Patients cope with changed roles in family, job, and/or society. Patients are able to resume activities within their own possibilities.	Knowledge; awareness, attitude, self-efficacy, subjective norm	Self-efficacy (subscale: social environment), Self-efficacy (subscale: setting boundaries)
Lifestyle (4 sessions)	Patients refrain from tobacco use. Patients refrain from (harmful) alcohol use.	Knowledge, awareness, attitude, self-efficacy, subjective norm, intention, habits, skills	FTND[c], AUDIT[d], Self-efficacy (subscale: smoking), Self-efficacy (subscale: alcohol)
Healthy nutrition (3 sessions)	Patients eat healthy.	Knowledge, awareness, attitude, self-efficacy, subjective norm, intention, habits, skills	DHD[e], Self-efficacy (subscale: diet)
Being physically active in a healthy way (3 sessions)	Patients are physically active.	Knowledge, awareness, attitude, self-efficacy, subjective norm, intention, habits, skills	IPAQ[f], Self-efficacy (subscale: physical activity)
Interaction with health professionals (4 sessions)	Patients interact with health professionals.	Knowledge, awareness, attitude, self-efficacy, subjective norm	PEPPI-5[g], Self-efficacy (subscale interaction)

[a]CVD: cardiovascular disease.

[b]BMQ: Beliefs about Medicine Questionnaire.

[c]FTND: Fägerstorm Test for Nicotine Dependence.

[d]AUDIT: Alcohol Use Disorders Identification Test.

[e]DHD: Dutch Health Diet Index.

[f]IPAQ: International Physical Activity Questionnaire.

[g]PEPPI-5: Perceived Efficacy in Patient-Physician Interactions.

XSL•FO
RenderX

Participants

In the third quarter of 2015, 600 consecutive patients who had visited the outpatient clinic because of an established cardiovascular event were invited by the treating medical specialist to participate in this study. Inclusion criteria were (1) a cardiovascular disease (myocardial infarction, cerebrovascular disease [stroke included], peripheral artery disease, or combination); (2) a CVD event within 2 months to 1 year before start of the study; (3) aged 18 years or older; (4) able to read and understand Dutch; and (5) have access to a computer, internet, and an email account. Patients with a psychiatric disorder were excluded. Patients received information about the content and aim of the study, a short questionnaire to assess the inclusion criteria, and an informed consent form from the medical specialist via postal letter. Patients were asked to sign and return the informed consent form and completed questionnaire to the researcher (ME). When patients agreed to participate and were eligible, they received an invitation to complete the online questionnaire baseline data collection.

Randomization

Randomization took place after the baseline measurement and was stratified for four patient diagnoses: myocardial infarction, cerebrovascular disease (stroke included), peripheral artery disease, and aneurysm. The most recent diagnosis was used in randomization for patients with comorbidity in CVD. A blinded and independent statistician executed the randomization using SAS version 9.4 (SAS Institute Inc), which is an automated randomization program. The researcher (ME) informed patients about their assignment to control or intervention group. All patients in both groups continued with their care as usual: regular visits, treatment at the outpatient clinics, and standardized cardiovascular risk management, which contained an evaluation of cardiovascular risk factors, including feedback to optimize lifestyle. In addition to the care as usual, patients in the intervention group received 12 months access to the intervention (October 2015 until October 2016) directly after randomization.

Intervention

The Vascular View program was systematically developed in collaboration with CVD patients and health care professionals [20]. By defining performance and change objectives in conformity with the IM steps [19], 6 topics were distinguished and incorporated into the modules included in Vascular View: (1) coping with CVD, (2) setting boundaries in daily life, (3) lifestyle in general with specific attention to tobacco and harmful alcohol use, (4) healthy nutrition, (5) being physically active in a healthy way, and (6) interaction with health care professionals. Moreover, patients had access to two diaries (exercise and nutrition) in which patients could register their behavior to get insight into their exercise and nutrition routines. Each module comprised 3 or 4 sessions, which were personalized and supported by written information, tailored feedback, quotes from and videos of patients with CVD, pictures, and exercises. The Vascular View program started with information about the content and objectives of the 6 modules. The program was unguided but patients could complete an assessment and receive tailored advice about which of the 6 available modules was recommended for them.

Patients could visit the web-based self-management program and different modules as often as they wanted. Three groups were determined to give insight in the use of the program: nonusers, minimal users, or frequent users. Nonusers were patients that never or only once visited the program, minimal users visited the program 2 to 20 times, and frequent users visited more than 20 times. A detailed description of Vascular View, implementation, and the process evaluation is described elsewhere [20], and an overview of the program can be seen in Table 1.

Measurements and Outcomes

Baseline Characteristics

All patients who completed the baseline questionnaire between August 2015 and October 2015 received a questionnaire after 6 months (T1) and after 12 months (T2). At baseline, the following demographic and disease-related characteristics were collected: age, sex, educational level, work participation, cultural background, diagnosis and comorbidity, duration of illness, body weight, height, computer use, and experience with rehabilitation programs. Patient-reported outcome measurements were assessed at baseline and during follow-ups (T1 and T2). These outcomes could be distinguished in generic and module-specific outcomes (Table 1). The great amount of outcomes resulted in a long questionnaire with a great demand of time on participants. Therefore, the questionnaire at T1 was shortened by omitting two questionnaires: the Illness Perception Questionnaire (IPQ) and Rand-36. When patients preferred a paper questionnaire, a version was sent by post.

General Patient-Reported Outcome Measurements

Patient's illness attributions were assessed by the "causes of my illness" section on the IPQ. It contains 18 items measured on a 5-point Likert scale measuring four dimensions: psychological attributions, risk factors, immunity, and accident or chance. A higher score indicates a higher level of attribution to the dimension [23].

The patient's general health status was measured with the Rand-36, consisting of 36 items measuring 8 dimensions: physical functioning, social functioning, physical role limitations, emotional role limitations, mental health, vitality, pain, and general health perception [24]. The subscales physical and emotional role limitations have dichotomous items. The other subscales contained Likert scale items, with a higher score indicating better perceived health-related quality of life. All subscale scores were transformed to a 0-100 point scale.

The Patient Activation Measure (PAM-13), which includes statements about an individual's knowledge, confidence, and skills for self-management of their chronic illness behavior and the level of activation, was used to measure participants' self-management ability. The PAM-13 includes 13 items on a 5-point scale with a higher score indicating a higher level of patient activation [25,26].

No validated questionnaire was available that corresponded to the aims of the intervention; therefore, patient's self-efficacy

XSL•FO
RenderX

was measured with a self-developed questionnaire (see Multimedia Appendix 1 for an overview of the scales and examples of items). The aim of this questionnaire was to measure how confident patients felt about self-managing CVD based on the performance objectives and corresponding determinants (step 2 of the IM framework). Four patients from the expert group, who were involved in the development of the Vascular View-program [20], were asked to participate in the Think Aloud procedure, a technique used to evaluate the questionnaire [27]. The final questionnaire included 26 items measuring 8 subscales. The items were scored on a 4-point scale with a higher score indicating a higher level of confidence about self-managing CVD. For each subscale, Cronbach was calculated using the baseline, 6-month, and 12-month data of both groups (control and intervention). Each subscale is related to a module in the Vascular View program (see Table 1): acceptance (Cronbach =.81), social environment (Cronbach =.87), interaction with professionals (Cronbach =.85), physical activity (Cronbach =.83), diet (Cronbach =.84), smoking (Cronbach =.83), alcohol (Cronbach =.90), and setting boundaries (Cronbach =.78).

Module-Specific Patient-Reported Outcome Measurements

To measure patient attitudes toward their prescribed medicine, the first scale on the Beliefs about Medicines Questionnaire (BMQ) was used, which contains 10 items on a 5-point Likert scale. Two subscales (concerns and necessity) each contained 5 items that summed up to a scale score. A higher score reflects higher levels of concerns or feelings of necessity concerning the prescribed medicine [28,29].

Patient physical activity was measured with the International Physical Activity Questionnaire (IPAQ, short version) [30]. The IPAQ contains 7 questions divided in 3 subscales: walking, moderate intensity activity, and vigorous intensity activity. These subscales are described in minutes per week.

The Dutch Healthy Diet Index is a 34-item questionnaire to estimate adherence to the 2006 Dutch guidelines for a healthy diet, containing 8 components: vegetables, fruit, dietary fiber, fish, saturated fats, trans fats, natrium, and alcohol. Per component the score ranges between 0 and 10, resulting in a total score between 0 (no adherence) and 80 (complete adherence) [31].

Patient tobacco dependence was measured with the Fagerström Test for Nicotine Dependence. This questionnaire consists of 6 items resulting in a total score between 1 and 10, in which a higher score reflects more dependence of nicotine [32].

Alcohol use was measured by the 3-item Alcohol Use Disorders Identification Test with a total score from 0 to 12. A score of 5 or higher indicates the possibility of increasing risk and higher risk of alcohol drinking [33].

Patient interaction with health care professionals was measured with the Perceived Efficacy in Patient-Physician Interaction, which contains 5 items on a 5-point Likert scale that are summed to determine the total score. A higher score reflects more confidence of the patient in interactions with their physician [34,35].

Biomedical Measurements

Electronic patient dossiers were searched for biomedical data on two time points: April 2015 to December 2015 (baseline measurement) and April 2016 to March 2017 (12-month measurement). The search of biomedical data included weight, BMI, systolic and diastolic blood pressure, total cholesterol, low-density lipoprotein–cholesterol, high-density lipoprotein–cholesterol, triglycerides, and non–high-density lipoprotein–cholesterol.

Statistical Analysis

For explorative RCTs such as this, sample sizes are not calculated based on formal power analyses. Therefore, a sample size of 200 patients was chosen for this trial, which was considered a sufficient size for a representation of the relevant variation in the target group. All quantitative data were analyzed using SPSS Statistics version 25 (IBM Corporation). Descriptive analyses were used to describe the control and intervention groups at baseline. The differences between patient characteristics in the intervention and control group were tested using *t* tests and chi-square tests. A *P* value of <.05 was determined as statistically significant in all analyses.

A linear mixed-model analysis with repeated measures on intention-to-treat (ITT) basis was used to determine the differences in outcome measures between the intervention and control group. In this model, the outcomes were the dependent variables and the patient was the random factor. The fixed factors were group (intervention/control) and time and the interaction between time and group. This method automatically uses the missing at random assumption to handle missing data.

Subsequently, a per-protocol analysis (PPA) was performed to compare the control and intervention group with only those patients who completed the treatment originally allocated. Patients in the control group were included if they completed the questionnaire at baseline and 6 and 12 months. Patients in the intervention group were included if they completed the questionnaire at baseline and 6 and 12 months and used Vascular View at least once. Moreover, the same tests and linear mixed-model analyses were used as described in the ITT section.

Identifying Outcome Measures

To identify outcome measures most likely to capture the potential benefits, floor and ceiling effects were explored for all outcome measurements using Likert scales at baseline. A floor effect indicates that most of the participants score near the minimum score and a ceiling effect indicates that most participants score near the maximum score on a questionnaire.

In this calculation, we considered floor and ceiling effects exceeding 20% to be significant [36]. Thereby, spaghetti plots were used to analyze the changes between T0 and T2 for all individuals.

Results

Participants

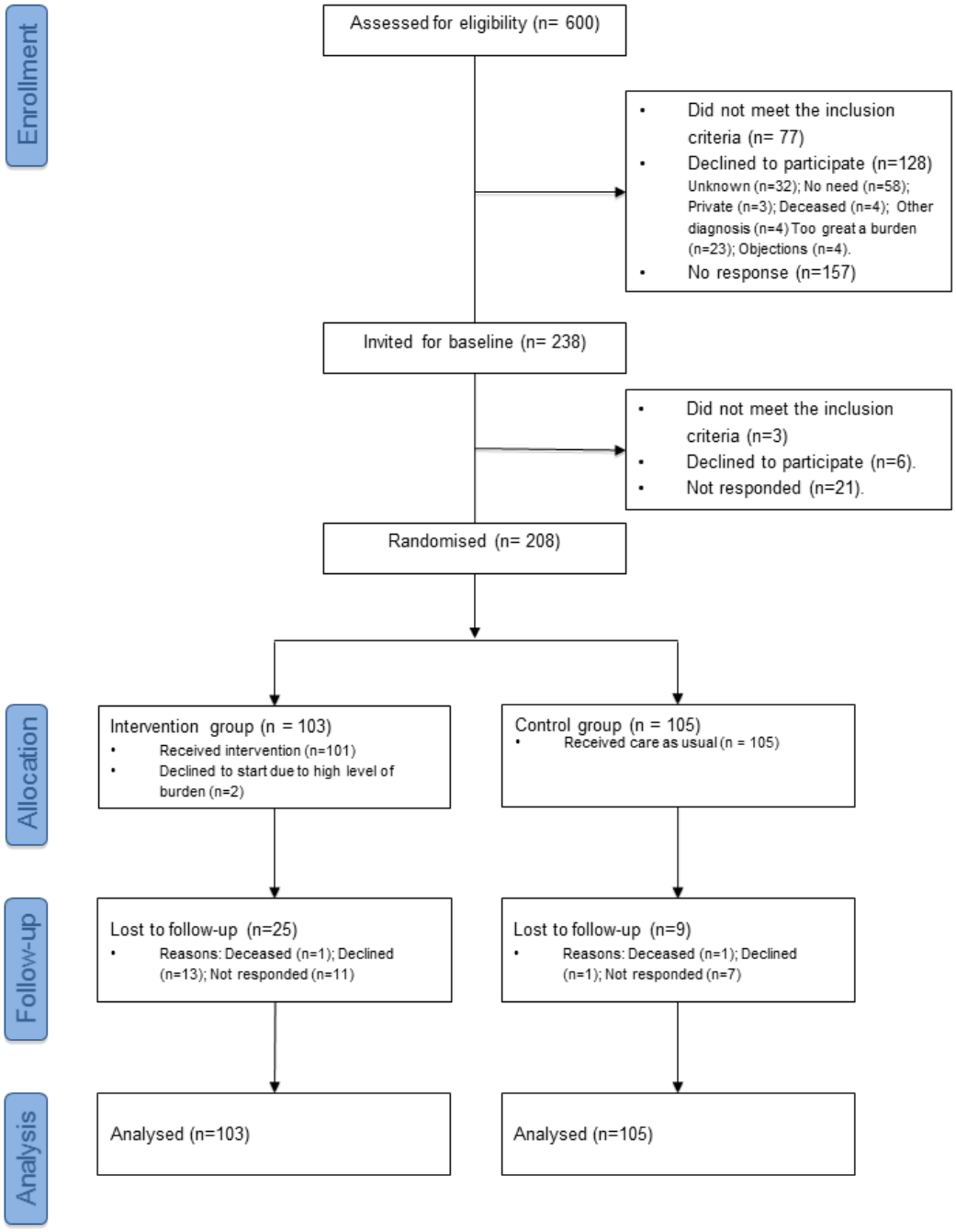
In total, 600 patients were eligible and invited by the medical specialist. Of these, 238 patients expressed interest, and 208

XSL•FO
RenderX

(87.4%) participated in the study (see Figure 1). Of the 30 patients who were not included, 3 did not meet the inclusion criteria, 6 declined, and 21 did not respond to the invitation. The 208 participating patients were randomized to the intervention (n=103) or control group (n=105) and stratified by CVD diagnosis. Two intervention group patients completed the baseline questionnaire but declined to participate in the intervention. At T1, 86 patients in the intervention group and 103 in the control group completed the questionnaire. At T2, 78 patients in the intervention group and 96 in the control group completed the questionnaire. More intervention group patients (25/103, 24.3%) compared with control group patients (9/105, 8.6%) were lost to follow-up.

All patients in the intervention group (n=101) had access to Vascular View, of which 37.6% (38/101) did not visit the program or only once, 27.7% (28/101) visited the program 2 to 20 times, and 34.7% (35/101) visited the program more than 20 times. The range of visits was between 1 and 43 visits per participant.

Figure 1. Flowchart of Vascular View study.



XSL•FO
RenderX

Outcomes of the Intention-To-Treat Analysis

Baseline Characteristics of Patients

Demographic and disease-related characteristics at baseline were compared for the control and intervention group, as shown in Table 2. The mean age of the patients in the intervention group was 63.3 (SD 10.0) years and 63.7 (SD 9.8) years in the control group. Both groups had more men than women: 69 men in the control and 73 in the intervention group. There were no statistically significant differences between the control and intervention group.

Table 2. Intention-to-treat analysis: patient characteristics at baseline.

Characteristics	Control (n=105)	Intervention (n=103)
Sex (male), n (%)	69 (66.0)	73 (71.0)
Age in years, mean (SD)	63.7 (9.8)	63.3 (10.0)
BMI (kg/m^2), mean (SD)	28.4 (4.8)	27.2 (4.9)
Disease duration, mean (SD)	4.9 (8.2)	4.3 (7.7)
Education level, n (%)		
Low	24 (22.9)	17 (16.5)
Intermediate	39 (37.1)	34 (33.0)
High	42 (40.0)	52 (50.5)
Work participation (yes), n (%)	35 (33.3)	40 (38.8)
Diagnosis group, n (%)		
Myocardial infarction	57 (54.3)	58 (56.3)
Cerebrovascular disease (stroke included)	35 (33.3)	33 (32.0)
Peripheral artery disease	13 (12.4)	12 (11.7)
Comorbidity within CVD[a] (yes), n (%)	21 (20.0)	18 (17.5)
Comorbidity, other (yes), n (%)	32 (30.5)	30 (32.0)
Cultural background (Dutch), n (%)	101 (96.2)	100 (97.1)

[a]CVD: cardiovascular disease.

Biomedical Measurements

An overview of the mean scores and standard deviations of patients in the intervention and control group at baseline and after 12 months is presented in Table 3. No statistically significant differences were found between the control and intervention group.

The biomedical data show a high level of missing values in both the control and intervention group. For example, triglycerides show a large group of missing values: 45% of values are gathered at T2 compared with T0.

XSL•FO
RenderX

Table 3. Biometrics of control and intervention group.

Biometrics	Control group (n=105)				Intervention group (n=103)			
	T0[a]		T2[b]		T0		T2	
	n	mean (SD)	n	mean (SD)	n	mean (SD)	n	mean (SD)
Weight (kg)	81	86 (16)	56	90 (20)	84	85 (15)	46	86 (15)
BMI (kg/m^2)	81	28.3 (4.5)	56	29.3 (5.3)	83	27.3 (4.0)	46	27.5 (4.3)
Systolic BP[c] (mm Hg)	96	135 (19)	84	135 (15)	94	135 (17)	68	134 (18)
Diastolic BP (mm Hg)	96	79 (10)	84	77 (11)	94	79 (9)	68	77 (10)
Total cholesterol	87	4.2 (1.1)	55	4.0 (0.9)	83	4.3 (1.2)	47	4.2 (1.1)
LDL-C[d]	67	2.2 (0.8)	33	2.1 (0.8)	65	2.3 (0.8)	29	2.2 (0.9)
HDL-C[e]	87	1.2 (0.3)	54	1.2 (0.4)	83	1.1 (0.3)	47	1.2 (0.3)
Triglycerides	71	1.9 (1.6)	33	1.8 (1.4)	69	1.9 (1.2)	30	1.9 (1.3)
non–HDL-C	87	3.0 (1.1)	53	2.7 (0.9)	82	3.2 (1.1)	46	3.0 (1.1)

[a]T0: baseline.

[b]T2: 12 months.

[c]BP: blood pressure.

[d]LDL-C: low-density lipoprotein–cholesterol.

[e]HDL-C: high-density lipoprotein–cholesterol.

Patient Outcome Measurements at Baseline and Follow-Up

Multimedia Appendix 2 provides an overview of the mean scores and standard deviations of outcome measurements of the patients in the intervention and control group at baseline and after 6 and 12 months. At baseline, the intervention group had more patients who smoked (n=13) compared with the control group (n=5), which was a statistically significant difference (mean difference 0.4; *P*=.04). Thereby, patients in the intervention group scored at baseline significantly higher on self-management behavior (PAM-13) compared with patients in the control group (mean difference 4.4; *P*=.03).

The estimated differences and *P* values between the intervention and control group after ITT analysis at 6 and 12 months after baseline are presented in Multimedia Appendix 2. Six out of 29 variables in the intervention group and 13 out of 29 in the control group decreased over time. Two questionnaires showed significant effects in favor of the intervention group. First, the IPQ scores showed that intervention group patients attributed the cause of their disease more to risk factors and immunity factors after 12 months. Patients in the control group showed a decrease on attribution to risk factors and a small increase on immunity factors. Second, the intervention group showed a statistically significant decrease for the dependency of nicotine after 6 months (–1.55; *P*=.01) and after 12 months (–1.67; *P*=.01) with respect to T0.

Outcomes Per-Protocol Analysis

Baseline Characteristics of Patients

There were no significant between-group differences found in the demographic and disease-related characteristics at baseline for the control and intervention group of the patients following the intervention per protocol (Table 4).

XSL•FO
RenderX

Table 4. Per-protocol analysis: patient characteristics at baseline.

Characteristics	Control (n=95)	Intervention (n=52)
Sex (male), n (%)	62 (65.3)	36 (69.2)
Age in years, mean (SD)	63.9 (9.4)	61.8 (9.3)
BMI (kg/m^2), mean (SD)	28.5 (4.9)	27.2 (5.2)
Disease duration, mean (SD)	4.7 (7.9)	4.0 (8.2)
Education level, n (%)		
Low	23 (24.2)	12 (23.1)
Intermediate	34 (35.8)	12 (23.1)
High	38 (40.0)	28 (53.8)
Work participation (yes), n (%)	19 (36.5)	29 (30.5)
Diagnosis group, n (%)		
Myocardial infarction	51 (53.7)	34 (65.4)
Cerebrovascular disease (stroke included)	32 (33.7)	14 (26.9)
Peripheral artery disease	12 (12.7)	1 (7.7)
Comorbidity within CVD[a] (yes), n (%)	19 (20.0)	6 (11.5)
Comorbidity, other (yes), n (%)	29 (30.5)	17 (32.7)
Cultural background (Dutch), n (%)	92 (96.8)	51 (98.1)

[a]CVD: cardiovascular disease.

Patient Outcome Measurements at Baseline and Follow-Up

Multimedia Appendix 3 gives an overview of the PPA with mean scores, standard deviations, effect sizes, and *P* values of outcome measurements of the patients in the intervention and control group at baseline and after 6 and 12 months. At baseline, patients in the intervention group scored significantly higher on self-management behavior (PAM-13) compared with patients in the control group (mean difference 5.6; *P*=.03). The subscale social environment showed a significantly higher score on self-efficacy for intervention group patients compared with the control group (mean difference 0.2; *P*=.04) at baseline.

The PPA showed effects on the same outcomes as the ITT analysis. The IPQ scores on risk factors (0.31; *P*=.02) and immunity factors (0.48; *P*<.001) showed a statistically significant difference between intervention and control group: the IPQ scores show that after 12 months, patients in the intervention group attributed the cause of their disease more to risk factors and immunity factors. Patients in the control group showed a decrease on attribution to risk factors and no change on immunity factors. Patients in the intervention group decreased the dependency of nicotine significantly toward the control group after 6 months (–1.87; *P*=.01) and after 12 months (–1.72; *P*=.02).

Suitable Outcome Measures

The percentages of patients scoring 0 (floor effects) or full marks (ceiling effects) on the 10 outcome measures were assessed at baseline. Significant floor effects were seen on the subscales IPQ psychological attributions (22.0%), IPQ immunity (35.3%), and IPQ accident or chance (26.6%). Significant ceiling effects were noticed on the subscales Rand social functioning (29.3%), Rand bodily pain (29.3%), patient self-efficacy with a self-constructed 26-item questionnaire (SE) interaction (26.9%), SE physical activity (33.8%), SE diet (30.9%), SE alcohol (35.2%), and SE setting boundaries (23.7%).

Discussion

Principal Findings

This explorative RCT aimed to evaluate the potential effectiveness and effect sizes of the Vascular View program on 10 patient outcome measures and identify the outcome measures most likely to capture potential benefits. The evaluation of potential effectiveness of Vascular View showed significant effects for illness attribution and nicotine dependence. At 12 months, patients in the intervention group attributed the cause of their diseases more often to risk factors and immunity factors than patients in the control group. Intervention group patients showed less dependency of nicotine after 6 and 12 months compared with the control group. It should be noted that we are not convinced that this effect was clinically meaningful due to the small number of participants who stopped or decreased smoking and because the effect is partly due to a high dropout of smokers in the intervention group (4 out of 13) compared with no dropout in the control group and one patient in the control group who started smoking after the baseline measure. Although the other outcomes showed no statistically significant differences between the intervention and control group, there seems to be a positive trend in the improvement of outcome measurements in favor of the intervention group. Overall, we did not expect to find a nonconvincing trend regarding the efficacy of Vascular View, since the theory-based intervention

XSL•FO
RenderX

was thoroughly designed with patients and health care professionals on the basis of patient support needs according to the IM steps [19]. Furthermore, the outcome measures were selected carefully and in line with the objectives of the intervention. Randomization was successful, and the number of missing values was limited.

To our knowledge, Vascular View is the first web-based self-management program for secondary care patients with CVD that aims to improve so many components and in which patients can decide which modules they want to use. Other (effective) eHealth interventions for CVD prevention focused primarily on one or two risk factors (eg, physical activity [15] and smoking cessation [37]). Results of web-based self-management programs with multiple components in other diseases show comparable results [38-40]. Vascular View tried to guide patients by using a questionnaire to advise them in choosing courses in the welcome module and lifestyle module [20]. However, the large amount of topics might have resulted in an overload of information and subsequently demotivation and minimal use of the program. The lack of efficacy of the Vascular View program is not in line with studies that show the potential of eHealth interventions for secondary prevention of CVD [12-16]. For example, Vascular View addressed patients' intrinsic processes through the determinants, which is seen as a successful strategy for face-to-face self-management programs [41]. However, it is hard to compare studies on eHealth self-management programs because of the various eHealth approaches and outcome measures available. Review studies show a large diversity in studies, looking at outcome measures, use of eHealth interventions, implementation, etc.

A discrepancy between needed self-management support by patients and provided self-management support by nurses is familiar in health care. Vascular View is an unguided self-management program in which nurses did not discuss the Vascular View program with patients, the program was not part of consultation, and nurses did not encourage patients to continue use of the program. The question should be raised whether use of the intervention should be supported and perhaps even used as a partial replacement of usual care. Providing self-management support is a core task of nurses in outpatient clinics, and patients expect health care professionals to fulfill a comprehensive role [42]. However, nurses seem to experience barriers in discussing all self-management categories (symptom management, treatment, biomedical cardiovascular risk factors, psychosocial consequences, and lifestyle changes). Physical components are often discussed, but psychological components are left behind [43]. Web-based self-management interventions could be a way to improve self-management support by focusing on the patient instead of the health care professional. More research needs to be conducted to determine the balance between support by the nurse and a self-guided self-management program.

Only half of the intervention group (n=51) adhered to the intervention: they completed all questionnaires and used Vascular View once or more often. Moreover, 38% did not use the program or used it only once. Although a previous questionnaire study showed that 77% of patients with CVD were interested in support through the internet [11], it seems to be important to match expectations of patients with the online program [44]. When patient expectations did not line up with the online program, patients refused to complete the measurements and revisit the online program [44]. Three striking differences between patients in the ITT and PPA give insight in the difference between users and nonusers. First, mostly patients with an intermediate education level drop out (difference between ITT and PPA: 9.9%) and a relatively high number of patients with a lower level of education used the program (difference between ITT and PPA: 6.6%). These results are contrary to other research suggesting that lower educational levels are a barrier for eHealth use [45]. Second, the percentage of patients with myocardial infarction is higher in the intervention group of the PPA (65.4%) compared with the ITT (56.3%). This suggests a higher need for self-management support in patients with a myocardial infarction diagnosis. Third, the Vascular View program seems to be more popular in patients scoring higher on patient-related outcomes such as self-management, physical activity, quality of life, diet, nicotine adherence, and alcohol adherence.Improvement is more difficult to achieve in this group.

The second aim was to identify the outcome measures most likely to capture the potential benefits. Most of the chosen outcome measures were likely to capture the potential effect, although floor and ceiling effects were seen in the IPQ, Rand-36, and patient's self-efficacy questionnaire. Patient's illness attributions (IPQ) [23] showed a right-skewed distribution, meaning that more than 20% of the patients attributed their disease not to psychological factors, immunity, accident, or chance at all. However, effects after 12 months were seen on the IPQ since the intervention group increased attribution to psychological factors, immunity, and risk factors. In spite of the floor effects, the IPQ is a valuable measurement to give insight in the efficacy of the Vascular View program. A left-skewed distribution was found in 2 subscales of the Rand-36 [24] (social functioning and bodily pain), indicating that more than 20% of the patients already experienced a high quality of life. Since CVD patients show relatively low pain levels and physical symptoms, for future research we suggest an instrument with a higher sensitivity such as the Seattle Angina Questionnaire [46,47]. The self-efficacy questionnaire also showed a left-skewed distribution, suggesting high self-efficacy on interaction, physical activity, diet, alcohol, and setting boundaries. More research should be conducted to determine the reliability and validity of this self-developed questionnaire. To conclude, a more sensitive instrument for quality of life is suggested for future research. All other questionnaires were sensitive to measure change in this population.

Limitations

We believe this explorative RCT has numerous methodological strengths, although some limitations need to be mentioned. First, more patients in the intervention compared with the control group were lost to follow-up. However, we have not established that patients with specific characteristics dropped out. Second, this research was conducted in the outpatient clinic of a university hospital in which high quality of CVD care is already delivered, and therefore it may be harder to achieve improvement. Last, all patients in the outpatient group were

XSL•FO
RenderX

informed but only the ones who were interested in the study were invited to participate. A limitation is that this might be a biased group because they were probably already interested in self-management.

Recommendations for Clinical Practice and Research

For future studies, we recommend further studying patient self-management needs and the possibilities of tailoring eHealth. Although Vascular View was developed using IM on the basis of the support needs of patients with CVD [20], individual participating patients were not asked for their intentions to improve self-management and change their behavior and what kind of support needs they preferred. In our study, patients with higher starting levels (eg, self-management) were more likely to use the program. So, the readiness to change or phase of the disease might be predictors for using the program (and the needed support type). To conclude, to increase use and efficacy of eHealth programs, more insight into characteristics of patients who could benefit from web-based self-management programs is needed. A process evaluation will be conducted and published using the components of Saunders (fidelity, dose, reach, recruitment, and context) to get more insight into the low compliance with and noneffectiveness of the intervention and fine-tune the inclusion criteria. Furthermore, we believe that the program should be embedded in treatment and supplement self-management support provided by nurses. More research needs to be conducted to determine the balance between nurse support and self-guided self-management programs.

Conclusions

This study contributes to our understanding of self-management support for patients with CVD using eHealth apps. Although we believe in the potential of the Vascular View program, there is no conclusive evidence for the efficacy. Using an unguided self-management program might not work for everyone, and the program might need to be embedded in health care more firmly. A detailed process evaluation of the program should be conducted to gain thorough insight into the working elements of the program, patient needs in eHealth, and the use of the program by patients. Finally, it should be investigated how Vascular View can be more tailored to the patient needs and become more embedded in treatment.

Acknowledgments

This study is funded by ZonMw, the Netherlands Organization for Health Research and Development (520001001). The authors would like to thank the patients with CVD and health care professionals for their contribution in the expert groups.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Self-efficacy questionnaire.
[DOCX File , 13 KB - jmir_v22i7e17422_app1.docx]

Multimedia Appendix 2
Intention-to-treat analysis: mean scores, standard deviation, effect sizes and *P* values of the outcome measures.
[DOCX File , 19 KB - jmir_v22i7e17422_app2.docx]

Multimedia Appendix 3
Per-protocol analysis: mean scores, standard deviation, effect sizes and *P* values of the outcome measures.
[DOCX File , 19 KB - jmir_v22i7e17422_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (v.1.6.1).
[PDF File (Adobe PDF File), 2521 KB - jmir_v22i7e17422_app4.pdf]

References

1. Mendis S, Puska P, Norrving B. Global atlas on cardiovascular disease prevention and control. World Health Organization. 2011. URL: http://apps.who.int/iris/bitstream/10665/44701/1/9789241564373_eng.pdf?ua=1 [accessed 2020-06-03]
2. Helping people help themselves: a review of the evidence considering whether it is worthwhile to support self-management. The Health Foundation. 2011. URL: https://www.health.org.uk/sites/default/files/HelpingPeopleHelpThemselves.pdf [accessed 2020-06-03]
3. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 2002 Oct 16;288(15):1909-1914. [Medline: 12377092]
4. Clark NM. Management of chronic disease by patients. Annu Rev Public Health 2003;24:289-313. [doi: 10.1146/annurev.publhealth.24.100901.141021] [Medline: 12415147]

XSL•FO
RenderX

5. Müller-Riemenschneider F, Meinhard C, Damm K, Vauth C, Bockelbrink A, Greiner W, et al. Effectiveness of nonpharmacological secondary prevention of coronary heart disease. Eur J Cardiovasc Prev Rehabil 2010 Dec;17(6):688-700. [doi: 10.1097/HJR.0b013e32833a1c95] [Medline: 20458247]
6. Whalley B, Thompson DR, Taylor RS. Psychological interventions for coronary heart disease: Cochrane systematic review and meta-analysis. Int J Behav Med 2014 Feb;21(1):109-121. [doi: 10.1007/s12529-012-9282-x] [Medline: 23179678]
7. Turk-Adawi KI, Grace SL. Narrative review comparing the benefits of and participation in cardiac rehabilitation in high-, middle- and low-income countries. Heart Lung Circ 2015 May;24(5):510-520. [doi: 10.1016/j.hlc.2014.11.013] [Medline: 25534902]
8. Bjarnason-Wehrens B, McGee H, Zwisler A, Piepoli MF, Benzer W, Schmid J, et al. Cardiac rehabilitation in Europe: results from the European Cardiac Rehabilitation Inventory Survey. Eur J Cardiovasc Prev Rehabil 2010 Aug;17(4):410-418. [doi: 10.1097/HJR.0b013e328334f42d] [Medline: 20300001]
9. Neubeck L, Redfern J, Fernandez R, Briffa T, Bauman A, Freedman SB. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review. Eur J Cardiovasc Prev Rehabil 2009 Jun;16(3):281-289. [doi: 10.1097/HJR.0b013e32832a4e7a] [Medline: 19407659]
10. Dalal HM, Zawada A, Jolly K, Moxham T, Taylor RS. Home based versus centre based cardiac rehabilitation: Cochrane systematic review and meta-analysis. BMJ 2010;340:b5631 [FREE Full text] [Medline: 20085991]
11. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Jul 19;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]
12. Pfaeffli DL, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
13. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
14. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
15. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity ehealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
16. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
17. Vegting IL, Schrijver EJM, Otten RHJ, Nanayakkara PWB. Internet programs targeting multiple lifestyle interventions in primary and secondary care are not superior to usual care alone in improving cardiovascular risk profile: a systematic review. Eur J Intern Med 2014 Jan;25(1):73-81. [doi: 10.1016/j.ejim.2013.08.008] [Medline: 24011771]
18. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
19. Bartholomew L, Markham C, Ruiter R, Fernandez M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th Edition. San Francisco: Jossey-Bass; 2016.
20. Puijk-Hekman S, van Gaal BG, Bredie SJ, Nijhuis-van der Sanden MW, van Dulmen S. Self-management support program for patients with cardiovascular diseases: user-centered development of the tailored, web-based program vascular view. JMIR Res Protoc 2017 Feb 08;6(2):e18 [FREE Full text] [doi: 10.2196/resprot.6352] [Medline: 28179214]
21. de Vriesh H, Kremers SPJ, Smeets T, Brug J, Eijmael K. The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors. Am J Health Promot 2008;22(6):417-425. [doi: 10.4278/ajhp.22.6.417] [Medline: 18677882]
22. De Bruin M, Schaalma H, Hospers H. Coding manual for behavioral change techniques. URL: https://osf.io/bsvxc/download [accessed 2020-06-03]
23. Moss-Morris R, Weinman J, Petrie K, Horne R, Cameron L, Buick D. The Revised Illness Perception Questionnaire (IPQ-R). Psychol Health 2002 Jan;17(1):1-16. [doi: 10.1080/08870440290001494]
24. VanderZee KI, Sanderman R, Heyink JW. Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. Int J Behav Med 1996;3(2):104-122. [doi: 10.1207/s15327558ijbm0302_2] [Medline: 16250758]
25. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]

XSL•FO
RenderX

26. Rademakers J, Nijman J, Heijmans M, Rijken M. Measuring patient activation in The Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13). BMC Public Health 2012;12:577 [FREE Full text] [doi: 10.1186/1471-2458-12-577] [Medline: 22849664]
27. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 2016 Jul;3(4):430-441. [doi: 10.1177/104973239300300403]
28. Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. J Psychosom Res 1999 Dec;47(6):555-567. [Medline: 10661603]
29. Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health 1999 Jan;14(1):1-24. [doi: 10.1080/08870449908407311]
30. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
31. van Lee L, Feskens EJM, Meijboom S, Hooft van Huysduynen EJC, van't Veer P, de Vries JHM, et al. Evaluation of a screener to assess diet quality in the Netherlands. Br J Nutr 2016 Feb 14;115(3):517-526. [doi: 10.1017/S0007114515004705] [Medline: 26628073]
32. Pomerleau CS, Carton SM, Lutzke ML, Flessland KA, Pomerleau OF. Reliability of the Fagerstrom Tolerance Questionnaire and the Fagerstrom Test for Nicotine Dependence. Addict Behav 1994;19(1):33-39. [Medline: 8197891]
33. Seth P, Glenshaw M, Sabatier JHF, Adams R, Du Preez V, DeLuca N, et al. AUDIT, AUDIT-C, and AUDIT-3: drinking patterns and screening for harmful, hazardous and dependent drinking in Katutura, Namibia. PLoS One 2015;10(3):e0120850 [FREE Full text] [doi: 10.1371/journal.pone.0120850] [Medline: 25799590]
34. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatr Soc 1998 Jul;46(7):889-894. [Medline: 9670878]
35. ten Klooster PM, Oostveen JCM, Zandbelt LC, Taal E, Drossaert CHC, Harmsen EJ. Further validation of the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns 2012 Apr;87(1):125-130. [doi: 10.1016/j.pec.2011.07.017] [Medline: 21889864]
36. Hsueh I, Hsieh C. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. Clin Rehabil 2002 Sep;16(6):617-624. [doi: 10.1191/0269215502cr530oa] [Medline: 12392336]
37. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
38. Zuidema R, van Dulmen S, Nijhuis-van der Sanden M, Meek I, van den Ende C, Fransen J, et al. Efficacy of a web-based self-management enhancing program for patients with rheumatoid arthritis: explorative randomized controlled trial. J Med Internet Res 2019 Apr 30;21(4):e12463 [FREE Full text] [doi: 10.2196/12463] [Medline: 31038461]
39. Beentjes TAA, Goossens PJJ, Vermeulen H, Teerenstra S, Nijhuis-van der Sanden MWG, van Gaal BGI. E-IMR: e-health added to face-to-face delivery of Illness Management & Recovery programme for people with severe mental illness, an exploratory clustered randomized controlled trial. BMC Health Serv Res 2018 Dec 12;18(1):962 [FREE Full text] [doi: 10.1186/s12913-018-3767-5] [Medline: 30541536]
40. Leenen LAM, Wijnen BFM, Kessels AGH, Chan H, de Kinderen RJA, Evers SMAA, et al. Effectiveness of a multicomponent self-management intervention for adults with epilepsy (ZMILE study): a randomized controlled trial. Epilepsy Behav 2018 Mar;80:259-265. [doi: 10.1016/j.yebeh.2018.01.019] [Medline: 29449140]
41. van Hooft SM, Been-Dahmen JMJ, Ista E, Boeije HR. A realist review: what do nurse-led self-management interventions achieve for outpatients with a chronic condition? J Adv Nurs 2017 Jun;73(6):1255-1271. [doi: 10.1111/jan.13189] [Medline: 27754557]
42. Dwarswaard J, Bakker EJM, Boeije HR. Self-management support from the perspective of patients with a chronic condition: a thematic synthesis of qualitative studies. Health Expect 2016 Apr;19(2):194-208 [FREE Full text] [doi: 10.1111/hex.12346] [Medline: 25619975]
43. Engelen MM, van Dulmen S, Nijhuis-van der Sanden MWG, Adriaansen MJM, Vermeulen H, Bredie SJH, et al. Self-management support in cardiovascular consultations by advanced practice nurses trained in motivational interviewing: an observational study. Patient Educ Couns 2020 Jan;103(1):159-164. [doi: 10.1016/j.pec.2019.07.018] [Medline: 31362901]
44. Crutzen R, Viechtbauer W, Kotz D, Spigt M. No differential attrition was found in randomized controlled trials published in general medical journals: a meta-analysis. J Clin Epidemiol 2013 Sep;66(9):948-954. [doi: 10.1016/j.jclinepi.2013.03.019] [Medline: 23790724]
45. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
46. Höfer S, Benzer W, Schüssler G, von Steinbüchel N, Oldridge NB. Health-related quality of life in patients with coronary artery disease treated for angina: validity and reliability of German translations of two specific questionnaires. Qual Life Res 2003 Mar;12(2):199-212. [doi: 10.1023/a:1022272620947] [Medline: 12639066]

XSL•FO
RenderX

47. Dougherty CM, Dewhurst T, Nichol WP, Spertus J. Comparison of three quality of life instruments in stable angina pectoris: Seattle Angina Questionnaire, Short Form Health Survey (SF-36), and Quality of Life Index-Cardiac Version III. J Clin Epidemiol 1998 Jul;51(7):569-575. [doi: 10.1016/s0895-4356(98)00028-6] [Medline: 9674663]

Abbreviations

BCT: behavior change technique
BMQ: Beliefs about Medicines Questionnaire
CVD: cardiovascular disease
eHealth: electronic health
IM: intervention mapping
IPAQ: International Physical Activity Questionnaire
IPQ: Illness Perception Questionnaire
ITT: intention to treat
mHealth: mobile health
PAM-13: Patient Activation Measure
PPA: per-protocol analysis
RCT: randomized controlled trial
SE: patient self-efficacy with a self-constructed 26-item questionnaire

Edited by G Eysenbach; submitted 11.12.19; peer-reviewed by T Muto, T Lakiang, A Hilt, D Cespedes, R Cruz Martínez; comments to author 14.02.20; revised version received 09.04.20; accepted 12.04.20; published 24.07.20.

Please cite as:
Engelen MM, van Dulmen S, Puijk-Hekman S, Vermeulen H, Nijhuis-van der Sanden MWG, Bredie SJH, van Gaal BGI
Evaluation of a Web-Based Self-Management Program for Patients With Cardiovascular Disease: Explorative Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17422
URL: http://www.jmir.org/2020/7/e17422/
doi:10.2196/17422
PMID:32706708

©Marscha M Engelen, Sandra van Dulmen, Saskia Puijk-Hekman, Hester Vermeulen, Maria WG Nijhuis-van der Sanden, Sebastian JH Bredie, Betsie GI van Gaal. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An mHealth Intervention (ReZone) to Help Young People Self-Manage Overwhelming Feelings: Cluster-Randomized Controlled Trial

Chloe Edridge[1,2], MPhil; Miranda Wolpert[1,2], DClinPsy; Jessica Deighton[1,2], PhD; Julian Edbrooke-Childs[1,2], PhD

[1]Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[2]Anna Freud National Centre for Children and Families, London, United Kingdom

Corresponding Author:
Chloe Edridge, MPhil
Clinical, Educational and Health Psychology
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 020 7443 2218
Email: Chloe.edridge@annafreud.org

Abstract

Background: Mental health difficulties in young people are increasing, and there is a need for evidence on the effectiveness of digital interventions to increase opportunities for supporting mental health in young people. Such studies are complicated due to issues of implementation and adoption, outcome measurement, and appropriate study designs.

Objective: The objective of this study was to examine the effectiveness of an mHealth intervention (ReZone) in reducing mental health difficulties in young people.

Methods: The cluster-randomized controlled trial enrolled 409 participants aged 10-15 years, and classes were allocated to ReZone or management as usual. Self-reported questionnaires were completed at baseline and 3-month follow-up.

Results: There were no significant differences between the ReZone condition and management as usual in the self-reported outcome measures. However, there were 3467 usage sessions, which corresponds to 16.9 times per student (total of 205 students) in classes allocated to ReZone.

Conclusions: It is essential to publish studies that do not show significant differences, as these findings can still contribute to the literature, help in learning, and inform the direction of future work. The results reported in this paper could be due to a range of reasons, including whether ReZone has the scope to impact change or limitations related to the setting, context, and appropriateness of an RCT. The findings of this study suggest that ReZone was implemented and adopted.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.7019

(J Med Internet Res 2020;22(7):e14223) doi:10.2196/14223

KEYWORDS

cluster trial; behavioral difficulties; schools; mHealth; digital; mental health

Introduction

Mental health difficulties are increasing among young people. Emotional disorders are the most prevalent and show the most significant increase over time and with age [1,2]. Recent figures show that 1 in 8 (12.8%) people aged 5-19 years had at least one mental disorder when assessed [3]. Mobile health (or mHealth) interventions have the potential to provide mental health support at scale, and there is a need for evidence as to the effectiveness of digital interventions for supporting mental health in young people [4]. However, non-adoption and implementation issues present substantial barriers to effectively doing so [5]. Additionally, questions about whether randomized controlled trials are always the most appropriate method of evaluation need to be considered in the context of practicality and feasibility factors [6-8].

The authors have co-designed an mHealth intervention, ReZone, with young people, parents/carers, and teachers, based on

XSL•FO
RenderX

cognitive behavioral therapy (CBT), mindfulness, and attention bias modification training (ABMT). This short paper will report the findings from a cluster randomized controlled trial (RCT) to test the hypothesis that young people in classes receiving ReZone have fewer behavioral difficulties than young people in classes receiving management as usual. Full details on the methods and development are reported in the protocol [9] and implementation paper [10].

Methods

Participants and Procedure

A total of 49 classes were recruited (27 mainstream, 22 alternative provision), resulting in a sample of 409 students (mean age 10.9, SD 1.44 years; 56% male; Table 1). Schools were recruited via newsletters and events published in our Schools in Mind network. Participants were enrolled from 12 mainstream primary schools (307 students) and 11 secondary alternative provision schools (103 students). A total of 66 people were lost to follow-up, resulting in an attrition rate of 16%. All students (aged 10-15 years) in schools taking part in the project were eligible to participate. Classes were randomized using a true random number generator in blocks of two by an independent trials unit, to management as usual or ReZone stratified by school type (alternative provision vs. mainstream) after classes were recruited and baseline measures assessed. Ethical approval was received from University College London (UCL) Research Ethics Committee (number: 7969/001). The present research is reported in line with CONSORT guidelines [11].

Table 1. Demographic characteristics and evaluation statistics.

Demographics	Overall	Mainstream primary			Alternative provision		
		Control	Intervention	*P* value	Control	Intervention	*P* value
Age (years), mean (SD)	10.9 (1.44)	10.25 (0.43)	10.28 (0.45)	.53	13.6 (1.48)	12.3 (1.45)	<.001
Male, n (%)	227 (56)	69 (43.9)	75 (50.3)	.22	38 (80.9)	45 (80.4)	.90
Ethnicity, n (%)				.74			.90
White	247 (60)	98 (62.4)	82 (55.0)		31 (66.0)	36 (64.3)	
Mixed-race	29 (7)	N/A[a]	6 (4.0)		7 (14.9)	9 (16.1)	
Asian	78 (19)	34 (21.7)	40 (26.9)		3 (6.4)	1 (1.8)	
Black	40 (10)	11 (7.0)	15 (10.1)		5 (10.6)	9 (16.1)	
Other	15 (4)	5 (3.2)	5 (3.4)		7 (14.9)	1 (1.8)	
Disabled, n (%)	28 (3.4)	2 (1.3)	2 (1.3)	.99	6 (13.0)	4 (7.3)	.28
Alternate provision schools, n (%)	103 (25)	N/A	N/A	N/A	N/A	N/A	N/A
English first language, n (%)	352 (87)	N/A	N/A	N/A	N/A	N/A	N/A
Usage, n							
Happy faces	1169	N/A	N/A	N/A	N/A	N/A	N/A
Breathing	811	N/A	N/A	N/A	N/A	N/A	N/A
Game	752	N/A	N/A	N/A	N/A	N/A	N/A
Stress bucket	404	N/A	N/A	N/A	N/A	N/A	N/A
Art	225	N/A	N/A	N/A	N/A	N/A	N/A
Time out	106	N/A	N/A	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

Intervention: ReZone App

Full details of the ReZone intervention are described in the protocol [9]. The intervention primarily aims to reduce behavioral difficulties. It is available as a web-based smartphone and tablet application flexibly designed to be used in groups or individually, with or without adult support depending on the user's capability and confidence. This flexibility also extends to the order and frequency in which elements of the tool are used. However, it is required to have an initial guided session or tutorial with an adult. This flexible approach to implementation aimed to fit within a variety of schools and their differing student characteristics, priorities, and timetables.

The primary features of ReZone (Textbox 1) are presented below.

XSL•FO
RenderX

Textbox 1. The primary features of ReZone.

Stress bucket. The stress bucket lets the user add any stressors that they are experiencing to a bucket. They are then able to introduce activities that help them cope with each stressor. They can see the water in the bucket rise and fall as they add and relieve stressors. If the bucket reaches 50 stress points, it will overflow.

Timeout. Timeout asks the user to think through a time when they have felt stressed, angry, or upset. The user works through the events leading up to feeling this way and what happened afterward. The user can also think through what they could have done differently, thus guiding them to develop a behavioral plan. The timeout visualization is a rocket, and each thought process creates an exhaust cloud.

Chill out. Chill out uses breathing to help the user calm down and relax. Each chill out activity is based around an object or animal (ie, rabbit, jellyfish, ball, or square), using breathing in different ways.

Art therapy. The user can choose between a castle, dinosaur, fish, goat, heart, helicopter, unicorn, rocket, footballer, sea, or turtle to color in. There is a range of colors and tools to complete the drawing.

Happy faces. The user is given 30 seconds to find as many happy faces as they can, amongst other faces depicting negative emotions.

Game. In the game Balloon Blast, the user taps the screen to move a balloon up, trying to avoid all the obstacles as hitting one will mean the game is over. The game provides a break or reward for the user in between uses of the other features.

Measures

All measures were self-reported by young people and administered in school by researchers or school staff in paper form. Baseline demographic data were collected and are reported in Table 1. Emotional (α=.78) and behavioral difficulties (α=.81) were measured using the 16-item Me and My School (M&MS) questionnaire [12]. Mental wellbeing was measured using the 7-item Short Warwick-Edinburgh Mental Wellbeing Scale (α=.83) [13]. Empowerment was measured using the 6-item "self" subscale of the Youth Empowerment Scale-Mental Health (α=.73) [14]. Health-related quality of life was measured using the 6-item (with global health visual analog scale) EQ-5D-Y (α=.63) [15].

Intervention Usage

Intervention usage data were collected in ReZone and exported to the data analysis software. There were 3467 usage sessions, which corresponds to 16.9 times per each of the 205 students in classes allocated to ReZone. The game (n=752), breathing exercises (n=811), and happy faces (n=1169) were used the most; stress bucket (n=404), art (n=225), and time out (n=106) were used the least. Based on teacher consultations, which are discussed elsewhere [10], usage methods and frequency varied across schools and students due to a variety of factors, including student characteristics, teacher priorities, and school environment.

Analytic Strategy

Means and standard deviations, for all measures, across conditions at baseline and follow-up, were calculated. Intention-to-treat and per-protocol analyses with four multilevel regressions were tested for each of the outcome variables with time nested within students within classrooms. Per-protocol analyses included only those that had completed follow-up. Model 0 was computed as the null model to examine the change in outcome over time, without predictors. The intraclass correlation coefficients for each measure were 17% (emotional difficulties), 24% (behavioral difficulties), 7% (empowerment), 9% (wellbeing), and 3% (quality of life). We then entered school type (ie, alternative provision vs mainstream) as a fixed effect predictor to Model 1. In Model 2, patient-level grand mean-centered age, ethnicity, gender, and disability were entered as predictors. Demographics were controlled for in the analyses as there were significant differences in gender between groups at baseline within mainstream primary schools and age between groups within alternative provision schools (Table 1). In the Final Model (Model 3), condition (ReZone vs management as usual) was entered as a predictor. The likelihood ratio test was used to compare the fit of Models 1, 2, and 3. For example, Model 1 was significant for behavioral difficulties and therefore compared to Model 2, which was significant. Models 1 and 2 were therefore retained in the final model, with Model 3 (final model) being compared to Model 2 using the likelihood ratio test. Models 1 and 2 were significant for empowerment and wellbeing and therefore retained in the final model. For emotional difficulties and quality of life, only Model 2 was significant and retained in the final Model.

Results

The means and standard deviations of all study measures at baseline and postintervention are presented in Table 2. The likelihood ratio test for the intention to treat analysis was not significant for any of the final models compared to the previous significant model. There were no significant differences between the ReZone condition and management as usual in emotional difficulties (χ^2_1=0, *P*>.05), behavioural difficulties (χ^2_1=0.01, *P*>.05), empowerment (χ^2_1=0.01, *P*>.05), wellbeing (χ^2_1=0.05, *P*>.05), or quality of life (χ^2_1=0.3154, *P*>.05). The per-protocol results showed similar patterns.

XSL•FO
RenderX

Table 2. Baseline and follow-up mean scores by condition. All values are provided as mean (SD).

Outcome	Mainstream primary				Alternative provision			
	Intervention – T1[a]	Intervention – T2[b]	Control – T1	Control – T2	Intervention – T1	Intervention – T2	Control – T1	Control – T2
Emotional difficulties	5.56 (3.38)	5.18 (3.74)	6.27 (3.38)	5.43 (3.27)	6.08 (3.91)	5.61 (4.12)	5.48 (3.77)	6.08 (4.03)
Behavioral difficulties	3.31 (2.27)	3.28 (2.44)	3.32 (2.24)	3.18 (2.04)	5.45 (2.37)	5.11 (2.34)	5.57 (2.35)	5.82 (2.52)
Empowerment	24.53 (4.57)	25.69 (4.25)	24.65 (4.09)	25.13 (4.03)	22.96 (4.96)	23.61 (5.74)	23.35 (4.02)	23.79 (4.25)
Wellbeing	24.11 (4.78)	24.94 (5.41)	24.41 (5.18)	24.94 (5.06)	21.09 (5.90)	23.61 (5.74)	21.76 (3.98)	21.93 (4.65)
Quality of life	6.13 (1.60)	6.46 (5.86)	6.28 (1.48)	6.29 (6.06)	6.20 (1.24)	4.04 (3.08)	6.30 (1.32)	2.91 (3.14)

[a]T1: Time 1.

[b]T2: Time 2.

Discussion

This cluster randomized controlled trial examined the effectiveness of ReZone in reducing behavioral and mental health difficulties in young people. The study did not show a significant difference between those using ReZone and management as usual in the reduction of behavioral and emotional difficulties and improving empowerment, wellbeing, and quality of life.

Potential limitations of the study, implementation, and adoption barriers to digital interventions within schools have been discussed in depth [10]. Implementation barriers to new digital interventions in schools include the technology itself, the adopter system, and the organization [5]. The process of embedding ReZone into the school system can sometimes leave minimal time to utilize it successfully. Implementation strategies differed, with varying levels of teacher encouragement and consistency of scheduled ReZone use. Competing teacher priorities, opinions, and seniority of the school advocate could have impacted the time and drive to use the tool [10].

ReZone was used a total of 3467 times throughout the 3-month study period, and current usage data show that ReZone is still being used 10 months poststudy (501 sessions over the last 3 months). Aggregated objective usage data is useful to indicate overall acceptability and use. However, teacher consultations (fully reported in [10]) suggested students either used ReZone frequently or not at all. Analysis of a dose-response effect could be improved by gathering further quantifiable individual school usage information related to implementation.

The usage statistics and teacher consultations indicated happy faces and breathing were the most used activities, while the stress bucket, art, and time out were used the least. Happy faces and breathing can be completed in only 1-2 minutes; they are a fun challenge/game type of activity, and most of the time can be completed without teacher assistance. The stress bucket and time out require more time to work through thought processes, requiring the user to be willing to think and talk about their feelings, needing teacher support (which is not always feasible).

The challenges discussed link to the emerging concept of gamification within digital mental health tools [16,17]. There are some basic elements of gamification within the most popular elements of ReZone. Further work on gamification combined with other crucial processes such as co-design, could help address the broader issue of nonadoption and sustainability, with many users reporting that they stop the use of a health app after only two weeks [18]. This lack of consistency could be particularly applicable when working with young people in such settings as alternative provision schools. Pupil behavior often places higher demands on staff in alternative provision schools than in mainstream schools, and disengagement and loss of concentration are common issues.

Measures used were related to overall mental health, wellbeing, and behavior leading us to question how often a tool such as ReZone would need to be used to effect changes in broad domains and specifically in schools. If tools such as ReZone do not provide large-scale, substantial longer-term changes, we must also ask whether there is any impact on immediate feelings, behavior, and interactions. We did attempt to measure this by asking teachers to complete a few short questions related to how well students refocused and re-engaged after using ReZone. However, due to practical issues such as teacher demand, these were not completed. On reflection, this could have also been something students were asked to complete. It is also possible that effects were delayed, and further long-term data collection, such as one-year post-trial, may have shown differences [19-21].

Due to the complex nature of digital interventions, traditional RCTs may be insufficient to assess efficacy [8]. As this study shows, there are multiple objective and subjective data sources available, including quantitative self-report data, objective usage data, and subjective qualitative data, while digital and school studies present particular challenges in adherence, sustainability, and roll-out [10]. As with psychotherapy research in general and a reliance on self-report measures, participant blinding was not possible, although true and independent randomization was used. Additionally, due to this being a pragmatic trial, there may have been other interventions the school and participants may have been using.

It is essential to publish negative results to avoid publication bias. These studies can still make a substantial contribution to the literature, learning, and informing the direction of future work [22]. The negative results of this study could indicate that ReZone has a limited scope to impact change, that its utility is hampered by issues with implementation and nonadoption, and limitations of mHealth research in a school context. The results

XSL•FO
RenderX

may also suggest the limited applicability of the RCT study design for evaluating tools such as ReZone.

Acknowledgments

The research was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care North Thames at Bart's Health NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health.

Conflicts of Interest

None declared.

References

1. Deighton J, Lereya ST, Casey P, Patalay P, Humphrey N, Wolpert M. Prevalence of mental health problems in schools: poverty and other risk factors among 28,000 adolescents in England. Br J Psychiatry 2019 Jan 30:1-3. [doi: 10.1192/bjp.2019.19] [Medline: 30698513]
2. Fink E, Patalay P, Sharpe H, Holley S, Deighton J, Wolpert M. Mental Health Difficulties in Early Adolescence: A Comparison of Two Cross-Sectional Studies in England From 2009 to 2014. J Adolesc Health 2015 May;56(5):502-507. [doi: 10.1016/j.jadohealth.2015.01.023] [Medline: 25907650]
3. Mental Health of Children Young People in England, 2017. London: NHS Digital; 2018. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 [accessed 2020-06-19]
4. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
5. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
6. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Am J Public Health 2004 Mar;94(3):416-422. [doi: 10.2105/ajph.94.3.416] [Medline: 14998805]
7. Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial. Journal of Clinical Epidemiology 2001 Jun;54(6):541-549. [doi: 10.1016/s0895-4356(00)00347-4]
8. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
9. Edbrooke-Childs J, Smith J, Rees J, Edridge C, Calderon A, Saunders F, et al. An App to Help Young People Self-Manage When Feeling Overwhelmed (ReZone): Protocol of a Cluster Randomized Controlled Trial. JMIR Res Protoc 2017 Nov 03;6(11):e213. [doi: 10.2196/resprot.7019]
10. Edridge C, Deighton J, Wolpert M, Edbrooke-Childs J. The Implementation of an mHealth Intervention (ReZone) for the Self-Management of Overwhelming Feelings Among Young People. JMIR Form Res 2019 May 02;3(2):e11958 [FREE Full text] [doi: 10.2196/11958] [Medline: 31045499]
11. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8(1):18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
12. Patalay P, Deighton J, Fonagy P, Vostanis P, Wolpert M. Clinical validity of the Me and My School questionnaire: a self-report mental health measure for children and adolescents. Child Adolesc Psychiatry Ment Health 2014;8(1):17 [FREE Full text] [doi: 10.1186/1753-2000-8-17] [Medline: 24963344]
13. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7(1):15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
14. Walker JS, Thorne EK, Powers LE, Gaonkar R. Development of a Scale to Measure the Empowerment of Youth Consumers of Mental Health Services. Journal of Emotional and Behavioral Disorders 2009 Jun 05;18(1):51-59. [doi: 10.1177/1063426609337388]
15. EuroQol Group. EQ-5D-Y user guide: basic information on how to use the EQ-5D-Y instrument. 2014. URL: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-Y_User_Guide_v1.0_2014.pdf [accessed 2017-01-01]

XSL•FO
RenderX

16. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
17. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using Mobile Health Gamification to Facilitate Cognitive Behavioral Therapy Skills Practice in Child Anxiety Treatment: Open Clinical Trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
18. Dorsey ER, Yvonne Chan Y, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. Academic Medicine 2017;92(2):157-160. [doi: 10.1097/acm.0000000000001205]
19. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. The Lancet 2018 Dec;392(10163):2452-2464. [doi: 10.1016/s0140-6736(18)31782-3]
20. Foshee V, Bauman K, Greene W, Koch G, Linder G, MacDougall J. The Safe Dates program: 1-year follow-up results. Am J Public Health 2000 Oct;90(10):1619-1622. [doi: 10.2105/ajph.90.10.1619] [Medline: 11029999]
21. Foshee VA, Bauman KE, Ennett ST, Linder GF, Benefield T, Suchindran C. Assessing the long-term effects of the Safe Dates program and a booster in preventing and reducing adolescent dating violence victimization and perpetration. Am J Public Health 2004 Apr;94(4):619-624. [doi: 10.2105/ajph.94.4.619] [Medline: 15054015]
22. Franco A, Malhotra N, Simonovits G. Social science. Publication bias in the social sciences: unlocking the file drawer. Science 2014 Sep 19;345(6203):1502-1505 [FREE Full text] [doi: 10.1126/science.1255484] [Medline: 25170047]

Abbreviations

ABMT: attention bias modification training
CBT: cognitive behavioral therapy
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 08.04.19; peer-reviewed by R Meiksin, B Yanez; comments to author 06.07.19; revised version received 31.10.19; accepted 23.03.20; published 27.07.20.

Please cite as:
Edridge C, Wolpert M, Deighton J, Edbrooke-Childs J
An mHealth Intervention (ReZone) to Help Young People Self-Manage Overwhelming Feelings: Cluster-Randomized Controlled Trial
J Med Internet Res 2020;22(7):e14223
URL: http://www.jmir.org/2020/7/e14223/
doi:10.2196/14223
PMID:32716299

©Chloe Edridge, Miranda Wolpert, Jessica Deighton, Julian Edbrooke-Childs. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of an Electronic Health Service on Child Participation in Pediatric Oncology Care: Quasiexperimental Study

Britt-Mari Gilljam[1], DPhil; Jens M Nygren[1], Prof Dr; Petra Svedberg[1], Prof Dr; Susann Arvidsson[1], DPhil

School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Britt-Mari Gilljam, DPhil
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 301 18
Sweden
Phone: 46 35 16 71 00
Email: Britt-mari.gilljam@hh.se

Abstract

Background: For children 6-12 years old, there is a shortage of electronic Health (eHealth) services that promote their participation in health care. Therefore, a digital communication tool, called Sisom, was developed to give children a voice in their health care. Children with long-term diseases want to be more involved in their health care and have the right to receive information, be listened to, express their opinions, and participate in decision making in health care. However, the outcomes of using Sisom in practice at pediatric oncology clinics have not been investigated.

Objective: The aim of this study was to investigate children's participation during appointments with pediatricians at pediatric oncology clinics, with or without the use of the eHealth service Sisom.

Methods: A quasiexperimental design with mixed methods was used. We analyzed 27 filmed appointments with pediatricians for 14 children (8 girls and 6 boys) aged 6-12 years (mean 8.3 years) with a cancer diagnosis. The intervention group consisted of children who used Sisom prior to their appointments with pediatricians at a pediatric oncology clinic, and the control group consisted of children who had appointments with pediatricians at 4 pediatric oncology clinics. Data from observations from the videos were quantitatively and qualitatively analyzed. The quantitative analysis included manual calculations of how many times the pediatricians spoke directly to the children, the proportion of the appointment time that the children were talking, and levels of participation by the children. For the qualitative analysis, we used directed content analysis to analyze the children's levels of participation guided by a framework based on Shier's model of participation.

Results: Pediatricians directed a greater proportion of their discussion toward the child in the intervention group (731 occasions) than in the control group (624 occasions), but the proportion of the appointment time the children talked was almost the same for both the intervention and control groups (mean 17.0 minutes vs 17.6 minutes). The levels of participation corresponded to the first three levels of Shier's participation model: children were listened to, children were supported to express their views, and children's views were taken into account. The results showed an increased level of participation by the children in the intervention group. Several codes that were found did not fit into any of the existing categories, and a new category was thus formed: children received information.

Conclusions: This study shows that the eHealth service Sisom can increase children's participation during appointments with health care professionals. Further studies employing a randomized control design focusing on the effects of eHealth services on children's health outcomes, perceived participation, and cost-effectiveness could make a significant contribution to guiding the implementation of eHealth services in pediatric care.

(J Med Internet Res 2020;22(7):e17673) doi:10.2196/17673

KEYWORDS

cancer; child care; communication; eHealth; patient participation

XSL•FO
RenderX

Introduction

Background

Advancements in electronic Health (eHealth) services to promote health, participation, and communication have been made in recent years [1-4]. These are in the form of mobile health technology (mobile devices) [5], and the majority are aimed at adults and adolescents [6,7], while the range of eHealth advancements for younger children is limited [8]. Most eHealth services for adolescents do not focus on participation, are not validated, and lack proof of effectiveness [6,9-11]. Existing eHealth services instead primarily focus on self-management [6,10-12], symptom assessment [9,12,13], social support, support for self-care [12,13], or medication adherence [11,13,14]. Only one study on an eHealth service promoting participation for young children, 3-5 years old, has been performed [15], while there are no studies that describe an eHealth service that promotes participation for children 6 years old and older [16,17].

Children's Participation in Health Care

Children with long-term diseases such as cancer want to be have greater involvement in their own health care [18-21].They want to receive more information about emerging symptoms, treatment, and prognosis [22] and have the possibility to decide whether and to what extent they want to participate in decision-making processes regarding their own health care [18,23]. It is, however, difficult for children to explain how their long-term disease makes them feel [24] and to accept and manage the consequences and lifestyle related to their long-term disease [17]. It is thus essential to include children in their own health care and support them in their self-management responsibility in order to improve their wellbeing and reduce fear and anxiety [25] as well as to strengthen their self-confidence and independence [17,18,26]. Even if children's participation is considered an essential part in health care, there is still a number of barriers to overcome in relation to health care professionals' attitudes and viewpoints [27]. Furthermore, it is emphasized in the United Nations Convention on the Rights of the Child [28] that children have the right to be involved in all matters affecting them. Despite the knowledge that children want to be more involved in their health care, that their participation has positive effects on treatment and health outcomes, and that laws and regulations emphasize the requirement of patient participation in health care, there is still a lack of knowledge, strategies, and methods for strengthening children's participation in pediatric health care and how to accomplish their participation [29,30].

eHealth Services: Example of Sisom

Sisom is an eHealth service helping children to communicate in health care by engaging them in a playful virtual world [31,32]. Sisom was developed with a participatory design that included children with cancer in the process [31]. Previous research conducted on Sisom has primarilyfocused on usability and content validity, resulting in different diagnosis-specific versions and country-specific versions: Norway, Sweden, United States, and Canada [33-36]. In Sisom, the children travel by boat in an archipelago of islands with different subjects where they can express their feelings and symptoms by answering a set of 82 questions (Multimedia Appendix 1) [33]. The 5 islands represent different topics for the questions: (1) to handle things, (2) my body, (3) thoughts and feelings, (4) things you may be afraid of, and (5) in the hospital (Figure 1). The intent is that the children answer the questions in Sisom prior to a health care appointment or when they are hospitalized. A report of the children's responses can be printed. Based on the children's answers, the health care professionals can talk to the children about how they perceive their situation and how they think about different health care–related and social issues. It has been shown that communication between children, their parents, and health care professionals is improved when using Sisom [37,38].

It has been highlighted that participatory approaches where patients have a real influence on the development of eHealth interventions lead to a more user-friendly and effective intervention [39]. This is especially important for interventions aimed at children, since adults lack a child's perspective. However, children's participation in the development of eHealth services is not common [40]. The use of eHealth services to enable and support children's participation in pediatric health care is scarce. We found only one review of eHealth services that was designed to support communication between children with cancer and health care professionals with the purpose of strengthening children's participation in care [12]. The implementation of eHealth services for children with chronic conditions is primarily focused on changing health behavior [7,14,41,42] and medical adherence [14], and most of the eHealth interventions that have been implemented in pediatric oncology clinics target emotional distress, health behaviors, health outcomes, and neurocognitive functioning [7], with insufficient evidence of effects. Furthermore, there is a need for more rigorous research and evaluations to determine eHealth service efficacy [6,7,14,42,43] especially in relation to eHealth services that are directed towards strengthening children's participation in care. There is a lack of evidence to support that eHealth services actually promote children's participation in their health care. It is also unclear to what extent children's participation is supported and promoted in practice in pediatric oncology clinics. Thus, the aim of this study was to investigate children's participation during appointments with pediatricians at pediatric oncology clinics, with or without the use of the eHealth service Sisom.

XSL•FO
RenderX

Figure 1. The 5 islands of Sisom.

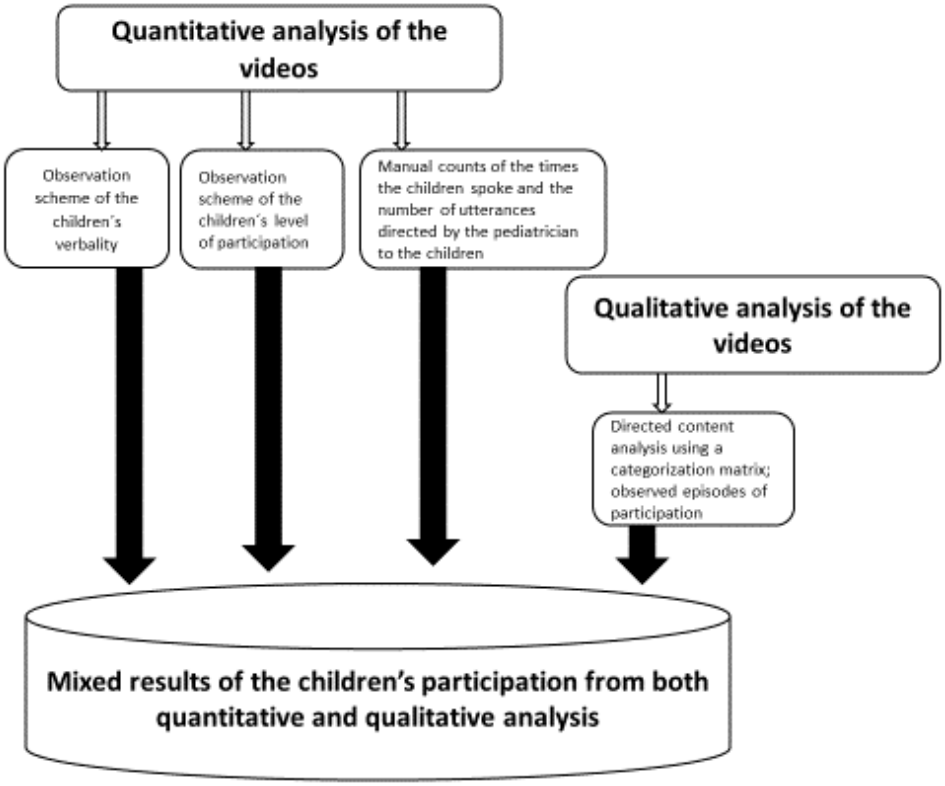


Methods

Design

A quasiexperimental design [44] with a convergent parallel mixed method [45,46] was used, where the frequencies and differences between groups were calculated with descriptive statistics and qualitative content analysis was used to determine how children's participation was manifested [47] (Figure 2).

Figure 2. Processes for quantitative and qualitative analysis.



XSL•FO
RenderX

Participants

Inclusion criteria for participation were children 6-12 years old undergoing cancer treatment at 4 pediatric oncology clinics in Norway. Exclusion criteria were children who did not speak Norwegian, diagnosis of brain tumors, mental disability, developmental delay, speech disorders, or hearing impairments. The demographic characteristics are provided in Table 1.

Table 1. Demographic characteristics of the participants.

Characteristics	Intervention group, n (%)	Control group, n (%)
Age (years)		
6-8	2 (33)	5 (63)
9-12	4 (67)	3 (37)
Gender		
Girl	2 (67)	6 (75)
Boy	4 (33)	2 (25)
Diagnosis		
Acute lymphoblastic leukemia (ALL)	3 (50)	3 (38)
Lymphoma	1 (17)	2 (25)
Carcinoma	1 (17)	0 (0)
Not defined	1 (16)	3 (37)

Recruitment

The children were successively recruited after receiving a cancer diagnosis and began the treatment. The head nurse identified eligible children and asked them and their parents if they were willing to be contacted about the study. The project assistant then informed the child and the parents, both verbally and in writing, about the content of the study including the voluntary nature of participation and the possibility to withdraw later in the process and asked if they were willing to participate. Families interested in participating signed an informed consent form. Of the 20 children who were asked to participate, 15 chose to opt-in. One child, who had disease-related complications, withdrew, resulting in 14 participants aged 6-12 years, of whom 6 were boys and 8 were girls.

Data Collection

Data were collected through video recordings of the appointments with the 15 pediatricians, who conducted 1-5 recorded appointments each, for a total of 27 video-recorded appointments in outpatient examination rooms. Although an additional 3 pediatrician appointments were planned to be filmed with each child, this was not possible for all the children due to medical complications or relocations to another hospital. All the control group appointments took place prior to Sisom being introduced at the hospitals. The group of children in the intervention then met the pediatricians after Sisom was introduced. The appointments, both for the children in the control group and children in the intervention group, were in the form of routine meetings, which started with a nurse taking standard measurements of length, weight, and pulse oximetry. The pediatrician started by reviewing the children's medication, reviewing infection status, and explaining the results of blood tests, supplemented by a medical examination. All those participating in the appointments were aware that they would be filmed. The children in the intervention group used Sisom while they were in the waiting room prior to meeting the pediatrician, and the project assistant was available for help. A report of the children's responses was then printed and given to the pediatrician as well as to the parent and child; then, the meeting with the pediatrician took place. It was only during one appointment that the pediatrician revealed (to the researcher analyzing the recording) that Sisom had been used using the words: "I saw in your schedule that one of your feet is hurting; is it?" The word Sisom was not mentioned during any appointment.

Data Analysis

The analysis was performed using a convergent parallel mixed method [46] to combine qualitative and quantitative data to be able to attain a more complete understanding [48] of the children's participation during the appointments with pediatricians, with or without using Sisom. The quantitative and qualitative data were collected concurrently, and the results were combined and compared for convergence and differences [46]. All the video recordings were viewed several times with the aim of trying to gain an overall picture and see and hear everything that occurred during the appointments. The analysis process was carried out by two researchers who had no knowledge about which recordings were from the appointments during which Sisom had been used. The data acquisition consisted of manual counting and observations in accordance with an observation scheme and a categorization matrix (Table 2).

XSL•FO
RenderX

Table 2. Summary of the research questions, data collection methods, and data analysis.

Research questions	How many times did the pediatricians speak directly to the children?	How long did the children speak?	Which levels of participation were achieved during the appointments?	Which levels of verbality were achieved in the appointments?	How did the children's participation manifest itself?
Data collection	Manual counting of the number of times the pediatricians spoke to the children.	Manual counting of the amount of times the children spoke.	Observations in accordance with two observation schemes with 4-degree scales were developed, and the children's participation was ranked.	Observations in accordance with two observation schemes with 4-degree scales were developed, and the children's level of verbality was ranked.	Observations according to a categorization matrix were developed and used for the collection of brief episodes of participation in the films.
Data analysis	Quantitative analysis	Quantitative analysis	Ranking	Ranking	Qualitative analysis

Quantitative Analysis

Frequencies and differences were calculated using quantitative analysis. The number of times the pediatrician spoke directly to the child was counted manually and recounted repeatedly. Frequency calculations were performed to calculate the percent values of the proportion of the total appointment time that the child was talking, the proportion of the time the child spoke to the pediatrician, and the proportion of the total number of times the pediatrician spoke that was addressed to the child [49]. The number of times the children spoke and the number of times the pediatrician spoke to the children were calculated. Descriptive statistics of the results were compiled and divided into the intervention group and control group, after the blinding was revealed.

An observation scheme was used to assess the children's verbality [50]. The level of the children's verbality during the appointments was ranked with a 4-degree verbality scale: (1) verbally inactive, gave <3 answers during the appointment and did not comment on anything or ask any questions; (2) limited verbality, gave >3 answers but no comments or questions; (3) moderate verbality, answered questions with ≥4 words but made <3 comments or questions; 4) verbally active, responded extensively to >3 questions, commenting or asking questions >3 times [50].

An observation scheme was developed by the authors to assess the children's level of participation. The observation scheme was based on the levels in Shier's participation model "Pathways to participation," a theoretical model intended to be used in practice when working with children up to 12 years old. The model consists of 5 levels: (1) children are listened to, (2) children are supported in expressing their views, (3) children's views are taken into account, (4) children are involved in decision-making processes, and (5) children share power and responsibility over decision making [51]. The 3 lowest levels present levels of participation that are the limit of what is stated in the children's Convention on the Rights of the Child as the meaning of participation. The 2 highest levels thus represent higher participation and are also consistent with the definition of shared decision making. The aim of the observation scheme was to rank the levels of the children's participation during the pediatrician appointments, in relation to Shier's 5 levels of participation. We defined in the observation scheme that the decision making should consider decisions concerning medication, caring, or procedure. To make the grading clear and understandable at each level, a 4-degree scale was used for each level in the observation scheme: (0) on a low level or not at all, (1) on a fairly low level, (2) on a high level, and (3) on a very high level (Multimedia Appendix 2). For example, at the first level of participation (children were listened to), "on a low level (0)" was assessed if the health care professionals only spoke to and listened to the parents during the conversation and if the health care professionals only took the initiative to ask the child on a few occasions or not at all. "On a fairly low level (1)" was assessed if the health care professionals spoke to and mostly listened to the parents and took the initiative to ask the child on some occasions but did not listen to what the child had to say. "On a high level (2)" was assessed if health care professionals spoke to and mostly listened to the child and took the initiative to ask the child something on several occasions but did not listen clearly to what the child was saying. "On a very high level (3)" was assessed if the health care professionals spoke to and mostly listened to the child and took the initiative to ask the child something on several occasions and listened clearly to what the child was saying. The grading for all levels is described in detail in Multimedia Appendix 2.

Face validity and content validity were assessed through the training of coders [52] to test the observation scheme and its reliability; two of the researchers analyzed the same video separately. Then all the researchers evaluated and made adjustments to the observation scheme. Interrater reliability was assessed in the next step [44], where a number of video recordings were examined by two of the researchers, who then compared their individual documentation.

Qualitative Analysis

In order to attain a greater understanding of the results, a directed, deductive qualitative content analysis was performed to discern how the children's participation manifested itself during the appointments [47]. A categorization matrix was developed from Shier's participation model [51], with which the data from the videos were reviewed for content, passages of participation, and brief episodes of participation [47]. The analysis process was performed in the following steps: (1) looking at the videos and highlighting all episodes that, on first impression, appeared to represent passages of participation; (2) coding all highlighted passages using the predetermined categories in Shier's model; and (3) giving a new category label that captures the essence of participation to any codes that could not be categorized with the predetermined categories [47]. Brief episodes in the films were found during the analysis of the video recordings where the children were actively excluded from participation. These film episodes were also included in the

XSL•FO
RenderX

results as negative codes in order to more clearly describe the degree of the children's participation. The blinding was released when the categorization process was completed, and the codes in each category were divided into the intervention group and control group. Two researchers completed the analysis, and to enhance the credibility and the dependability of the analysis, the data analysis was discussed continuously with all authors.

Ethical Considerations

The videos were collected between 2006 and 2009 by the Center for Patient Involvement and Collaboration Research, a research group at Oslo University Hospital. This study was approved by the Regional Ethical Review Board in Oslo (2014/1599), with approval for use of the collected videos. The research group at Oslo University Hospital offered the data to the first author to use for this study. The difficulties in doing research on already collected data and to have to depend on the specific setting and design of the data collection were discussed by the authors as a limitation of the study. However, not analyzing already collected data could be considered unethical, which was considered in the decision to conduct this study. The participating children had approved their participation and expected that their contribution to the research would improve care for children with cancer. The funders Region Halland and the Center for Research on Welfare, Health, and Sport at Halmstad University had no access to the study data and have not influenced the way in which the data were analyzed.

Results

Sample

In the intervention group, 13 appointments with pediatricians were performed at 1 pediatric oncology clinic, with 2 girls and 4 boys participating (mean age 9.5 years). In the control group, 14 appointments with pediatricians were performed at 4 pediatric oncology clinics, with 6 girls and 2 boys (mean age 8.1 years) participating who received care as usual.

Quantitative Results

Verbality

The appointments lasted from 6 to 41 minutes. The average times that the children and the parents met the pediatricians were 17.8 minutes for the intervention group (out of 231 minutes and 13 appointments) and 17.6 minutes for the control group (out of 246 minutes and 14 appointments; Multimedia Appendix 3). The children spoke to the pediatricians, on average, 1.14 minutes in the intervention group and 1.78 minutes in the control group (Multimedia Appendix 3). Except for answers to specific questions, the children responded by saying "yes," "no," or "hmmm" in most of the appointments. Short and unclear expressions were difficult to record and thus not included. The pediatricians in the intervention group spoke directly to the children on more occasions and for longer total time (average 56.23 minutes from 731 occasions) compared with the control group (average 44.57 minutes from 624 occasions; Multimedia Appendix 4). The children´s verbality was slightly higher in the control group than in the intervention group (Multimedia Appendix 4).

Participation

There was a higher score for the children's participation, based on Shier's participation levels [51], when using Sisom compared with the control group. The higher scores are particularly noticeable for the following 3 levels: children were listened to, children were supported in expressing their views, and children's views were taken into account (Table 3).

XSL•FO
RenderX

Table 3. The ranges (0-3) of the levels of children's participation during the appointments with the pediatricians, according to Shier's model.

Group and code number of the child	5 levels in Shier's model of participation[a]				
	Children were listened to	Children were supported into expressing their views	Children's views were taken into account	Children were involved in decision-making processes	Children shared power and responsibility over decision making
Intervention group					
I.1	3	3	0	0	0
I.1	3	3	0	0	0
I.1	3	3	0	0	0
I.2	3	3	0	0	0
I.3	3	3	0	0	0
I.3	2	3	0	0	0
I.3	3	3	3	0	0
I.4	3	3	3	3	0
I.4	3	3	3	0	0
I.4	3	3	3	3	0
I.5	3	3	3	0	0
I.6	3	3	3	0	0
I.6	3	3	3	3	0
Control group					
C.1	3	3	0	0	0
C.1	3	1	0	0	0
C.2	0	0	0	0	0
C.3	0	0	0	0	0
C.3	0	0	0	0	0
C.4	0	0	0	0	0
C.5	1	0	0	0	0
C.6	1	1	1	0	0
C.7	3	3	3	0	0
C.7	1	0	0	0	0
C.7	3	3	3	0	0
C.8	1	0	0	0	0
C.8	3	3	3	0	0
C.8	3	3	3	0	0

[a]To grade the level of children's participation, we used a 4-degree scale: (0) on a low level or not at all, (1) on a fairly low level, (2) on a high level, and (3) on a very high level.

Qualitative Results

The findings from the directed qualitative content analysis showed passages and episodes of participation that were in line with the first 3 levels in the model presented by Shier: (1) children were listened to, (2) children were supported in expressing their views, and (3) children's views were taken into account. However, the analysis of the videos did not reveal any passages or episodes that could be included in the fourth or fifth level of participation. One new category, which was not included in Shier's model, named "children received information" emerged in the analysis (Table 4).

XSL•FO
RenderX

Table 4. Analysis of the videos: examples and number of codes in each group.

Group	New category	5 levels in Shier's model of participation				
	Children received information	Children were listened to	Children were supported in expressing their views	Children's views were taken into account	Children were involved in decision- making processes	Children shared power and responsibility for decision making
Intervention group						
Examples	The pediatrician informed about what would happen. The pediatrician informed the child about a medicine.	The pediatrician gave the child time and waited for an answer. The pediatrician showed through body language and position that the child was in focus. The parent answered on behalf of the child.	The pediatrician asked the child a medical question in order to get the child's perspective. The pediatrician asked if the child wanted to talk about something special.	The pediatrician followed the child's wishes about the medical examination.	The child was involved in a decision about a blood test.	None
Number of positive codes	84	21	100	2	1	0
Number of negative codes	0	4	4	0	0	0
Control group						
Examples	The pediatrician informed the child about the medical examination. Information was only directed to a parent.	The pediatrician talked over the head of the child. Children asked and said something without receiving attention from health care professionals.	The child expressed a desire on a medical aspect. The pediatrician asked the child about symptoms.	The physician followed the child's wishes about the medical examination.	None	None
Number of positive codes	36	5	32	3	0	0
Number of negative codes	6	34	28	0	0	0

Children Received Information

Receiving information meant that the children were provided with medical information about what the pediatricians were doing and what could be expected to happen. The children also received information in response to their direct questions about, for example, test results, whether they could eat before a specific examination, or whether a peripheral venous catheter was required before an operation. In the intervention group, the pediatricians gave the children information about examinations and medication during the appointments. The disease-related information was often repeated several times during the same appointment, and almost all the children were informed by the pediatrician about the medical examination, what the pediatrician would do, and the result of the examination. The children were also often informed about the disease, symptoms, and side effects. Some children received health advice, such as the importance of eating so that the body gets nutrients. No child from the intervention group asked for further information during the appointments, but in the control group, the children asked for information on a number of occasions, such as "Why did you press there?" (when the pediatrician palpated his stomach; boy, 6 years old, control group).

The information the children in the control group asked for during the appointments was whether the child had to spend Christmas at the hospital or about examinations, tests, their anatomy, and things they did not understand. One exception to this was one child in the control group who was informed about the course of the disease and the future. The explanation during this appointment was given slowly and in a language adapted to the child's level of development. During another appointment in the control group, the pediatrician had two accompanying medical colleagues. The conversation during this appointment was conducted in English and was thus not understandable to the Norwegian-speaking child and was only directed to the medical colleagues. The pediatrician informed his colleagues about what he did in English during the medical examination but said nothing to the child. The pediatricians informed the children about the results of the medical examination and the medicines and their side effects in only a few of the appointments in the control group, and the information was aimed only at the parents.

Children Were Listened To

The children were listened to when the pediatrician paid full attention to the child and gave the child time to think before answering. The pediatricians often turned to the children during the conversation in the appointments in the intervention group. The pediatricians' body language, sitting close to the child and looking at the child, showed that they were really listening. The

XSL•FO
RenderX

pediatricians sat either squatting or near the child and thus had better eye contact.

The pediatricians waited for and gave the child time to think before answering when talking with the children during the appointments in the intervention group. The pediatricians listened to the children's answers and asked further questions or confirmed that they were listening with an "Hmm." The child's response was followed up by the pediatrician in the cases when parents and children responded differently to the same question. One physician moved closer to the child to get better eye contact and repeated the questions after the parent repeatedly responded on behalf of the child. For example, a pediatrician asked "Doesn't the food taste good anymore?" When there was no answer from the child, the pediatrician squatted twice in order to get better eye contact with the boy, pats him on the blanket, and repeats the question (boy, 7 years old, intervention group).

The children were not listened to when the pediatrician spoke quickly without giving the child time to reflect and answer or when the pediatrician used language that was difficult for a child to understand. These types of communication occurred during many appointments in the control group, although there were also appointments when the pediatrician listened to the child. The pediatricians were more often focused on the parents than on the child during the appointments in the control group. The pediatricians looked quickly at the parent and seemed to expect the parent to respond even though the question was initially addressed to the child. The conversation between the child and pediatrician was performed through a parent, leaving the child ignored during some of the appointments in the control group. On one occasion, the child was playing with the blood pressure equipment and said he would like to take his blood pressure, but none of those present, including the pediatrician, nurse, and parent, responded or reacted. On one occasion, when another pediatrician asked a child what medications the child took, the latter began to respond, but the pediatrician did not listen to the child and turned to the mother instead. The child got upset and went and put on music at a high volume. During other appointments, the pediatrician asked the parents how the child was feeling, even though the child was sitting next to them. One pediatrician even expressed his desire for the child to be quiet. Several children during the appointments in the control group said or asked things without anyone listening and responding. On one occasion, a child started talking to her mother, and the pediatrician drowned the child by telling her mother a story about his private holiday trip. This child protested loudly without any reaction from the pediatrician or support from the mother.

Children Were Supported in Expressing Their Views

The children were supported in expressing their views, which entailed the child being asked and invited into the conversation, but also that the children themselves initiated discussions about for example, how the child wanted the procedure or process to be conducted during a medical examination or that the child wanted to return home. The children often appeared to be tired and silent in both the intervention and control groups; thus, support from the pediatrician was needed. The children were often invited into the conversation during the intervention group's appointments. The pediatricians were attentive in their interactions with the child and gave the latter time to express themselves, which in some appointments gave the children courage and capability to express their opinions. The pediatrician offered their support and built trust by trying to get the children to talk about their problems, thoughts, and reflections. Some pediatricians asked the children questions about their social life, friends, social exclusion, and mental health, while others only asked questions about the child's disease and treatment. The pediatricians had nearly all their attention directed to the child during some of the appointments in the intervention group. Although some children were mostly silent, they were then invited into the conversation through questions directed at them and being given enough time for answering and expressing their views. The conversations and questions during the various appointments mainly concerned the children's condition and possible symptoms. There were often open questions posed by the pediatricians.

When the parents in the intervention group said, during one appointment, that their child had a stomachache, the pediatrician turned to the child and asked questions about it. During some of the other appointments in the intervention group, the pediatricians asked how the children wanted the medical examination to be carried out, and sometimes the children expressed their views spontaneously. A girl said that she preferred to lie down during the examination. The children in the intervention group were often directly asked questions about medication, such as which medicines they took. The problems the children spoke about were, for example, pain when measuring blood pressure and a medicine that was only available as a tablet that the child had difficulties in swallowing. The children often demonstrated their knowledge about medicines; for example, they could list all the medicines they took and often the doses. Moreover, they spoke about incidents such as infections and blood transfusions since the last appointment, as well as new symptoms, such as bruises. For example, a 9-year-old girl in the intervention group said "The medicine is a bit awkward; we have to crush it and mix it with yoghurt.

The pediatricians offered emotional support to the children during many of the appointments in the intervention group by encouraging them to come to them if problems arose. The pediatricians generally wanted to know if the children had difficult or unpleasant experiences or if they had other questions and concerns. Several of the pediatricians explained to the children that they wanted to understand how the children thought and felt. A pediatrician explained at one appointment that it was sad but normal that the child could no longer ride a bike or ski downhill, but that it would be possible to do it again in the future, while another pediatrician praised the child's strength in showing her baldness despite everyone in school knowing and being able to see. The pediatricians also asked open questions about participation where they offered support to the children by allowing them to express their views: "Is there anything you have forgotten to ask about or anything else you want to say?" (pediatrician to girl, 9 years old, intervention group).

XSL•FO
RenderX

The pediatricians rarely invited the children into the conversation during the control group's appointments; for example, at one appointment, many questions were asked very quickly, not giving the child time to answer. The pediatricians gave instructions rapidly during the medical examination in some of the control group's appointments, not allowing the children to express their opinions.

Talking about private things helped the children relax and feel secure. The pediatrician talked to the children about private things in approximately half of the appointments in the intervention group. They talked about how the children spent the summer, whether they had swum, and whether they would be going away on holiday. Other subjects were also raised, such as being unable to go to school, difficulties in sports activities, and talking about computer games and Christmas presents. The pediatrician only talked to the children about private things, such as pets and school, in a few of the control group's appointments.

Children's Views Were Taken Into Account

The level in Stier's model concerning the children's views being taken into account pertains to the pediatricians' consideration of the children's views. Only a few brief episodes in the videos were found for this level. The children's views were taken into account during the interventions group's appointments. For example, at one appointment when a child asked her parent if she could go out and listen to music, it was noted by the pediatrician who said to the child that she could go out if that was what she preferred, and at another appointment, a child expressed a wish to lie down during the medical examination.

> *You can do it [the examination] when I'm lying down, that's better.* [girl, 9 years old, intervention group]
>
> *Jump up and lie down then [on the examination table].* [pediatrician]

A child in the control group was informed during an appointment of the need to stay and have treatment, and the child became very upset. The child's opinion was taken into account in a respectful and supportive way when the child started to scream and cry. The pediatrician, nurse, and parents were completely silent, allowing the child to express her disappointment. The adults showed empathy through eye contact and body language expressing that they were also sad on behalf of the child. A child protested about having to remove clothes during another appointment in the control group. This child's opinion was taken into account by the pediatrician who lifted the clothes up a little instead.

Children Were Involved in Decision Making

The children were not involved in decision making in their health care. The pediatricians made most of the decisions, purely medical, both in the control and intervention groups, and neither the parents nor the children participated in these decisions. The decisions concerned continuation or temporary suspension of treatment, a need for blood transfusion, removal of a subcutaneous venous port, and planning for new blood tests and were expressed as already having been determined. Only one brief episode was found, which was in the intervention group, where a nurse suggested that the child allowed blood samples to be taken prior to talking to the pediatrician, and the latter then asked the child if this was a good idea. The child participated in the decision by deciding whether blood samples could be taken prior to or after the appointment with the pediatrician.

Children Shared Power and Responsibility for Decision Making

No brief episode was found showing that children shared power and responsibility for decision making.

Discussion

Principal Findings

The overall results showed increased participation for the children in the intervention group using Sisom compared to the control group, especially in relation to the lower levels of participation (levels 1-3) in Shier's model [51]. Levels 4 and 5 concern the child's possibilities for sharing decision making and responsibility with health care professionals. These levels were barely noticeable in any of the appointments in the two groups. These findings confirm previous research that has shown that children's possibilities for participating in decisions that concern them in health care are almost nonexistent [50,53,54]. It has been highlighted both in research and clinical practice that there is a need for guidelines, strategies, and methods to enable a higher level of children's participation in clinical practice. This study showed that the use of the eHealth service Sisom can increase the health care professional's ability to promote children's participation, but there is much more to do to reach a higher level where the children are co-actors and share decisions about issues in their care. To change and restructure the working procedures for children's participation in practice, it is crucial to consider both the individual and contextual factors such as attitudes, values, knowledge, readiness for behavioral change, culture, resources, and priorities [55]. For example, it has been maintained that pediatric care needs guidance and support [56], education [57], training [58], and interventions and methods (for example, eHealth services such as Sisom) [29,53,59,60] to increase children's participation. This entails extensive effort by both the health care services provided for children and the professionals who work there to increase their ability to involve children in decision making. It is also necessary for them to access adapted and person-centered information for children to have the ability to participate in decisions [19,20]. The children in the intervention group received more information from the pediatrician, expressed in a child-centered way and adapted to the needs of the child, in comparison with the control group. However, in the control group, the children asked for more and further information than those in the intervention group. One explanation for these differences between the groups could be that the pediatricians in the intervention group were more focused on and paid greater attention to the child's problems expressed through the use of Sisom. The pediatricians thus provided information based on the children's perspectives. However, information in pediatric health care is a complex aspect. The children's preference for and ability to absorb information could vary depending on different factors such as personality, age, maturity, and disease

XSL•FO
RenderX

state [20]. Another factor could be that the child is afraid of appearing to be stupid or feeling too shy to ask for information [61]. The child's preferences could also vary in different situations and change over time [23,28,62,63]. Previous research has shown that children speak infrequently to health care professionals during pediatric appointments [64]. Information in pediatric health care has also been highlighted by children with different diseases as being important for their possibilities for participating in health care [65,66]. The directed qualitative analysis in our study revealed one further category termed "children received information," which is not included in Shier's model [51]. The health care professionals were giving the children information with the purpose of letting them know and understand what would be happening and to experience control over the situation.

It appears that health care professionals find it more difficult to give children space and opportunity to speak during appointments without having the support from an eHealth service such as Sisom. The children can convey problems and thoughts through Sisom, which can provide the health care professionals with a different way to engage with the children and to facilitate interaction. The compelling ambition, as emphasized in the United Nations Convention on the Rights of the Child [27], is that pediatric care and professionals' interaction with the child should lead to the children being listened to and supported in expressing their views, these views being taking into account, and children being involved in the decision-making processes [51]. The decision-making level was not, however, achieved in this study. Even in care situations where there is a strict treatment protocol that needs to be followed in order to succeed with the care, informing and interacting about the treatment and care plan with the child are important for the child to understand, be able to ask questions [19], and express how they want to be involved in decision making [19,23]. The pediatricians in the intervention group were more focused on the children and seemed to be more aware of their own behavior when interacting with the child, which is important for the children to have positive experiences with care [67,68]. The use of an eHealth service for children has the potential of facilitating transition towards more trustful and supportive behavior from health care professionals [37,38,69,70].

Methodological Considerations

The results in this study would have shown higher levels of participation for the children in the intervention group if based solely on the quantitative methodology, but the reasons for these results and how they came about would have remained unknown. Supplementing the methodology with a qualitative approach helped to gain an understanding of how the eHealth service, Sisom, can influence health care professionals through the visualization of the children's needs and promotion of interactions, thus demonstrating the benefit of a mixed method approach [48].

Video observations are used in studies to observe how people behave towards each other, how they act and interact, to see body language and glances, and to hear what they talk about [71]. A video-recorded observation can be preferred in complex processes, such as appointments with pediatricians in health care settings, where many different things can happen simultaneously, and the researcher is able to review the videos several times [72]. It was sometimes difficult to see facial expressions and hear who was talking and what they were saying in the analyzed videos, partly due to technical reasons but also due to the participants' location in the room. This led to the videos being reviewed many times to ensure that the content was perceived correctly. The pediatricians' behavior towards the children can vary based not only on the use of Sisom but also on the different personalities and attitudes of the pediatricians and the different clinics' organization and working procedures. However, all pediatricians have specialist expertise and lengthy experience working and interacting with children. The children were between 6 and 12 years old, which means different levels of development and verbality that may have affected their interactions in the appointments. However, this was not obvious in relation to the children's statements to the pediatrician or their verbalities. This may impair the study's credibility as we draw conclusions based on the differences between the control and intervention groups. A strategy was applied to strengthen the credibility and dependability of the study, whereby the coders were trained and the similarity in the analysis of the two researchers was assessed by interrater reliability. Detailed information about the analysis process has been provided, and examples of citations in the qualitative results are presented for confirmation.

Study Limitations

One limitation could be that the data were collected several years ago. The fact that the data were collected by other researchers led to difficulties in obtaining facts about the execution of this data collection. For example, it was unclear if the child had met the current pediatrician before. The sample size was too small to perform significant statistical calculations. It is possible that health care professionals are more child-centered today since the importance of this approach has received greater attention in recent years. However, even if we are aware of some limitations in this study, it is important to investigate the implementation of eHealth services such as Sisom in daily practice since there is still a lack of similar resources to promote children's participation in health care. This study can thus make an important contribution.

Conclusions

eHealth services, such as Sisom, can increase children's participation in health care and influence health care professionals' ways of communicating with children in health care services. The overall results showed increased participation for the children in the intervention group compared to the control group, but a higher level of participation where the child could share decision making and responsibility with health care professionals was not visible in any of the groups. Sisom provides opportunities for enhanced improvement in health care, but its capabilities are far from being fully utilized. Further studies employing a randomized control design focusing on the effects of eHealth services on children's health outcomes, perceived participation, and cost-effectiveness could make a

XSL•FO
RenderX

significant contribution to guiding implementation of eHealth services in pediatric care.

Acknowledgments

Funding for the research for the first authors from Region Halland and the Center for Research on Welfare, Health, and Sport at Halmstad University is gratefully acknowledged. The authors would like to acknowledge Professor Cornelia Ruland and the Center for Shared Decision Making and Collaborative Care Research at Oslo University Hospital for valuable support during the design of the study and for sharing recorded data used for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of questions in Sisom.
[DOCX File , 17 KB - jmir_v22i7e17673_app1.docx]

Multimedia Appendix 2
Observation scheme of the children's levels of participation.
[DOCX File , 20 KB - jmir_v22i7e17673_app2.docx]

Multimedia Appendix 3
Statements during appointments in the intervention and control groups.
[DOCX File , 17 KB - jmir_v22i7e17673_app3.docx]

Multimedia Appendix 4
Verbality in the intervention and control groups.
[DOCX File , 17 KB - jmir_v22i7e17673_app4.docx]

References

1. Furaijat G, Kleinert E, Simmenroth A, Müller F. Implementing a digital communication assistance tool to collect the medical history of refugee patients: DICTUM Friedland - an action-oriented mixed methods study protocol. BMC Health Serv Res 2019 Feb 06;19(1):103 [FREE Full text] [doi: 10.1186/s12913-019-3928-1] [Medline: 30728030]
2. Holtz BE, Murray KM, Hershey DD, Richman J, Dunneback JK, Vyas A, et al. The design and development of MyT1DHero: A mobile app for adolescents with type 1 diabetes and their parents. J Telemed Telecare 2019 Apr 01;25(3):172-180. [doi: 10.1177/1357633X17745470] [Medline: 29228854]
3. Nanah A, Bayoumi AB. The pros and cons of digital health communication tools in neurosurgery: a systematic review of literature. Neurosurg Rev 2020 Jun 17;43(3):835-846. [doi: 10.1007/s10143-018-1043-0] [Medline: 30334173]
4. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014 Jan 28;1(1):CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
5. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
6. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and Adolescents: A Systematic Review of Adolescents' Use of Mobile Phone and Tablet Apps That Support Personal Management of Their Chronic or Long-Term Physical Conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
7. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
8. Schalkers I, Parsons CS, Bunders JF, Dedding C. Health professionals' perspectives on children's and young people's participation in health care: a qualitative multihospital study. J Clinical Nursing 2016 Feb 16;25(7-8):1035-1044. [doi: 10.1111/jocn.13152] [Medline: 26879522]
9. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental Healthcare Technologies for Anxiety and Depression in Childhood and Adolescence: Systematic Review of Studies Reporting Implementation Outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]

XSL•FO
RenderX

10. Raaff C, Glazebrook C, Wharrad H. A systematic review of interactive multimedia interventions to promote children's communication with health professionals: implications for communicating with overweight children. BMC Med Inform Decis Mak 2014 Jan 22;14:8 [FREE Full text] [doi: 10.1186/1472-6947-14-8] [Medline: 24447844]
11. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
12. Mehdizadeh H, Asadi F, Mehrvar A, Nazemi E, Emami H. Smartphone apps to help children and adolescents with cancer and their families: a scoping review. Acta Oncol 2019 Jul;58(7):1003-1014. [doi: 10.1080/0284186X.2019.1588474] [Medline: 30915872]
13. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
14. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text Messaging and Mobile Phone Apps as Interventions to Improve Adherence in Adolescents With Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
15. Stålberg A, Sandberg A, Söderbäck M, Larsson T. The child's perspective as a guiding principle: Young children as co-designers in the design of an interactive application meant to facilitate participation in healthcare situations. J Biomed Inform 2016 Jun 2;61:149-158 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.024] [Medline: 27050824]
16. Coda A, Sculley D, Santos D, Girones X, Brosseau L, Smith DR, et al. Harnessing interactive technologies to improve health outcomes in juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2017 May 16;15(1):40 [FREE Full text] [doi: 10.1186/s12969-017-0168-y] [Medline: 28511689]
17. Nightingale R, McHugh G, Kirk S, Swallow V. Supporting children and young people to assume responsibility from their parents for the self-management of their long-term condition: An integrative review. Child Care Health Dev 2019 Mar;45(2):175-188. [doi: 10.1111/cch.12645] [Medline: 30690751]
18. Coyne I, Sheehan A, Heery E, While AE. Healthcare transition for adolescents and young adults with long-term conditions: Qualitative study of patients, parents and healthcare professionals' experiences. J Clin Nurs 2019 Nov;28(21-22):4062-4076. [doi: 10.1111/jocn.15006] [Medline: 31327174]
19. Day E, Jones L, Langner R, Bluebond-Langner M. Current understanding of decision-making in adolescents with cancer: A narrative systematic review. Palliative Medicine 2016 Jul 10;30(10):920-934. [doi: 10.1177/0269216316648072] [Medline: 27160700]
20. Davies A, Randall D. Perceptions of Children's Participation in Their Healthcare: A Critical Review. Issues Compr Pediatr Nurs 2015;38(3):202-221. [doi: 10.3109/01460862.2015.1063740] [Medline: 26331450]
21. Siembida EJ, Kadan-Lottick NS, Moss K, Bellizzi KM. Adolescent cancer patients' perceived quality of cancer care: The roles of patient engagement and supporting independence. Patient Education and Counseling 2018 Sep;101(9):1683-1689. [doi: 10.1016/j.pec.2018.04.002]
22. Wang J, Jacobs S, Dewalt DA, Stern E, Gross H, Hinds PS. A Longitudinal Study of PROMIS Pediatric Symptom Clusters in Children Undergoing Chemotherapy. J Pain Symptom Manage 2018 Feb;55(2):359-367. [doi: 10.1016/j.jpainsymman.2017.08.021] [Medline: 28865867]
23. Ruhe KM, Badarau DO, Brazzola P, Hengartner H, Elger BS, Wangmo T. Participation in pediatric oncology: views of child and adolescent patients. Psycho-Oncology 2015 Dec 17;25(9):1036-1042. [doi: 10.1002/pon.4053] [Medline: 26677126]
24. Woodgate RL. Feeling states: a new approach to understanding how children and adolescents with cancer experience symptoms. Cancer Nurs 2008;31(3):229-238. [doi: 10.1097/01.NCC.0000305731.95839.ca] [Medline: 18453880]
25. Gårdling J, Edwinson Månsson M, Törnqvist E, Hallström I. Caring for children undergoing radiotherapy treatment: Swedish radiotherapy nurses' perceptions. Eur J Oncol Nurs 2015 Dec;19(6):660-666. [doi: 10.1016/j.ejon.2015.04.004] [Medline: 26002987]
26. Wo SW, Ong LC, Low WY, Lai PSM. Exploring the needs and challenges of parents and their children in childhood epilepsy care: A qualitative study. Epilepsy Behav 2018 Nov;88:268-276. [doi: 10.1016/j.yebeh.2018.09.018] [Medline: 30321756]
27. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why Do They Need to Check Me?" Patient Participation Through eHealth and the Doctor-Patient Relationship: Qualitative Study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
28. UN. United Nations Convention on the Rights of the Child. United Nations Convention on the Rights of the Child. URL: https://www.unicef.org/child-rights-convention: [accessed 2019-12-20]
29. Coyne I, Amory A, Gibson F, Kiernan G. Information-sharing between healthcare professionals, parents and children with cancer: more than a matter of information exchange. Eur J Cancer Care (Engl) 2016 Jan;25(1):141-156. [doi: 10.1111/ecc.12411] [Medline: 26537295]
30. Quaye AA, Coyne I, Söderbäck M, Hallström IK. Children's active participation in decision‐making processes during hospitalisation: An observational study. J Clinical Nursing 2019 Sep 17;28(23-24):4525-4537. [doi: 10.1111/jocn.15042] [Medline: 31430412]

XSL•FO
RenderX

31. Ruland CM, Slaughter L, Starren J, Vatne TM, Moe EY. Children's contributions to designing a communication tool for children with cancer. Stud Health Technol Inform 2007;129(Pt 2):977-982. [Medline: 17911861]
32. Ruland CM, Starren J, Vatne TM. Participatory design with children in the development of a support system for patient-centered care in pediatric oncology. J Biomed Inform 2008 Aug;41(4):624-635 [FREE Full text] [doi: 10.1016/j.jbi.2007.10.004] [Medline: 18082468]
33. Arvidsson S, Gilljam B, Nygren J, Ruland CM, Nordby-Bøe T, Svedberg P. Redesign and Validation of Sisom, an Interactive Assessment and Communication Tool for Children With Cancer. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e76 [FREE Full text] [doi: 10.2196/mhealth.5715] [Medline: 27343004]
34. Baggott C, Baird J, Hinds P, Ruland CM, Miaskowski C. Evaluation of Sisom: A computer-based animated tool to elicit symptoms and psychosocial concerns from children with cancer. Eur J Oncol Nurs 2015 Aug;19(4):359-369. [doi: 10.1016/j.ejon.2015.01.006] [Medline: 25691300]
35. Tsimicalis A, Stone PW, Bakken S, Yoon S, Sands S, Porter R, et al. Usability testing of a computerized communication tool in a diverse urban pediatric population. Cancer Nurs 2014;37(6):E25-E34 [FREE Full text] [doi: 10.1097/NCC.0000000000000115] [Medline: 24457227]
36. Tsimicalis A, Rennick J, Stinson J, May SL, Louli J, Choquette A, et al. Usability Testing of an Interactive Communication Tool to Help Children Express Their Cancer Symptoms. J Pediatr Oncol Nurs 2018;35(5):320-331. [doi: 10.1177/1043454218777728] [Medline: 29848164]
37. Vatne TM, Finset A, Ørnes K, Holmstrøm H, Småstuen MC, Ruland CM. Effects of an interactive symptom communication tool for children with heart disease on patient-provider communication in outpatient care: Preliminary results. Journal of Communication in Healthcare 2013 Nov 15;6(2):106-114. [doi: 10.1179/1753807612Y.0000000020]
38. Svedberg P, Arvidsson S, Larsson I, Carlsson I, Nygren JM. Barriers and Enablers Affecting Successful Implementation of the Electronic Health Service Sisom: Multicenter Study of Child Participation in Pediatric Care. J Med Internet Res 2019 Nov 15;21(11):14271. [doi: 10.2196/14271] [Medline: 31730040]
39. Badawy SM, Thompson AA, Kuhns LM. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
40. Larsson I, Staland-Nyman C, Svedberg P, Nygren JM, Carlsson I. Children and young people's participation in developing interventions in health and well-being: a scoping review. BMC Health Serv Res 2018 Jun 28;18(1):507 [FREE Full text] [doi: 10.1186/s12913-018-3219-2] [Medline: 29954392]
41. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
42. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
43. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
44. Grove S, Burns N, Gray J. The practice of nursing research: Appraisal, synthesis, and generation of evidence. 7th ed: Elsevier Health Sciences; 2013.
45. Chiang-Hanisko L, Newman D, Dyess S, Piyakong D, Liehr P. Guidance for using mixed methods design in nursing practice research. Appl Nurs Res 2016 Aug;31:1-5. [doi: 10.1016/j.apnr.2015.12.006] [Medline: 27397810]
46. Creswell JD, Creswell JW. Research design: Qualitative, quantitative, and mixed methods approaches. Thousand Oaks, CA: Sage Publications, Inc; 2014.
47. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
48. Klassen AC, Creswell J, Plano Clark VL, Smith KC, Meissner HI. Best practices in mixed methods for quality of life research. Qual Life Res 2012 Apr;21(3):377-380. [doi: 10.1007/s11136-012-0122-x] [Medline: 22311251]
49. Polit D, Beck T. Essentials of nursing research. Appraising evidence for nursing practice. Ninth edition. Philadelphia, PA: Wolters Kluwer; 2018.
50. Vatne TM, Ruland CM, Ørnes K, Finset A. Children's expressions of negative emotions and adults' responses during routine cardiac consultations. J Pediatr Psychol 2012 Mar 09;37(2):232-240. [doi: 10.1093/jpepsy/jsr074] [Medline: 21908544]
51. Shier H. Pathways to participation: openings, opportunities and obligations. Child. Soc 2001 Apr;15(2):107-117. [doi: 10.1002/chi.617]
52. Goodell LS, Stage VC, Cooke NK. Practical Qualitative Research Strategies: Training Interviewers and Coders. J Nutr Educ Behav 2016 Sep;48(8):578-585.e1. [doi: 10.1016/j.jneb.2016.06.001] [Medline: 27395426]
53. Wyatt KD, List B, Brinkman WB, Prutsky Lopez G, Asi N, Erwin P, et al. Shared Decision Making in Pediatrics: A Systematic Review and Meta-analysis. Acad Pediatr 2015;15(6):573-583 [FREE Full text] [doi: 10.1016/j.acap.2015.03.011] [Medline: 25983006]

XSL•FO
RenderX

54. Gondek D, Edbrooke-Childs J, Velikonja T, Chapman L, Saunders F, Hayes D, et al. Facilitators and Barriers to Person-centred Care in Child and Young People Mental Health Services: A Systematic Review. Clin Psychol Psychother 2017 Jul 02;24(4):870-886. [doi: 10.1002/cpp.2052] [Medline: 27910173]
55. Scholl I, LaRussa A, Hahlweg P, Kobrin S, Elwyn G. Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. Implement Sci 2018 Mar 09;13(1):40 [FREE Full text] [doi: 10.1186/s13012-018-0731-z] [Medline: 29523167]
56. Harder M, Söderbäck M, Ranheim A. Health care professionals' perspective on children's participation in health care situations: encounters in mutuality and alienation. International Journal of Qualitative Studies on Health and Well-being 2018 Dec 20;13(1):1555421. [doi: 10.1080/17482631.2018.1555421] [Medline: 30704374]
57. Kluthe C, Isaac DM, Hiller K, Carroll M, Wine E, van Manen M, et al. Qualitative Analysis of Pediatric Patient and Caregiver Perspectives After Recent Diagnosis With Inflammatory Bowel Disease. J Pediatr Nurs 2018;38:106-113. [doi: 10.1016/j.pedn.2017.11.011] [Medline: 29357985]
58. Wiering BM, Noordman J, Tates K, Zwaanswijk M, Elwyn G, De Bont ESJM, et al. Sharing decisions during diagnostic consultations; an observational study in pediatric oncology. Patient Educ Couns 2016 Jan;99(1):61-67. [doi: 10.1016/j.pec.2015.07.026] [Medline: 26298218]
59. Kaziunas E, Hanauer DA, Ackerman MS, Choi SW. Identifying unmet informational needs in the inpatient setting to increase patient and caregiver engagement in the context of pediatric hematopoietic stem cell transplantation. J Am Med Inform Assoc 2016 Jan;23(1):94-104 [FREE Full text] [doi: 10.1093/jamia/ocv116] [Medline: 26510878]
60. Lipstein EA, Brinkman WB, Fiks AG, Hendrix KS, Kryworuchko J, Miller VA, et al. An emerging field of research: challenges in pediatric decision making. Med Decis Making 2015 Apr 21;35(3):403-408 [FREE Full text] [doi: 10.1177/0272989X14546901] [Medline: 25145576]
61. Gutman T, Hanson CS, Bernays S, Craig JC, Sinha A, Dart A, et al. Child and Parental Perspectives on Communication and Decision Making in Pediatric CKD: A Focus Group Study. Am J Kidney Dis 2018 Oct;72(4):547-559. [doi: 10.1053/j.ajkd.2018.05.005] [Medline: 29980375]
62. Coyne I, Amory A, Kiernan G, Gibson F. Children's participation in shared decision-making: children, adolescents, parents and healthcare professionals' perspectives and experiences. Eur J Oncol Nurs 2014 Jun;18(3):273-280. [doi: 10.1016/j.ejon.2014.01.006] [Medline: 24629505]
63. Weaver MS, Baker JN, Gattuso JS, Gibson DV, Sykes AD, Hinds PS. Adolescents' preferences for treatment decisional involvement during their cancer. Cancer 2015 Sep 08;121(24):4416-4424. [doi: 10.1002/cncr.29663] [Medline: 26348790]
64. Vigilante VA, Hossain J, Wysocki T, Sharif I. Correlates of type and quantity of child communication during pediatric subspecialty encounters. Patient Educ Couns 2015 Nov;98(11):1352-1359. [doi: 10.1016/j.pec.2015.05.001] [Medline: 26160039]
65. Gilljam B, Arvidsson S, Nygren JM, Svedberg P. Promoting participation in healthcare situations for children with JIA: a grounded theory study. Int J Qual Stud Health Well-being 2016;11:30518 [FREE Full text] [doi: 10.3402/qhw.v11.30518] [Medline: 27172512]
66. Sjöberg C, Amhliden H, Nygren JM, Arvidsson S, Svedberg P. The perspective of children on factors influencing their participation in perioperative care. J Clin Nurs 2015 Oct;24(19-20):2945-2953. [doi: 10.1111/jocn.12911] [Medline: 26215896]
67. Ambresin A, Bennett K, Patton GC, Sanci LA, Sawyer SM. Assessment of youth-friendly health care: a systematic review of indicators drawn from young people's perspectives. J Adolesc Health 2013 Jun;52(6):670-681. [doi: 10.1016/j.jadohealth.2012.12.014] [Medline: 23701887]
68. Koller D. 'Kids need to talk too': inclusive practices for children's healthcare education and participation. J Clin Nurs 2017 Sep;26(17-18):2657-2668. [doi: 10.1111/jocn.13703] [Medline: 28001330]
69. Downing J, Gleeson H, Clayton P, Davis J, Dimitri P, Wales J, et al. Communication with young people in paediatric and adult endocrine consultations: an intervention development and feasibility study. BMC Endocr Disord 2017 Jun 15;17(1):33 [FREE Full text] [doi: 10.1186/s12902-017-0182-6] [Medline: 28619024]
70. Kaplan HC, Thakkar SN, Burns L, Chini B, Dykes DM, McPhail GL, et al. Protocol of a Pilot Study of Technology-Enabled Coproduction in Pediatric Chronic Illness Care. JMIR Res Protoc 2017 Apr 28;6(4):e71 [FREE Full text] [doi: 10.2196/resprot.7074] [Medline: 28455274]
71. Hammersley M, Atkinson P. Ethnography, principles in practice. London and New York: Routledge; 2007.
72. Yanes AF, McElroy LM, Abecassis ZA, Holl J, Woods D, Ladner DP. Observation for assessment of clinician performance: a narrative review. BMJ Qual Saf 2016 Jan 30;25(1):46-55. [doi: 10.1136/bmjqs-2015-004171] [Medline: 26424762]

Abbreviations

eHealth: electronic health.

XSL•FO
RenderX

Edited by G Eysenbach; submitted 03.01.20; peer-reviewed by S Badawy, A Ahmadvand, R Knoerl; comments to author 10.03.20; revised version received 04.05.20; accepted 13.05.20; published 28.07.20.

Please cite as:
Gilljam BM, Nygren JM, Svedberg P, Arvidsson S
Impact of an Electronic Health Service on Child Participation in Pediatric Oncology Care: Quasiexperimental Study
J Med Internet Res 2020;22(7):e17673
URL: http://www.jmir.org/2020/7/e17673/
doi:10.2196/17673
PMID:32720907

©Britt-Mari Gilljam, Jens M Nygren, Petra Svedberg, Susann Arvidsson. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Teens Taking Charge: A Randomized Controlled Trial of a Web-Based Self-Management Program With Telephone Support for Adolescents With Juvenile Idiopathic Arthritis

Jennifer N Stinson[1], PhD, RN-EC, CPNP; Chitra Lalloo[1,2], PhD, BHSc; Amos S Hundert[1], BSc; Sarah Campillo[3], MD, FRCPC, MSc; Tania Cellucci[4], MD, FRCPC, MScCH; Paul Dancey[5], MD, FRCPC; Ciaran Duffy[6], MB, BCh, MSc, FRCPC, FRCPI; Janet Ellsworth[7], MD, FRCPC; Brian M Feldman[1], MD, MSc, FRCPC; Adam M Huber[8], BA, MD, FRCPC Pediatrics, FRCPC Rheumatology, MSc; Nicole Johnson[9], MD, RCPSC; Geert't Jong[10], MD, PhD; Kiem Oen[11], MD; Alan M Rosenberg[12], BA, MD, FRCPC; Natalie J Shiff[13], MD; Lynn Spiegel[14], MD, FRCPC; Shirley M L Tse[1], MD, FRCPC; Lori Tucker[15], MD; Joseph Charles Victor[16], MSc, BSc

[1]The Hospital for Sick Children, Toronto, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Montreal Children's Hospital, Montreal, QC, Canada

[4]Department of Pediatrics, Division of Rheumatology, McMaster University, Hamilton, ON, Canada

[5]Department of Pediatrics, Division of Rheumatology, Memorial University of Newfoundland, St. John's, NL, Canada

[6]Department of Pediatrics, Division of Rheumatology, McGill Research Institute, Montreal, QC, Canada

[7]Department of Pediatrics, Division of Rheumatology, University of Alberta, Edmonton, AB, Canada

[8]IWK Health Centre, Halifax, NS, Canada

[9]Alberta Children's Hospital, Calgary, AB, Canada

[10]Children's Hospital Research Institute of Manitoba, Winnipeg, MB, Canada

[11]Pediatrics and Child Health, University of Manitoba, Winnipeg, MB, Canada

[12]University of Saskatchewan, Saskatoon, SK, Canada

[13]Biobehavioral Nursing Science, University of Florida, Gainesville, FL, United States

[14]Hospital for Sick Children, Toronto, ON, Canada

[15]Department of Pediatrics, Division of Rheumatology, University of British Columbia, Vancouver, ON, Canada

[16]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

Corresponding Author:
Jennifer N Stinson, PhD, RN-EC, CPNP
The Hospital for Sick Children
555 University Ave
Toronto, ON
Canada
Phone: 1 (416) 8131500
Email: jennifer.stinson@sickkids.ca

Abstract

Background: Juvenile idiopathic arthritis (JIA) is a serious and potentially debilitating pediatric illness. Improved disease self-management may help to improve health outcomes.

Objective: This study aimed to evaluate the effectiveness of the *Teens Taking Charge* Web-based self-management intervention in reducing symptoms and improving health-related quality of life (HRQL) in adolescents with JIA compared with a Web-based education control condition.

Methods: Adolescents with JIA aged 12 to 18 years were recruited from 11 Canadian pediatric rheumatology centers. Caregivers were invited to participate along with their child. In addition to standard medical care, participants were randomized to receive either (1) the *Teens Taking Charge* self-management intervention or (2) a Web-based education control condition for a period of 12 weeks. Adolescents in the intervention group completed website modules addressing cognitive behavioral coping skills, stress management, and other self-management topics, while also receiving monthly telephone calls from a trained health coach.

XSL•FO
RenderX

Adolescents in the education control group were instructed to view a series of preselected public JIA educational websites and received monthly calls from a coach who asked about *their own best efforts* at managing JIA. Caregivers in the intervention group completed website modules related to promoting independence and disease self-management in their child. Caregivers in the education control group were instructed to view a series of preselected public JIA educational websites. Outcome assessment occurred at baseline, 12 weeks (posttreatment), and at 6 and 12 months postrandomization. The primary outcomes were pain intensity, pain interference, and HRQL. Secondary outcomes were emotional symptoms, adherence, coping, knowledge, and self-efficacy.

Results: In total, 333 adolescents and 306 caregivers were enrolled. Significant overall reductions in pain intensity (P=.02) and pain interference (P=.007) were observed for intervention group participants compared with those in the education control group, after adjusting for baseline levels. There was a significant overall improvement in HRQL related to problems with pain (P=.02) and problems with daily activities (P=.01). There was also a significant difference in the intervention group over time (P=.008) for HRQL related to treatment problems, with the intervention group participants demonstrating improved HRQL by 12 months compared with education control group participants. Both groups showed nonsignificant improvements compared with baseline in other primary outcomes. There were no significant differences between the groups in any secondary outcomes or caregiver-reported outcomes.

Conclusions: The results of this randomized trial suggest that the *Teens Taking Charge* Web-based intervention is effective at reducing both pain intensity and pain interference, as well as improving HRQL in adolescents with JIA, compared with education control. These effects are sustained for up to 12 months following program completion. The *Teens Taking Charge* program is now publicly available at no cost.

Trial Registration: ClinicalTrials.gov NCT01572896; https://clinicaltrials.gov/ct2/show/NCT01572896

(J Med Internet Res 2020;22(7):e16234) doi:10.2196/16234

KEYWORDS

eHealth; randomized controlled trial; adolescents; juvenile idiopathic arthritis; self-management; self-efficacy; technology; patient education; internet; pediatric pain

Introduction

Background

Juvenile idiopathic arthritis (JIA) is a serious and potentially debilitating pediatric illness. It is the most common rheumatic disease in childhood, affecting approximately 1 in 1000 children [1]. The disease course can be unpredictable, and children commonly experience both physical (eg, pain, fatigue, stiffness) and emotional (eg, stress, anxiety, depression) symptoms that may restrict day-to-day function [2-7]. Persistent pain is common and can contribute to adverse psychological effects and functional disability [8,9]. As children mature, they are expected to assume increasing responsibility for disease self-management concomitant with their growing independence and autonomy. However, adherence to disease management is typically suboptimal [10,11]. Poor adherence and inappropriate strategies for self-management may reduce the potential benefits of treatment and impact health-related quality of life (HRQL) [10,11].

Improved disease self-management early in the disease trajectory may help to improve health outcomes. Web-based self-management interventions are a promising strategy for improving the accessibility and availability of education, psychosocial interventions, and social support for adolescents with JIA [12,13]. There is preliminary evidence that self-management interventions can improve symptoms and health status in certain childhood illnesses [14-20].

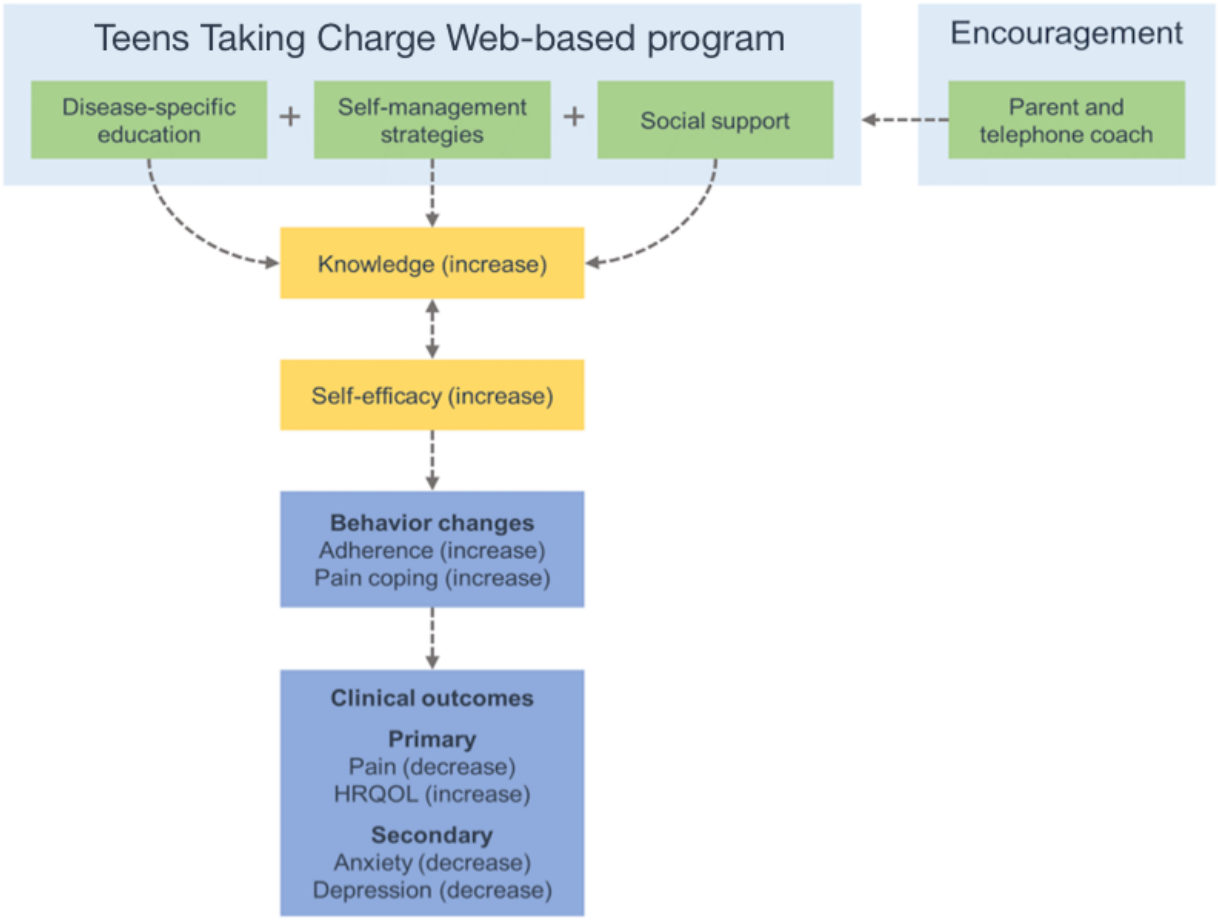
Teens Taking Charge is a Web-based treatment that was developed to meet the need for a high-quality, accessible, empirically grounded, and multilingual self-management program for adolescents with JIA [21-23]. The program was developed using a sequential phased approach. In phase 1, a qualitative needs assessment was conducted to identify the self-management needs of adolescents with JIA [21]. Adolescents articulated a universal need for more JIA-specific knowledge, self-management strategies, and meaningful social support to better manage their JIA. In phase 2, a prototype website was developed, and it underwent iterative cycles of usability testing to ensure ease of use and understanding of the program [22]. In phase 3, a pilot feasibility randomized controlled trial (RCT) was completed with 46 adolescents with JIA and their caregivers [23]. The pilot RCT supported the feasibility (high acceptability, user satisfaction, and compliance) and initial effectiveness of the program in improving knowledge and decreasing pain in adolescents with JIA compared with an education control condition.

Objective

The objective of this study was to evaluate the effectiveness of the *Teens Taking Charge* Web-based self-management intervention in reducing symptoms and improving HRQL in adolescents with JIA compared with a Web-based education control condition. We hypothesized that adolescents randomized to the intervention group would demonstrate (1) reduced pain and improved HRQL; (1) reduced emotional (anxiety and depression) symptoms; and (3) increased treatment adherence, pain coping, JIA-specific knowledge, and self-efficacy compared with adolescents randomized to the control group. See Figure 1 for a summary of the hypothesized mechanism of action of the *Teens Taking Charge* intervention on health outcomes. This model is based on the published work of Murray et al [24].

XSL•FO
RenderX

Figure 1. Conceptual model outlining the hypothesized mechanism of action of the Teens Taking Charge: Managing JIA Online Web-based intervention. JIA: juvenile idiopathic arthritis. HRQL: health-related quality of life.



Methods

Trial Design

A 2-arm parallel group RCT design was conducted to evaluate the effectiveness of the *Teens Taking Charge* program in comparison with a Web-based education control condition. The allocation ratio was 1:1. The trial was registered on ClinicalTrials.gov with the identifier, NCT01572896. The study was approved by the local research ethics boards at each of the participating institutions.

Participants

Adolescents were recruited in person from 11 Canadian pediatric rheumatology centers. A local research coordinator explained the study to eligible adolescents and obtained written consent.

Parents and caregivers were invited to participate with their child. Adolescents were eligible to participate if they were (1) between 12 and 17.9 years of age, (2) diagnosed with JIA as per their rheumatologist, (3) able to speak and read English or French, and (4) able to complete Web-based outcome measures as per self-report. Adolescents were excluded if they had (1) major comorbid illnesses or cognitive impairments that could affect their ability to use and understand the Web-based program as per their health care provider or (2) were currently participating in other cognitive behavioral therapy interventions.

Parents and caregivers were eligible if they were (1) able to speak and read English or French and (2) able to complete parent modules and outcome measures. Participants received gift cards (total value Can $100 [US $75]) in recognition of their time and commitment to the study.

Study Conditions

Intervention Condition

The intervention was based on cognitive behavioral principles of improving pain and HRQL through providing disease education, empowering patients with coping skills to help manage disease symptoms and stress, and providing validation and encouragement through opportunities for social support.

The intervention condition had three components: (1) a 12-module adolescent website consisting of JIA-specific education, self-management strategies, and social support, (2) telephone health coaches for the adolescent only, and (3) a 2-module caregiver website consisting of education about the impact of JIA and strategies to support their adolescent in self-management.

The adolescent and caregiver websites were available in English and French. The module content is summarized in Table 1. All content was developed by a team of experts from across Canada including study investigators, allied health team members, and adolescent medicine health care professionals, and patient

XSL•FO
RenderX

advocates. Modules were written at a grade 6 to grade 7 level, were developmentally appropriate for those aged 12 to 18 years, and were geared to address the needs identified in the phase 1 qualitative study [21]. See Figure 2 for screenshots of the *Teens Taking Charge* adolescent website.

Participants were instructed to work through the modules over a 12-week period at their own pace to allow sufficient flexibility to accommodate for exams, holidays, and illnesses. Participants received automated standard weekly emails to remind them to complete their next module and to congratulate them on module completion.

Adolescents in the intervention group also received monthly telephone calls from a health coach who was a trained non–health care professional [23]. The telephone coach (1) reviewed content from the previous 4 weeks (eg, assignments, knowledge quizzes, and goals); (2) determined whether the participant completed the modules and answered questions regarding the material and or self-management skills; and (3) provided guidance and helped problem-solve any issues with the program content that may have arisen in the previous 4 weeks. If participants asked questions that the coach could not answer, the coach redirected them to their rheumatology health care team. All calls were audio recorded and audited every 3 months to ensure integrity of the intervention delivery.

Table 1. Module breakdown of the Teens Taking Charge intervention for adolescents and caregivers.

Module number and title	Content description
Adolescent modules	
1. About JIA[a]	Clinical features of JIA
2. Understanding diagnosis	How JIA is diagnosed, coping, and health check-ups
3. Managing your symptoms	Understanding and managing pain, fatigue, and stiffness
4. Managing stress	Understanding and managing stress
5. Relaxation	Belly breathing, relaxation with and without tension, mini relaxation, and behavioral reversal
6. JIA medications	Overview of medications for JIA and medication management
7. Distraction	Attention focusing, imagery, mental games, and pleasant activities
8. Other types of care	Physical activity, nutrition, orthotics, occupational therapy, psychological therapy, complementary therapy, surgical procedures, eye care, and dental care
9. Managing your thoughts	Understanding stress and thinking; changing negative thoughts
10. Therapies, self-monitoring, and supports	Getting the most from your treatment, self-monitoring, and communication with the health care team
11. Your lifestyle	Staying active, healthy eating, and lifestyle choices
12. Looking ahead	Transitioning to adult care, preparing for higher education and employment, and maintaining a treatment program
Caregiver modules	
1. Impact of arthritis	Impact of JIA on the family and finances; JIA disease education
2. Letting go	Communicating with your teen and helping your teen to take control of disease management

[a]JIA: juvenile idiopathic arthritis.

XSL•FO
RenderX

Figure 2. An example screenshot of the Teens Taking Charge: Managing JIA Online adolescent Web intervention.



Education Control Condition

This condition was designed to control for the potential effects on outcomes of time, attention, and computer use during the intervention and through the follow-up period. In addition to standard medical care, adolescents and caregivers in the education control group were provided with access to a self-guided education study website called the *JIA Resource Centre*. It featured links to 12 publicly available JIA websites (identified from a systematic review) that provided patient education, but did not offer self-management strategies or opportunities for social support [25]. The information in these websites was similar to the materials that are routinely provided at the time of diagnosis in the participating recruitment centers. The educational websites were monitored on a weekly basis to ensure that they did not add any *active ingredients* during the trial. See Multimedia Appendix 1 for a list of websites in the *JIA Resource Centre.* Adolescents and caregivers were encouraged to browse one website per week from the *JIA Resource Centre*. They received automated email reminders each week to access the education content.

Each adolescent in the education control group also received monthly telephone calls from a coach. The coaches were trained non–health care professionals who encouraged participants to view the websites and also discussed adolescents' *own best efforts* at managing their JIA over the 12-week period. The coach used standardized scripts for these calls. They did not provide any advice or self-management support to participants, and the interaction was minimal. If a participant asked for information related to self-management of their JIA, the coach redirected them to their rheumatology health care team. All calls were recorded and audited to ensure integrity of the control calls. Participants from the education control group were offered the full intervention following their trial completion (at 12-month time point after completion of all outcome measures) for a period of 3 months.

Outcomes

Outcome data included a combination of adolescent self-reports and parent-proxy reports, as outlined in Multimedia Appendix 2 [26-34]. Measures were completed at four different time points: baseline (after consent, before randomization; T_1), condition completion (3 months after randomization; T_2), 6 months after randomization (T_3), and 12 months after randomization (T_4). Measures were completed online by study participants through the Research Electronic Data Capture secure Web-based system hosted at the Hospital for Sick Children. The lead study coordinator contacted participants via telephone, email, and mail to remind them to complete follow-up measures. All measures have evidence of validity and reliability for adolescents with arthritis and were available in English and French.

XSL•FO
RenderX

Sample Size

Sample size was calculated by assuming an effect size (standardized group difference) of 0.31 units on the outcome of average pain intensity over a 12-month follow-up because this magnitude of difference is considered the minimally clinically significant important group difference for pain intensity [35]. It is also congruent with the treatment effect observed during pilot work [23]. Assuming a type I error rate (alpha) of .05 and the maximum correlation between repeated pain measurements to be 0.90, a sample size of 117 participants per group (234 in total) was required to achieve 80% power to detect an effect of this magnitude or larger. To account for the lack of compliance in completing online measures, we assumed a loss to follow-up of 20% based on our pilot RCT (13%) [23] and systematic review [36]. Therefore, we planned to randomize 294 adolescents to one of the two groups.

Randomization

Randomization was centrally controlled, concealed, and balanced by physician-rated disease activity into those with low (<3/10) and moderate-to-severe (>4/10) disease activity [37] and by study center (ie, to control for center-specific education, transitional care programs, and treatment approaches). A secure, Web-based randomization service was used for allocating participants to the trial groups. Following randomization, participants were sent instructions on how to access their assigned program (website link and log-in) and were contacted by their assigned coach.

Blinding

This was a single blind study (adolescent and caregiver only). Both groups received a website and monthly telephone calls. Participants were informed that they would be chosen at random to either receive the *Taking Charge: Managing JIA Online* internet education program or the *JIA Resource Centre* internet education program. They were not explicitly told which program was considered the intervention condition.

Statistical Methods

As per an intent-to-treat approach, all participants were included in the final analysis and according to the arm (intervention or control) to which they were randomized. Only participants who completed at least one follow-up outcome assessment after baseline were included in the analysis. In addition, outcome assessments were required to be completed within 35 days of each time point (ie, 3 months, 6 months, 12 months) to be included in the analysis for that time point. This 35-day window was used because the *PedsQL Rheumatology Module* asks respondents to describe their function over the previous month [38]. Thus, participants who completed their outcome assessments beyond the assigned time point were reporting their function within a different window of time compared with participants who completed assessments on schedule and the data were considered missing.

Data were analyzed using the SAS software (Version 9.1.3. developed by SAS Institute Inc, Cary, North Carolina, 2006). [39]. Descriptive statistics were used to describe characteristics of the sample. A significance level of .05 was used for the primary question. A Bonferroni-adjusted alpha level of .007 was used to maintain an overall level of .05 for all secondary and other study outcomes. Linear mixed models were used to assess the effects of the intervention on primary, secondary, and other outcomes using the baseline scores as covariates. Models used an autoregressive first order covariance structure, which allows for correlations between measurements to decline as they are further apart in time [40,41]. To assess the effects of disease-related variables (eg, disease activity, duration of illness, and disease subtype) and age on primary and secondary outcomes, separate linear mixed models of each posttreatment measure were constructed using pretreatment disease-related variables and age as covariates [42].

Data Safety and Monitoring

A Data Safety and Monitoring Board (DSMB) was formed to regularly review the safety and treatment-specific efficacy data of the trial. The board included experts in translational research, pediatric rheumatology, chronic pain, cognitive behavioral therapy, RCTs, and biostatistics. The DSMB met via teleconference every 6 months for the duration of the trial.

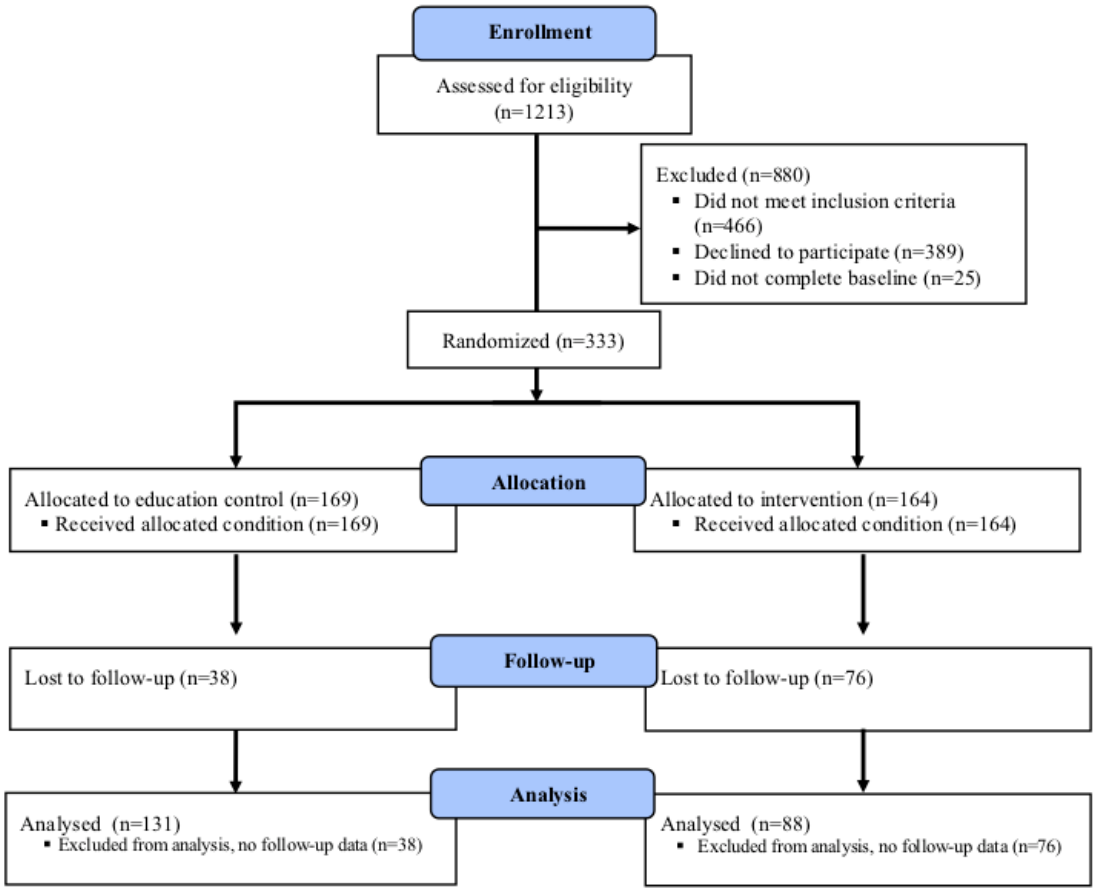
Results

Participants

The Consolidated Standards of Reporting Trials (CONSORT) [43] flow diagram detailing participant enrolment, allocation, follow-up, and analysis is provided in Figure 3. The randomized study sample consisted of 333 adolescents. As illustrated in the CONSORT diagram, 76/164 (46.3%) participants in the intervention group and 38/169 (22.5%) in the control group were lost to follow-up. Those participants lost to follow-up did not have any significant differences in baseline characteristics with the exception of the PedsQL worry subscale (lost to follow-up=40.0; analytic=45.9; *P*=.03).

XSL•FO
RenderX

Figure 3. Consolidated Standards of Reporting Trials flow diagram.



Across the entire sample, 100/604 (16.6%) outcome assessments were completed beyond the 35-day window for that time point. The breakdown of overdue assessments across time points was as follows: 3 months (24/197, 12.2%); 6 months (38/200, 19.0%); and 12 months (38/207, 18.4%). Comparing the groups, 45/251 (17.9%) intervention outcome assessments and 55/353 (15.6%) control outcome assessments were completed outside the window. Similarly, 101/541 (18.7%) caregiver outcome assessments were completed beyond the 35-day window for that time point. The breakdown of overdue caregiver assessments across time points was: 3 months (26/169, 15.4%); 6 months (35/182, 19.2%); 12 months (40/190, 21.1%). Comparing the caregiver study groups, 46/216 (21.3%) intervention assessments and 55/325 (16.9%) control assessments missed the allowed window.

Characteristics of the analyzed adolescent sample by assigned condition are provided in Table 2. The analyzed caregiver sample consisted of biological mothers (158/197, 80.2%), biological fathers (36/197, 18.3%), and adoptive parents or grandparents (3/197, 1.5%). More than half of the caregiver sample (129/197, 65.5%) had graduated from college or graduate school.

XSL•FO
RenderX

Table 2. Adolescent demographic and disease characteristics.

Characteristic	Total (N=219)	Intervention (n=88)	Control (n=131)
Age (years), mean (SD)	14.4 (1.6)	14 (1.5)	14.5 (1.7)
Sex, n (%)			
Female	154 (70.3)	63 (72)	91 (69.5)
Male	65 (29.7)	25 (28)	40 (30.5)
Language, n (%)			
English	214 (97.7)	87 (99)	127 (96.9)
French	5 (2.3)	1 (1)	4 (3.1)
Annual household income (Can $), n (%)[a]			
Less than $25,000 (US $18,263)	9 (4.6)	2 (3)	7 (5.7)
$25,000-49,999 (US $18,263-US $36,526)	25 (12.7)	9 (12)	16 (13.1)
$50,000-74,999 (US $36,526-US $54,790)	28 (14.2)	6 (8)	22 (18.0)
$75,000-99,999 (US $54,790-US $73,053)	29 (14.7)	13 (17)	16 (13.1)
$100,000-150,000 (US $73,053-US $109,580)	33 (16.8)	12 (16)	21 (17.2)
Did not answer	44 (22.3)	18 (24)	26 (21.3)
JIA[b] category, n (%)			
Systemic	5 (2.3)	1 (1)	4 (3.1)
Oligoarthritis	47 (21.5)	21 (24)	26 (19.8)
Oligoarthritis—extended	24 (10.9)	9 (10)	15 (11.5)
Polyarthritis (RF-)	50 (22.8)	23 (26)	27 (20.6)
Polyarthritis (RF+)	19 (8.7)	5 (6)	14 (10.7)
Psoriatic arthritis	23 (10.5)	9 (10)	14 (10.7)
Enthesitis-related arthritis	35 (16.0)	13 (15)	22 (16.7)
Undifferentiated	9 (4.1)	3 (3)	6 (4.6)
Other	7 (3.2)	4 (5)	3 (2.3)
Disease severity, n (%)			
Low (0-3 PGA)[c]	181 (82.6)	75 (85)	106 (80.9)
Moderate to severe (4-10 PGA)	38 (17.4)	13 (15)	25 (19)
Duration of illness (years), mean (SD)	5.7 (4.6)	6 (5)	5.6 (4.6)
Expectation about intervention effectiveness at baseline	6.1 (2.1)	6 (2)	5.9 (2.0)

[a]Parent report, N=197.

[b]JIA: juvenile idiopathic arthritis.

[c]PGA: physician global assessment [23].

Adherence to Assigned Study Condition

Telephone coach calls were considered *complete* if a participant had reviewed their assigned website module content for that month. Thus, coach calls were used to define participant adherence to their assigned condition. In total, 72.4% (241/333) of participants completed at least two of three coach calls. Across all participants, coach call completion was as follows: 0 calls (75/333, 22.5%); 1 call (17/333, 5.1%); 2 calls (4/333, 1.2%); or 3 calls (237/333, 71.2%). A higher proportion of participants in the control condition met the minimum criteria for being considered adherent compared with the intervention condition (148/169, 87.6% vs 93/164, 56.7%). Overall, 146/169 (86.4%) control participants and 91/164 (55.5%) intervention participants completed all 3 calls.

Primary and Secondary Outcomes: Adolescents

There was a significant effect of condition on pain intensity (P=.02) and pain interference (P=.007) after adjusting for baseline differences between the groups, with improved scores (lower values) observed in the *Teens Taking Charge* group. The effect of condition was stable over time (from 3 to 12 months), as there were no significant conditions by time interactions. Similar results were seen for the HRQL domains of problems with pain (P=.02) and problems with daily activities (P=.01). Although there was no overall condition effect, there was a

XSL•FO
RenderX

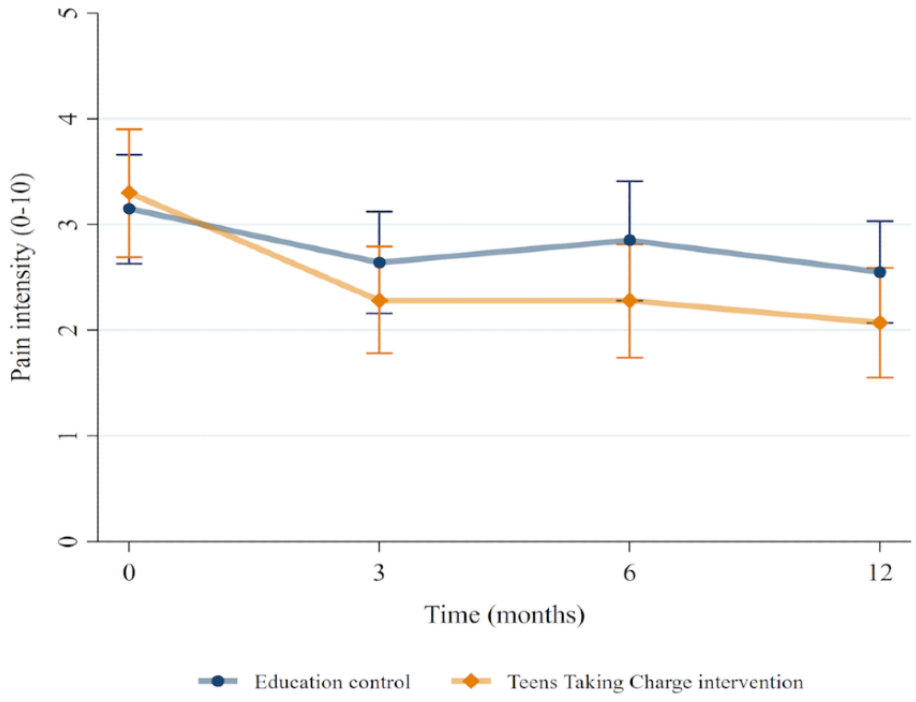
significant condition by time interaction (P=.008) for treatment problems, suggesting that the difference between the groups changed over time. No significant differences were seen on the PedsQL subscales of worry or communication problems. Results from the linear mixed models are presented in Table 3, and the mean values in both groups over time are depicted in Figures 4 to 6. See Multimedia Appendices 3-5 for additional analytic outputs from the adolescent analyses.

Table 3. Adolescent reported primary outcomes, linear mixed models.

Outcome measure	Linear mixed model					
	Time		Condition		Time by condition	
	F value (*ddf*[a]=281)	*P* value	*F* value (*ddf*=281)	*P* value	*F* value (*ddf*=281)	*P* value
Pain						
Pain intensity	1.12	.33	5.44	.02	0.49	.61
Pain interference	1.23	.30	7.40	.007	0.40	.67
Quality of life						
Problems with pain	3.68	.03	5.40	.02	0.50	.61
Problems with daily activities	0.12	.89	6.39	.01	0.19	.83
Treatment problems	4.30	.02	0.12	.73	4.94	.008
Worry	2.64	.07	0.43	.51	0.20	.82
Communication problems	2.93	.06	1.49	.22	0.42	.66

[a]ddf: denominator degrees of freedom.

Figure 4. Pain intensity scores over time in each treatment group, mean and 95% CI.



XSL•FO
RenderX

Figure 5. Pain interference scores over time in each treatment group, mean and 95% CI.

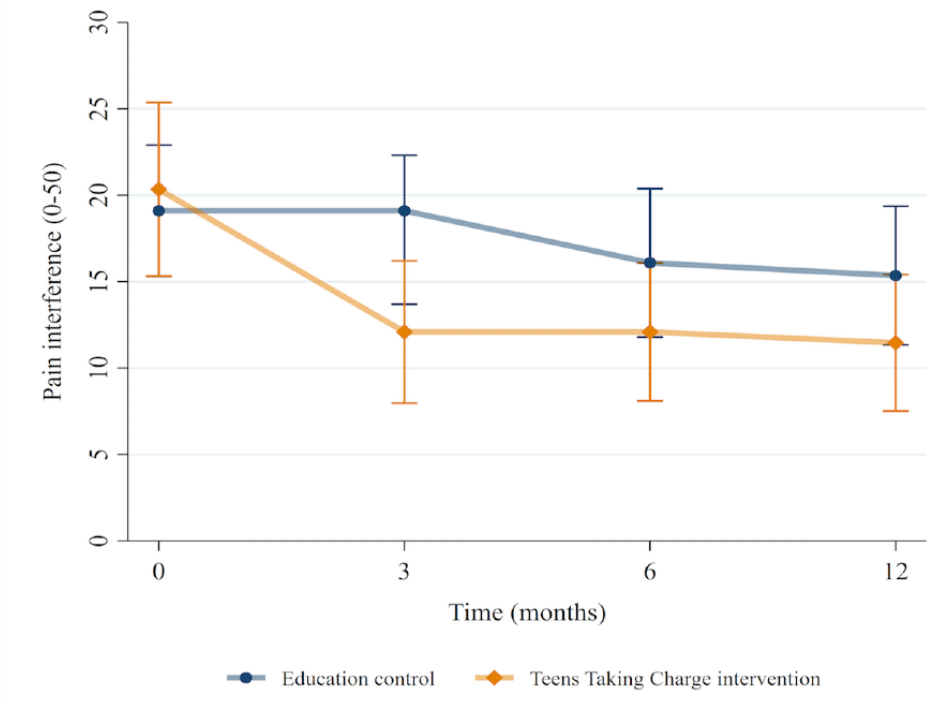
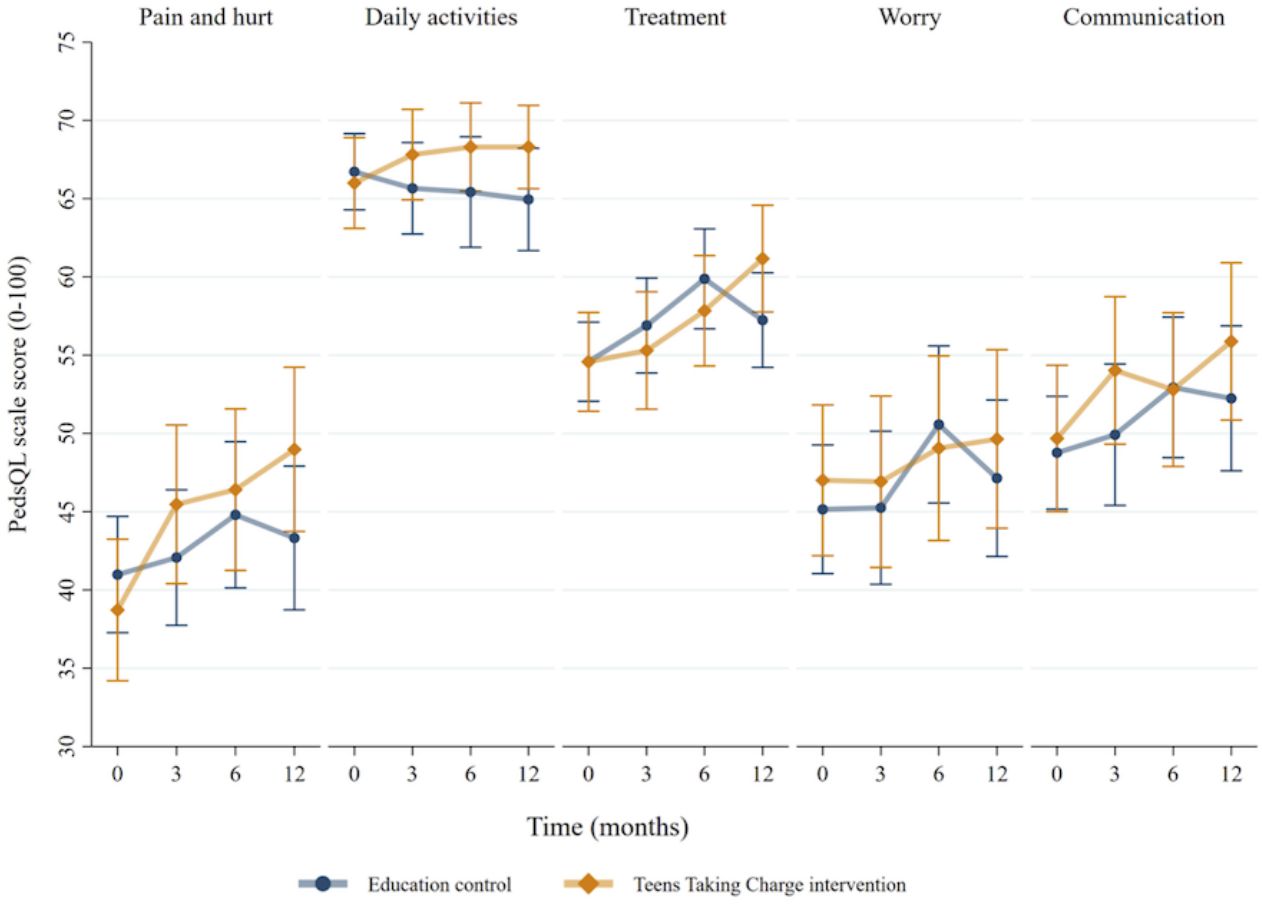


Figure 6. PedsQL subscale scores over time in each treatment group, mean and 95% CI.



Primary and Secondary Outcomes: Caregivers

The primary and secondary caregiver outcome variables as a function of condition and assessment time point are displayed in Multimedia Appendices 6 and 7. There were no significant differences between study groups on primary and secondary caregiver outcomes.

XSL•FO
RenderX

Discussion

Principal Findings

This study sought to evaluate the effectiveness of the *Teens Taking Charge* intervention compared with Web-based education in improving symptoms and HRQL in adolescents with JIA. Our results indicate that the *Teens Taking Charge* self-management intervention has evidence of effectiveness at improving pain intensity and pain interference. These improvements were sustained for up to 12 months postrandomization. The intervention group also demonstrated significant improvements in HRQL related to pain and hurt, daily activities, and treatment problems compared with the control. There were improvements, yet no significant group differences in the secondary outcomes of emotional symptoms, treatment adherence, pain coping, JIA knowledge, and self-efficacy. The data trends support the conceptual model for how JIA self-management education could be foundational to promote improvement in health outcomes.

Comparison With Previous Work

As expected, there is concordance between the results of this trial and the *Teens Taking Charge* pilot RCT. The pilot trial demonstrated initial program effectiveness in improving JIA knowledge and decreasing pain intensity compared with education control [23].

Results of this RCT should also be compared with published results by Connelly et al [44], who evaluated an adapted version of *Teens Taking Charge* in English and Spanish-speaking adolescents with JIA in the United States. Data from the Connelly trial indicated that participants in both study groups had comparable and statistically significant improvements in pain intensity, pain interference, and HRQL over the study period, with no significant between-group differences. The effect size estimates associated with outcome improvement were small.

The findings of these two RCTs should be considered in the context of several factors. First, this study was conducted in Canada while the Connelly study was conducted in the United States. Given that the *Teens Taking Charge* program was delivered in addition to usual care, it is possible that regional differences in baseline care affected the relative impact of the supplemental intervention. Although both RCTs demonstrated improvements in primary outcomes in the control and intervention groups, only this study found significant improvements in the intervention participants compared with control participants. Connelly hypothesized that the choice to compare a self-directed Web-based self-management intervention with an *active* control arm (ie, online education) rather than a wait list or standard care condition may have contributed to the lack of between-group differences in his trial. Participants may have engaged differently with the *active control* condition—with participants in the American trial benefiting more from the provided JIA education and/or attention from a telephone coach. This may be a function of the type of education routinely offered as part of usual clinic care and/or how much participants actually viewed and used the *active control* content. Unfortunately, as data about how often control participants perused the educational websites were not captured, direct comparisons between the trials in this regard is not possible.

This trial's results also concur with two systematic reviews, which found that remotely delivered self-management interventions can be effective in improving painful symptoms in children and adolescents [45,46]. However, a meta-analysis of pediatric chronic pain studies [46], which used an active comparator condition rather than usual care, found that Web-based cognitive behavioral therapy was not significantly better than active control (pooled g=0.10; 95% CI −0.32 to 0.52; P=.64). Although this study did demonstrate significant between-group differences with an active comparator group, our findings should be considered in the context of previous similar studies. Of the three active control studies included in the meta-analysis study [46], the trial by Palermo et al [47] had the largest sample size (N=273 vs N=18 and N=48) and the highest Moncrieff quality rating [48] (37 vs 16 and 26, respectively). The Palermo trial evaluated a program called Web-based Management of Adolescent Pain (Web-MAP2). Their data demonstrated that, from baseline to 6-month follow-up, adolescents with chronic pain who received the Web-MAP2 intervention achieved greater reductions in the primary outcome of daily activity limitations than the education control group. However, while both groups improved, there was no significant difference between groups from baseline to the immediate posttreatment period. Most participants in the Palermo trial reported a baseline pain intensity of moderate-to-severe intensity, while most participants in this trial (181/219, 82.6%) had low disease severity at baseline and low pain intensity (see Table 2). As they started their trial with higher dysfunction, intervention participants in the Palermo trial may have needed more time to consolidate and apply their learned self-management skills, which may have contributed to a lack of differentiation in initial pain outcomes compared with control group participants who received pain education alone.

Overall, this study demonstrates that a Web-delivered cognitive behavioral intervention can significantly improve health outcomes in adolescents with JIA, compared with active control, and that these improvements can be sustained from the immediate posttreatment period for up to 12 months.

Trial Strengths

The extended follow-up period over 1 year allowed for the examination of the maintenance of treatment effects. This is an important aspect of trial quality to address in pediatric Web-based self-management interventions, given that most previous studies have only reported outcomes in the immediate posttreatment period. The use of an education control condition is also an important contribution as most pediatric Web-based RCTs have employed usual care or wait-list control conditions, making it difficult to separate the effects of active treatment from those of increased attention and access to an online program.

XSL•FO
RenderX

Trial Limitation

In this study, telephone coach calls were used as a proxy for program adherence. The website platform, which was built and supported by hospital IT infrastructure, did not support the capture of user-level usage analytics at the time of development. Thus, the impact of program usage level on outcomes could not be examined.

As per the sample size calculation, this trial aimed to analyze a minimum of 117 participants per group. The study analytic sample included 88 intervention and 131 control participants. This loss of participants reduced our statistical power to detect small differences between the study groups. Although we did identify significant group differences on primary outcomes owing to sufficiently large effect sizes, it is possible that smaller effect sizes on other outcomes were missed owing to the underpowered sample.

The higher rate of loss to follow-up in the intervention vs control group may have been the result of increased participant burden (eg, greater time commitment required to review intervention content vs control). Intervention group participants were also less likely to complete all coach telephone coach calls than those in the control group. In light of the high rate of loss to follow-up and lack of data on program usage, the generalizability of study findings should be interpreted with caution.

Significance

This study shows the beneficial effects of a digital approach to disease self-management for the youth with JIA. This paradigm could have far-reaching implications for self-management of other chronic health conditions experienced by the youth. Given that the *Teens Taking Charge* program has been found effective, we envision that it should be offered as a resource for all youths with JIA to supplement their medical care. We believe that this adjunctive self-guided online program will help to overcome current barriers that prevent adolescents with JIA and their families from receiving adequate support in disease self-management.

Conclusions and Next Steps

The results of this RCT indicate that the *Teens Taking Charge* Web-based intervention is effective at reducing both pain intensity and pain interference and improving HRQL in adolescents with JIA compared with education control. These effects are sustained for up to 12 months following program completion. The *Teens Taking Charge* program is now publicly available at no cost [49]. Efforts are underway to widely disseminate this resource to increase the accessibility of self-management care for adolescents and families living with JIA. Efforts are also underway to culturally adapt the intervention for use in different settings.

Acknowledgments

The authors thank the members of the DSMB: Dr Susmita Kashikar-Zuck (Cincinnati Children's Hospital), Dr Chandima Karunanayake (University of Saskatchewan), and Dr Ilona Szer (Rady Children's Hospital). Research coordination by Stephanie Luca and Carol Chan (the Hospital for Sick Children), Michele Gibbon (Children's Hospital of Eastern Ontario), Angelyne Sarmiento and Audrea Chen (BC Children's Hospital), Aleasha Warner, Cindy Campbell, Victoria Dempsey (IWK Health Centre), Jeannine Schellenberg and Shirley Frederickson (Winnipeg Children's Hospital), Ronda Danchak (Stollery Children's Hospital), Mary Gardner, Karen Murphy, Rana Aslanov (Janeway Children's Health and Rehabilitation Centre), Dorothy McKelvey and Joanna Mazza (Montreal Children's Hospital), Dwaraka Veeramreddy (Alberta Children's Hospital), Lois Saunders (McMaster Children's Hospital), and Joan Dietz (Royal University Hospital). The telephone coaches were Lidia Giosa and Maria Fernanda Frances (The Hospital for Sick Children). Dr Mark A Connelly (Children's Mercy Kansas City) provided a helpful manuscript review. Mallika Makkar (The Hospital for Sick Children) aided with manuscript submission. The authors thank the families who took part in this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Publicly accessible arthritis education websites provided in the "JIA Resource Centre" for education control participants.
[DOCX File , 13 KB - jmir_v22i7e16234_app1.docx]

Multimedia Appendix 2
Measures.
[DOCX File , 19 KB - jmir_v22i7e16234_app2.docx]

Multimedia Appendix 3
Adolescent reported primary outcomes over time by condition.
[DOCX File , 16 KB - jmir_v22i7e16234_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Adolescent reported secondary outcomes over time by condition.
[DOCX File , 15 KB - jmir_v22i7e16234_app4.docx]

Multimedia Appendix 5
Adolescent reported secondary outcomes, linear mixed models.
[DOCX File , 13 KB - jmir_v22i7e16234_app5.docx]

Multimedia Appendix 6
Parent-reported outcomes over time by condition.
[DOCX File , 17 KB - jmir_v22i7e16234_app6.docx]

Multimedia Appendix 7
Parent reported outcomes, linear mixed models.
[DOCX File , 14 KB - jmir_v22i7e16234_app7.docx]

Multimedia Appendix 8
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3623 KB - jmir_v22i7e16234_app8.pdf]

References

1. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: a systematic review. Joint Bone Spine 2014 Mar;81(2):112-117. [doi: 10.1016/j.jbspin.2013.09.003] [Medline: 24210707]
2. Shaw KL, Southwood TR, Duffy CM, McDonagh JE. Health-related quality of life in adolescents with juvenile idiopathic arthritis. Arthritis Rheum 2006 Apr 15;55(2):199-207 [FREE Full text] [doi: 10.1002/art.21852] [Medline: 16583399]
3. Ens A, Lang B, Ramsey S, Stringer E, Huber AM. The financial burden of juvenile idiopathic arthritis: a Nova Scotia experience. Pediatr Rheumatol Online J 2013 May 29;11(1):24 [FREE Full text] [doi: 10.1186/1546-0096-11-24] [Medline: 23714134]
4. Bernatsky S, Duffy C, Malleson P, Feldman DE, St Pierre Y, Clarke AE. Economic impact of juvenile idiopathic arthritis. Arthritis Rheum 2007 Feb 15;57(1):44-48 [FREE Full text] [doi: 10.1002/art.22463] [Medline: 17266088]
5. Gutiérrez-Suárez R, Pistorio A, Cruz AC, Norambuena X, Flato B, Rumba I, Pediatric Rheumatology International Trials Organisation (PRINTO). Health-related quality of life of patients with juvenile idiopathic arthritis coming from 3 different geographic areas. The PRINTO multinational quality of life cohort study. Rheumatology (Oxford) 2007 Feb;46(2):314-320. [doi: 10.1093/rheumatology/kel218] [Medline: 16877459]
6. Haverman L, Grootenhuis MA, van den Berg JM, van Veenendaal M, Dolman KM, Swart JF, et al. Predictors of health-related quality of life in children and adolescents with juvenile idiopathic arthritis: results from a web-based survey. Arthritis Care Res (Hoboken) 2012 May;64(5):694-703 [FREE Full text] [doi: 10.1002/acr.21609] [Medline: 22238240]
7. Tong A, Jones J, Craig JC, Singh-Grewal D. Children's experiences of living with juvenile idiopathic arthritis: a thematic synthesis of qualitative studies. Arthritis Care Res (Hoboken) 2012 Sep;64(9):1392-1404 [FREE Full text] [doi: 10.1002/acr.21695] [Medline: 22504867]
8. Bromberg MH, Connelly M, Anthony KK, Gil KM, Schanberg LE. Self-reported pain and disease symptoms persist in juvenile idiopathic arthritis despite treatment advances: an electronic diary study. Arthritis Rheumatol 2014 Feb;66(2):462-469 [FREE Full text] [doi: 10.1002/art.38223] [Medline: 24504820]
9. Rashid A, Cordingley L, Carrasco R, Foster HE, Baildam EM, Chieng A, et al. Patterns of pain over time among children with juvenile idiopathic arthritis. Arch Dis Child 2018 May;103(5):437-443 [FREE Full text] [doi: 10.1136/archdischild-2017-313337] [Medline: 29175824]
10. April KT, Feldman DE, Zunzunegui MV, Duffy CM. Association between perceived treatment adherence and health-related quality of life in children with juvenile idiopathic arthritis: perspectives of both parents and children. Patient Prefer Adherence 2008 Feb 2;2:121-128 [FREE Full text] [Medline: 19920952]
11. Feldman DE, de Civita M, Dobkin PL, Malleson PN, Meshefedjian G, Duffy CM. Effects of adherence to treatment on short-term outcomes in children with juvenile idiopathic arthritis. Arthritis Rheum 2007 Aug 15;57(6):905-912 [FREE Full text] [doi: 10.1002/art.22907] [Medline: 17665485]
12. Lavigne JV, Ross CK, Berry SL, Hayford JR, Pachman LM. Evaluation of a psychological treatment package for treating pain in juvenile rheumatoid arthritis. Arthritis Care Res 1992 Jun;5(2):101-110. [doi: 10.1002/art.1790050209] [Medline: 1390962]
13. Barlow J, Shaw KL, Southwood TR. Do psychosocial interventions have a role to play in paediatric rheumatology? Br J Rheumatol 1998 May;37(5):573-578. [doi: 10.1093/rheumatology/37.5.573] [Medline: 9651088]

XSL•FO
RenderX

14. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
15. Marks R, Allegrante JP, Lorig K. A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part I). Health Promot Pract 2005 Jan;6(1):37-43. [doi: 10.1177/1524839904266790] [Medline: 15574526]
16. Lineker S, Badley E, Dalby D. Unmet service needs of children with rheumatic diseases and their parents in a metropolitan area. J Rheumatol 1996 Jun;23(6):1054-1058. [Medline: 8782140]
17. Elgar FJ, McGrath PJ. Self-administered psychosocial treatments for children and families. J Clin Psychol 2003 Mar;59(3):321-339. [doi: 10.1002/jclp.10132] [Medline: 12579548]
18. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: In review, in use, and into the future. Prof Psychol Res Pract 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
19. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
20. Drotar D, Greenley R, Hoff A, Johnson C, Lewandowski A, Moore M, et al. Summary of issues and challenges in the use of new technologies in clinical care and with children and adolescents with chronic illness. Child Health Care 2006;35(1):91-102. [doi: 10.1207/s15326888chc3501_8]
21. Stinson JN, Toomey PC, Stevens BJ, Kagan S, Duffy CM, Huber A, et al. Asking the experts: exploring the self-management needs of adolescents with arthritis. Arthritis Rheum 2008 Jan 15;59(1):65-72 [FREE Full text] [doi: 10.1002/art.23244] [Medline: 18163408]
22. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010 Jul 29;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
23. Stinson JN, McGrath PJ, Hodnett ED, Feldman BM, Duffy CM, Huber AM, et al. An internet-based self-management program with telephone support for adolescents with arthritis: a pilot randomized controlled trial. J Rheumatol 2010 Sep;37(9):1944-1952. [doi: 10.3899/jrheum.091327] [Medline: 20595280]
24. Murray E, Burns J, See T, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
25. Stinson JN, Tucker L, Huber A, Harris H, Lin C, Cohen L, et al. Surfing for juvenile idiopathic arthritis: perspectives on quality and content of information on the internet. J Rheumatol 2009 Aug;36(8):1755-1762. [doi: 10.3899/jrheum.081010] [Medline: 19531749]
26. Goosens ME, Vlaeyen JW, Hidding A, Kole-Snijders A, Evers SM. Treatment expectancy affects the outcome of cognitive-behavioral interventions in chronic pain. Clin J Pain 2005;21(1):18-26; discussion 69. [doi: 10.1097/00002508-200501000-00003] [Medline: 15599128]
27. Moretti C, Viola S, Pistorio A, Magni-Manzoni S, Ruperto N, Martini A, et al. Relative responsiveness of condition specific and generic health status measures in juvenile idiopathic arthritis. Ann Rheum Dis 2005 Feb;64(2):257-261 [FREE Full text] [doi: 10.1136/ard.2003.016519] [Medline: 15647433]
28. Stinson JN, Stevens BJ, Feldman BM, Streiner D, McGrath PJ, Dupuis A, et al. Construct validity of a multidimensional electronic pain diary for adolescents with arthritis. Pain 2008 Jun;136(3):281-292. [doi: 10.1016/j.pain.2007.07.002] [Medline: 17723279]
29. Varni JW, Seid M, Knight TS, Burwinkle T, Brown J, Szer IS. The PedsQL in pediatric rheumatology: reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory Generic Core Scales and Rheumatology Module. Arthritis Rheum 2002 Mar;46(3):714-725 [FREE Full text] [doi: 10.1002/art.10095] [Medline: 11920407]
30. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai J, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res 2010 May;19(4):595-607 [FREE Full text] [doi: 10.1007/s11136-010-9619-3] [Medline: 20213516]
31. Feldman DE, de Civita M, Dobkin PL, Malleson P, Meshefedjian G, Duffy CM. Perceived adherence to prescribed treatment in juvenile idiopathic arthritis over a one-year period. Arthritis Rheum 2007 Mar 15;57(2):226-233 [FREE Full text] [doi: 10.1002/art.22534] [Medline: 17330298]
32. Reid GJ, Gilbert CA, McGrath PJ. The Pain Coping Questionnaire: preliminary validation. Pain 1998;76(1):83-96. [doi: 10.1016/s0304-3959(98)00029-3]
33. Huguet A, Miró J, Nieto R. The factor structure and factorial invariance of the Pain-Coping Questionnaire across age: evidence from community-based samples of children and adults. Eur J Pain 2009 Sep;13(8):879-889. [doi: 10.1016/j.ejpain.2008.10.004] [Medline: 19084438]
34. André M, Hedengren E, Hagelberg S, Stenström CH. Perceived ability to manage juvenile chronic arthritis among adolescents and parents: development of a questionnaire to assess medical issues, exercise, pain, and social support. Arthritis Care Res 1999 Aug;12(4):229-237. [doi: 10.1002/1529-0131(199908)12:4<229::aid-art1>3.0.co;2-j] [Medline: 10689987]

XSL•FO
RenderX

35. Barlow JH, Shaw KL, Wright CC. Development and preliminary validation of a children's arthritis self-efficacy scale. Arthritis Rheum 2001 Apr;45(2):159-166 [FREE Full text] [doi: 10.1002/1529-0131(200104)45:2<159::AID-ANR169>3.0.CO;2-2] [Medline: 11324780]
36. Barlow JH, Shaw KL, Wright CC. Development and preliminary validation of a self-efficacy measure for use among parents of children with juvenile idiopathic arthritis. Arthritis Care Res 2000 Aug;13(4):227-236. [doi: 10.1002/1529-0131(200008)13:4<227::aid-anr7>3.0.co;2-n] [Medline: 14635277]
37. Cepeda MS, Carr DB. Clinical meaning of a decline in pain intensity in children and its implications in care and research. Pediatric Pain Letter 2006 Dec;8(3):1-5 [FREE Full text]
38. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009 Jun;34(5):495-510. [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
39. SAS Support: SAS Customer Support Site. Cary, North Carolina: SAS Institute Inc; 2006. SAS 9.1.3 Documentation URL: https://support.sas.com/documentation/installcenter/913/index.html [accessed 2020-01-01]
40. Brown H, Prescott R. Applied Mixed Models in Medicine. New York, New York: John Wiley & Sons, Inc; 1999.
41. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag; 2000.
42. Simpson HB, Petkova E, Cheng J, Huppert J, Foa E, Liebowitz MR. Statistical choices can affect inferences about treatment efficacy: a case study from obsessive-compulsive disorder research. J Psychiatr Res 2008 Jul;42(8):631-638 [FREE Full text] [doi: 10.1016/j.jpsychires.2007.07.012] [Medline: 17892885]
43. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
44. Connelly M, Schanberg LE, Ardoin S, Blakley M, Carrasco R, Chira P, et al. Multisite randomized clinical trial evaluating an online self-management program for adolescents with juvenile idiopathic arthritis. J Pediatr Psychol 2019 Apr 1;44(3):363-374. [doi: 10.1093/jpepsy/jsy066] [Medline: 30204919]
45. Fisher E, Law E, Palermo T, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2014 May 14;2014:- [FREE Full text] [doi: 10.1002/14651858.CD011118] [Medline: 25221436]
46. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
47. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
48. Moncrieff J, Churchill R, Drummond DC, McGuire H. Development of a quality assessment instrument for trials of treatments for depression and neurosis. Int J Method Psychiat Res 2001 Aug;10(3):126-133. [doi: 10.1002/mpr.108]
49. AboutKidsHealth for Teens.: The Hospital for Sick Children; 2020. The Taking Charge: Managing JIA Online Program URL: https://teens.aboutkidshealth.ca/jiateenhub [accessed 2020-03-12]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DSMB: Data Safety and Monitoring Board
HRQL: health-related quality of life
JIA: juvenile idiopathic arthritis
RCT: randomized controlled trial
Web-MAP2: Web-based Management of Adolescent Pain

Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by M Bonnert, A Pasaoglu Bozkurt; comments to author 28.10.19; revised version received 03.01.20; accepted 03.02.20; published 29.07.20.

Please cite as:
Stinson JN, Lalloo C, Hundert AS, Campillo S, Cellucci T, Dancey P, Duffy C, Ellsworth J, Feldman BM, Huber AM, Johnson N, Jong G, Oen K, Rosenberg AM, Shiff NJ, Spiegel L, Tse SML, Tucker L, Victor JC
Teens Taking Charge: A Randomized Controlled Trial of a Web-Based Self-Management Program With Telephone Support for Adolescents With Juvenile Idiopathic Arthritis
J Med Internet Res 2020;22(7):e16234
URL: https://www.jmir.org/2020/7/e16234
doi:10.2196/16234
PMID:32723728

XSL•FO
RenderX

©Jennifer N Stinson, Chitra Lalloo, Amos S Hundert, Sarah Campillo, Tania Cellucci, Paul Dancey, Ciaran Duffy, Janet Ellsworth, Brian M Feldman, Adam M Huber, Nicole Johnson, Geert't Jong, Kiem Oen, Alan M Rosenberg, Natalie J Shiff, Lynn Spiegel, Shirley M L Tse, Lori Tucker, Joseph Charles Victor. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing a Personalized Digital Patient Support Program for Patients Treated With Growth Hormone: Key Design Considerations

Sumaira Malik[1], BSc, MSc, PhD; Clare Moloney[1], BSc, MSc, PhD; Ekaterina Koledova[2], MD, PhD; Jonathan Reston[1], BA, MSc, PhD; John Weinman[3], PhD

[1]Atlantis Healthcare, London, United Kingdom

[2]Endocrinology, Global Medical, Safety and CMO, Merck KGaA, Darmstadt, Germany

[3]Institute of Pharmaceutical Science, Kings College London, London, United Kingdom

Corresponding Author:
Sumaira Malik, BSc, MSc, PhD
Atlantis Healthcare
Building 5 Chiswick Park
Chiswick High Road
London, W4 5YA
United Kingdom
Phone: 44 2087474360
Email: sumaira.malik@atlantishealthcare.com

Abstract

Background: Recombinant human growth hormone treatment can optimize growth potential; however, optimal outcomes are not always achieved for several reasons, including poor adherence. The overall objective of this project was to design a patient support program to maximize the chances of treatment success for people being treated with somatropin injection. An approach known as the behavior change wheel was used to enhance the development of the patient support program. The behavior change wheel provided a comprehensive framework to support the design of interventions.

Objective: The aim of this paper was to describe how the steps of the behavior change wheel were applied to the development of a patient support program for individuals with growth hormone deficiency undergoing treatment with somatropin.

Methods: We followed a series of steps that align to tenets of the behavior change wheel, namely, a narrative literature review to identify which behaviors needed to change and the potential drivers of and barriers to the behaviors, the selection of an intervention strategy and discrete behavior change techniques, and, finally, intervention specification.

Results: A recent systematic review identified a range of potentially modifiable factors found to have an influence on patient adherence to growth hormone treatment. Insights from the systematic review were used to guide the development of a patient support program. The final design of the patient support program consisted of four elements: (1) a personalization questionnaire to tailor support for each individual, (2) tailored reminder and support SMS text messages, (3) nurse-led phone calls, and (4) Easypod connect, an automated electronic autoinjector drug-delivery device with a transmitter and connection platform for Saizen (somatropin) that allows automatic recording, storage, and transmission of drug-usage data, thus providing insight into suboptimal adherence.

Conclusions: The patient support program that was designed is currently being piloted with patients to assess engagement with the program and determine its impact on patient outcomes. Results from the pilot will be used to further refine the program to ensure it meets user needs.

(J Med Internet Res 2020;22(7):e18157) doi:10.2196/18157

KEYWORDS

growth hormone; behavior change wheel; somatropin; adherence

XSL•FO
RenderX

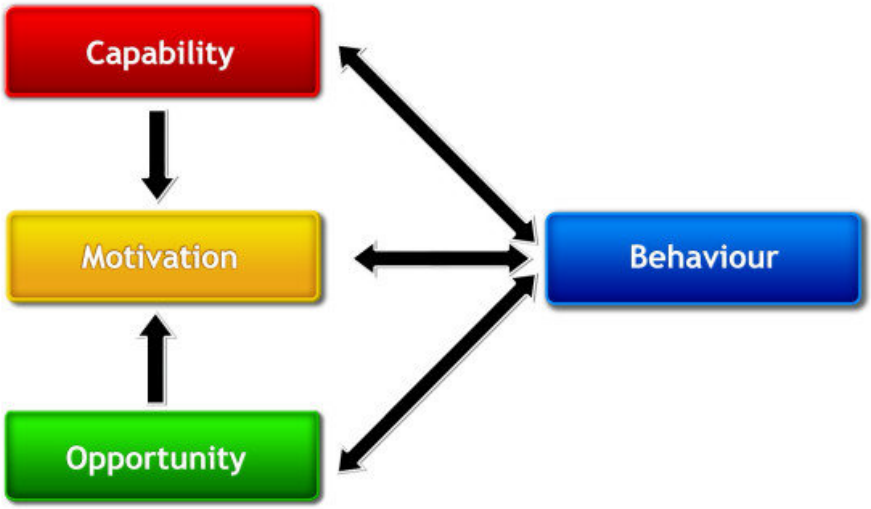
Introduction

Recombinant human growth hormone (r-hGH; somatropin) treatment is well established for children and adolescents with growth hormone deficiency [1]. The goal of this treatment is to optimize growth potential, so that the expected target height (based on a normal growth curve for each individual) is achieved by adulthood [2]. Evidence has demonstrated the efficacy of growth hormone treatments to improve growth in children with growth hormone deficiency compared to that of children whose growth hormone deficiency is not treated [3].

The impact of growth hormone deficiency goes beyond reduced growth potential. Research has shown that children with growth hormone deficiency may also experience a psychosocial burden as a result of being physically different from their peers [3,4]. This burden can continue into adulthood, particularly if short stature persists, limiting both personal and professional success [3] as well as potentially having an impact on long-term health, since, as adults, they may also have an increased risk of cardiovascular diseases [5]. With prevalence estimates for growth hormone deficiency from 1.8 to 2.9 per 10,000 in Europe and the United States [6-8], the potential of treatment success to benefit the health economy as well as patients and their families is significant [3].

The overall objective of this project was to design a patient support program to maximize the chances of treatment success for people treated with somatropin by injection. An approach known as the behavior change wheel [9] was used to enhance the development of the patient support program. The behavior change wheel provided a comprehensive framework to support the design of the interventions allowing designers to move from behavior analysis to intervention design using evidence-based techniques and behavior change theory. Central to the behavior change wheel is a framework called the capability, opportunity, motivation, and behavior model (COM-B) [9] (Figure 1). The COM-B model proposes that a person's health behaviors are driven by a range of factors, which can be grouped under three broad headings: *capability*, their physical or psychological ability to engage in the behavior, for example, in the context of adherence to medication, factors such as reduced cognitive functioning, mobility problems, or poor treatment comprehension may affect an individual's ability to understand how to physically administer their treatment; *motivation*, the internal thoughts and emotions that influence individual decision making in relation to the behavior, for example, beliefs about the need for treatment or the seriousness of the illness may affect how motivated an individual is to try a new treatment or depression as a result of their health condition can reduce their motivation to adhere to a treatment regimen; and *opportunity*, the external factors which make it possible to engage in the behavior, for example, having a supportive social network and easy access to healthcare resources. In addition to each component influencing medication-related behavior directly, opportunity and capability may affect motivation, and thus also influence behavior in this manner. Jackson and colleagues [10] successfully applied this model to treatment adherence behaviors. While there are similarities in factors for adherence across a patient population, it is important to note that each patient has their own unique needs and set of beliefs that have an impact on their ability to adhere to treatment [11]. Offering personally tailored self-management or adherence interventions to suit their individual needs is the most effective way to ensure patients understand and adopt targeted behaviors, including adherence to r-hGH treatment [11-13]. Tailored interventions and support have also been shown to be more cost-effective [14], an important consideration in the sustainability of support services.

Figure 1. Capability, opportunity, motivation, and behavior framework.



Another crucial component of the behavior change wheel is the selection of appropriate techniques to change behavior. In this paper, we describe how the steps of the behavior change wheel were applied to the development of a patient support program for individuals with growth hormone deficiency who were prescribed somatropin; the duration of individual treatments varies depending on age at time of diagnosis (usually early childhood), and in order to reach full height potential, are continued up to the age when bone growth stops (as the individual reaches adult height). The patient support program was designed to incorporate a large digital component, while also maintaining a strong human support element to personalize

XSL•FO
RenderX

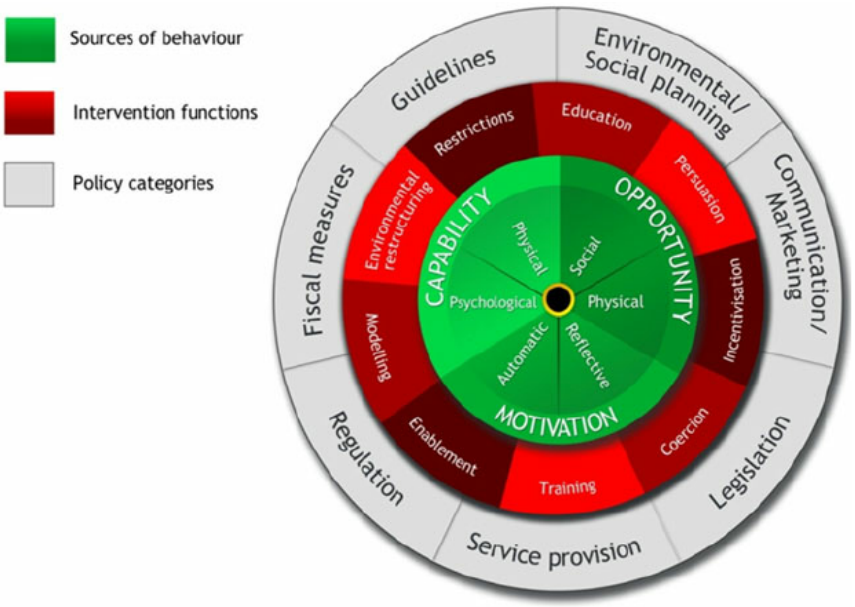
and complement the digital aspects of the intervention. Digital interventions that incorporate a wide range of communication channels such as SMS text messaging, interactive websites, apps, and electronic devices are thought to have great potential for this patient population, given the popularity of digital media among young people [15]; however, despite this, maintaining engagement with digital interventions can be difficult with many studies reporting high levels of dropouts or nonusage [16]. Personalized messaging and multicomponent interventions including both digital and nondigital aspects have been shown to be more effective for behavior change than nontailored messaging and single-component interventions [17]. Yardley et al [16] proposed several general principles and considerations for maximizing engagement with digital behavior-change interventions; they stressed that, most importantly, the intervention should be relevant to the user, ie, it should address an unmet need and should be tailored to the specific situation, needs, and motivations of the user [16]. Utilizing the behavior change wheel and the COM-B framework allowed us to identify the drivers that underlie treatment-related behavior in growth hormone deficiency patients and caregivers, and thus, to create a program with relevant tailored content. Further suggestions [16] included adding a human element to help maintain motivation and ongoing adherence to the digital components of the intervention, ensuring the intervention content was accessible and engaging for people with different levels of health literacy, and taking an iterative approach to intervention development and evaluation to refine the intervention to ensure it would continue to meet user needs. These considerations were also used to guide the design of the patient support program.

Methods

Overview

We followed a series of steps that were aligned to tenets of the behavior change wheel [9], namely, identifying behaviors that needed to change, identifying the potential drivers of and barriers to the behavior, selecting an intervention strategy and discreet behavior change techniques, and specifying an intervention (Figure 2).

Figure 2. Behavior change wheel.



Step 1

A narrative review of the literature was carried out to understand problems associated with treatment management in growth hormone deficiency and to select and specify the target behavior. Literature searches were conducted using CINAHL, Medline, PsychINFO (via EBSCOHost), and the Cochrane Register for Controlled Trials and Systematic Reviews. The following keywords were used: “growth hormone,” “human growth hormone” (MeSH), “child” (MeSH), “adolescent” (MeSH), “patient compliance” (MeSH). Literature was limited to papers written in English and no date restrictions were applied. The reference lists and citations of publications that were identified were also searched for additional papers on the topic.

Step 2

Underlying barriers and drivers were defined using insights from a 2018 systematic review [18] looking at the range of potentially modifiable factors that have been found to influence levels of adherence to growth hormone treatment; these factors were grouped according to the COM-B framework. The findings from this review provided direction for selecting the types of interventions that would be needed to effectively target these factors in digital interventions.

Step 3

Intervention options were identified using the behavior change wheel, which proposes nine intervention functions: education, persuasion, incentivization, coercion, training, restriction, environmental restructuring, modeling, and enablement. We used APEASE criteria (acceptability, practicability, effectiveness or cost-effectiveness, affordability, safety or side effects, and equity) [19] to guide our selection of the most relevant intervention functions. We also engaged a team of specialists in health psychology with expertise in behavior change.

XSL•FO
RenderX

Step 4

Based on the factors identified by Graham et al [18] and the intervention functions that were selected, a list of relevant behavior change techniques was defined to address each COM-B factor.

Step 5

The full intervention was mapped out, which involved selecting delivery channels (SMS text messages, nurse-led calls, Growlink app), applying the appropriate behavior change techniques to develop content, and designing a personalization questionnaire to tailor the patient experience. In addition, a measurement strategy was designed to evaluate the effectiveness of the intervention.

Results

Step 1: Defining the Problem

A narrative review of the literature on growth hormone deficiency treatments found that despite the potential of r-hGH treatment to optimize growth potential, optimal outcomes were not always achieved; it was proposed that poor adherence was a key contributor to this [2,20,21]. The levels of nonadherence that were reported varied, particularly since definitions and assessment methods differed between studies, but with the proportion of nonadherent patients ranging from 5% to 82% across several indications for which growth hormone therapy is used [2] and with 7% to 71% of patients with growth hormone deficiency found to be nonadherent to treatment [18], it was clear that there was room to improve adherence in this population [2]. Adherence to somatropin injections was selected as a target behavior and outcome for the design of the patient support program. Table 1 further describes the target behaviors.

Table 1. Target behavior description.

Target behavior	Adherence to somatropin injection
Who needs to perform the behavior?	Pediatric patients with growth hormone deficiency or their caregiver
Where will they do it?	At home
How often will they do it?	This will vary depending on individual needs and will be determined by the prescription issued by their health care provider

Step 2: Behavior Diagnosis

In work relevant to this study, Graham et al [18] conducted a systematic review of pediatric nonadherence to r-hGH treatment using the COM-B framework to identify potentially modifiable factors. Key factors were identified in the review: within the scope of capability—knowledge and understanding of the condition, a lack of understanding of the consequences of missed r-hGH doses, forgetting to administer the medication, or poor administration technique; within the scope of opportunity—inadequate contact with health care providers and the quality of the health care provider–patient relationship, as well as the discomfort and pain associated with daily injections; and within the scope of motivation—the long duration of treatment and dissatisfaction with growth response results.

These findings were largely reflective of the insights gained from our narrative review of growth hormone and adherence literature; they suggested a need to address a range of factors related to patients receiving growth hormone therapy, their caregivers, and growth hormone deficiency and its treatment to ensure optimal adherence.

Steps 3 and 4: Implementation Strategy and Behavior Change Technique Selection

While all nine of the interventions listed in the behavior change wheel were found to be relevant to nonadherence to treatment in growth hormone deficiency, we selected the five intervention functions that we felt were the most relevant and the most appropriate to address within the patient support program; these were education, persuasion, training, modeling, and enablement. This decision was based on a combination of practical considerations such as affordability and cost-effectiveness as well as on behavior change expertise from a team of specialists in health psychology. Table 2 shows the relationship between the factors identified in the COM-B, intervention functions, and selected behavior change techniques. An expanded version of Table 2 including COM-B categories, intervention functions, and intervention content is available in Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. COM-B adherence factors and behavior change techniques.

Adherence factors	Behavior change techniques
Long duration of treatment	• Framing or reframing
Dissatisfaction with treatment outcome	• Self-monitoring • Framing or reframing • Reduce negative emotion
Knowledge and understanding of condition	• Credible source • Information on health consequences
Discomfort or pain from daily injection	• Instruction of how to perform the behavior • Problem solving • Verbal persuasion of capability • Reduce negative emotion
Lack of understanding of the consequences of missed doses	• Information on health consequence • Salience of consequences
Forgetting	• Feedback on behavior • Self-monitorin • Prompts or cues • Problem solving
Health care provider–patient communication	• Demonstration of the behavior • Problem solving • Verbal persuasion of capability
Poor injection technique	• Instruction on how to perform the behavior • Demonstration of the behavior

Step 5: Mapping the Design of the Patient Support Program

Overview

Based on these findings, as well as on insights from digital behavior change intervention design principles and behavior change theory, a multicomponent digital patient support program was designed for patients receiving treatment with somatropin injection (see Table 2 for example content). The patient support program consists of four components: personalization screening, SMS text messages, nurse-led phone calls, and an eHealth component.

Personalization Screening Questionnaire

A set of personalization questions was created based on the drivers of adherence that were identified in step 2. The design includes different questions for caregivers and patients, allowing for support to be tailored to both or either, depending on who is taking part in the patient support program. Patients and caregivers will be asked to complete the screening questionnaire when they join the patient support program and will be rescreened every 12 weeks. This is to ensure that support can be flexible to changing needs and to allow for review of intervention effectiveness by monitoring changes to beliefs and to support needs. The personalization screening questionnaire will be used to determine topic priority and focus of the SMS text messages and of the nurse-led components of the program.

SMS Text Messages

These include medication reminders as well as tailored intervention messages to address adherence factors that have been identified as important through the personalization screening questionnaire.

Nurse-Led Phone Calls

Patient support program nurses will be trained to deliver brief, evidence-based interventions to address the adherence barriers that were identified. The content and order of these telephone calls will be tailored to the needs of each individual patient or caregiver based on personalization screening questionnaire responses.

eHealth Component

The Easypod is an automated electronic autoinjector drug-delivery device with a transmitter and web-based connection platform for Saizen (somatropin) that automatically inserts a needle and delivers a preset dose to the patient. The device allows automatic recording, storage, and transmission of drug adherence data, thus providing insight into suboptimal adherence (through dose frequency) and its resulting effect on growth. Patients and caregivers can access the information via a patient app (Growlink), and physicians and nurses can access the information via the web-based Easypod connect platform.

XSL•FO
RenderX

Discussion

Principal Findings

The purpose of this paper was to describe the development of a patient support program aimed at supporting patients who have been prescribed Saizen in order to maximize their growth potential. A structured approach based on the behavior change wheel and the COM-B framework was used to identify the drivers of behavior and to systematically explore how these could be addressed through a pharma-funded patient support program.

The final design of the patient support program incorporated a patient-centric digital component featuring an online app linked to the injection device and frequent SMS text message delivery. The digital platform offers a convenient method for patients, caregivers, and health care providers to interact, to share, and to review adherence data for each patient. The addition of SMS text messaging allows adherence reminder messages and intervention content that was designed to address nonadherence to be sent to individual patients. SMS text messages have been successfully used in behavior change interventions as both reminders for appointments or for treatments [22], as well as to deliver interventions to change patient beliefs and to improve adherence [2]. A 2011 meta-analysis [23] of the efficacy of SMS text message reminders showed that they were effective across all age groups; there were also no differences in effect based on the timing of the messages or the rate that the messages were sent [23]. Messaging interventions have been shown to improve adherence to medications across a wide variety of clinical applications including asthma, antiretroviral treatments, and schizophrenia, in both adult and pediatric populations [22,24,25]. The effect of personalized messaging on behavior has also been shown to be more effective than that of nontailored messaging [17]. In line with the recommendations of Yardley et al [16], tailored content may also help increase engagement with digital interventions. In this patient support program, tailored content was achieved through a personalization screening questionnaire designed to determine the content and sequence of SMS text messages and nurse-led phone calls for delivery of the intervention; the online app was also designed to provide personalized adherence feedback.

Personalized telephone calls with a specialist nurse add a human element to the patient support program which may help to support continued engagement. Evidence has shown that nurses can successfully be trained to use behavior change techniques and to implement motivational interviewing techniques by a simple, brief course [26,27]. Nurse-led calls that implement motivational interviewing principles and that teach behavior change techniques have been shown to result in meaningful behavior change [28,29]. In addition, telephone-based support has been shown to be an effective delivery channel for promoting behavior change across different health conditions such as smoking cessation, increasing physical activity, and improving diet [30], as well as demonstrating a positive impact on treatment adherence [31,32]. Research has also shown that, in general, multicomponent interventions are more successful in having an impact on behavior than single-component interventions are [33]. Thus, the nurse-led coaching phone calls are likely to enhance and complement the support provided through the digital components.

Another important consideration for any patient support program is the level of health literacy. Yardley et al [16] argued that digital behavior change interventions should be accessible and engaging for people with low levels of health literacy and should also be acceptable and usable for those with higher levels of health literacy. All content that was developed for the patient support program was reviewed by an experienced health psychologist and writer, and by a creative team to ensure the pitch, tone, and level was appropriate for the range of health literacy levels in the patient group.

Future Work

Our patient support program will be piloted with the aim of establishing its perceived acceptability and usefulness. The evaluation of the pilot patient support program will be conducted using an observational (real-world evidence) within-subjects design. Three categories of data will be captured: operational data, user experience data, and impact data. The primary outcome will be patient perception of the helpfulness of the patient support program. Secondary outcomes will include change in adherence and change in quality of life over the course of the patient support program, and multiple regression modeled predictors of these outcomes. Finally, data will be captured to identify ways in which various components of the patient support program could be improved, in line with the principles of continuous development in patient support program design, behavior change technique, and digital interventions proposed by Michie et al [34]. These data will be captured during routine operation via the patient support program (such as number and length of sessions), supplementary behavior-changing customer relationship management, and follow-up evaluation surveys.

Strengths and Limitations

The patient support program was developed using a systematic and structured approach, by drawing on relevant literature, and using appropriate evidence-based behavior change theories and frameworks. To the best of our knowledge, this represents a novel approach to the development of patient support programs in growth hormone deficiency; however, despite this, the use of the behavior change wheel inevitably required that the intervention development team make subjective and practical decisions regarding the most appropriate strategies and delivery channels. Since there was limited data on the most appropriate intervention channels and implementation of behavior change techniques for this patient population; the team drew from their own experience working in this patient population to guide some of these decisions.

The first part of the design process involved a narrative review of relevant literature. While this review captured a large volume of relevant literature, the methods were not exhaustive. Key search terms were provided, but to ensure that all relevant papers had been captured in an entirely reproducible format, a full systematic review was required. To be efficient, this paper focused, where possible, on meta-analyses and systematic

XSL•FO
RenderX

reviews such as [18] which partially summarized large quantities of the literature.

Conclusions

Using the approach outlined by the behavior change wheel, a multicomponent patient support program for patients with growth hormone deficiency (and their caregivers) to achieve their growth potential through adherence using the Easypod connect platform was developed. By incorporating personalization screening questionnaires into the patient support program, personalized intervention messages and nurse-led phone support can be offered to patients. Work is underway to implement and validate the patient support program, to establish whether it can improve adherence in a real-world setting.

Acknowledgments

This work was funded by Merck KGaA. Editorial support was provided by Jorge Cancela who, at the time of writing, was an employee of Ares Trading (an affiliate of Merck KGaA) and by Steven Goodrick of inScience Communications, Springer Healthcare Ltd.

Conflicts of Interest

JR and SM are employees of Atlantis Healthcare and received funding from Merck KGaA to carry out this work. At the time of writing, CM was an employee of Atlantis Healthcare and received funding from Merck KGaA to carry out this work. JW is a consultant to Atlantis Healthcare and received funding from Merck KGaA to carry out this work. EK is an employee of Merck KGaA.

Multimedia Appendix 1
Expanded Table 2.
[DOC File , 39 KB - jmir_v22i7e18157_app1.doc]

References

1. Al Herbish AS, Al Alwan I, Al Mutair A, Al Twaim A, Al Agha A, Deeb A, et al. Growth hormone therapy and treatment outcomes: current clinical practice of the Gulf Cooperation Council. Expert Rev Endocrinol Metab 2014 Jul;9(4):319-325. [doi: 10.1586/17446651.2014.921115] [Medline: 30763992]
2. Fisher BG, Acerini CL. Understanding the growth hormone therapy adherence paradigm: a systematic review. Horm Res Paediatr 2013;79(4):189-196 [FREE Full text] [doi: 10.1159/000350251] [Medline: 23635797]
3. Takeda A, Cooper K, Bird A, Baxter L, Frampton GK, Gospodarevskaya E, et al. Recombinant human growth hormone for the treatment of growth disorders in children: a systematic review and economic evaluation. Health Technol Assess 2010 Sep;14(42):1-209, iii. [doi: 10.3310/hta14420] [Medline: 20849734]
4. Brod M, Alolga SL, Beck JF, Wilkinson L, Højbjerre L, Rasmussen MH. Understanding burden of illness for child growth hormone deficiency. Qual Life Res 2017 Jul;26(7):1673-1686 [FREE Full text] [doi: 10.1007/s11136-017-1529-1] [Medline: 28247315]
5. Bates AS, Van't Hoff W, Jones PJ, Clayton RN. The effect of hypopituitarism on life expectancy. J Clin Endocrinol Metab 1996 Mar;81(3):1169-1172. [doi: 10.1210/jcem.81.3.8772595] [Medline: 8772595]
6. Lindsay R, Feldkamp M, Harris D, Robertson J, Rallison M. Utah Growth Study: growth standards and the prevalence of growth hormone deficiency. J Pediatr 1994 Jul;125(1):29-35. [doi: 10.1016/s0022-3476(94)70117-2] [Medline: 8021781]
7. Thomas M, Massa G, Craen M, de Zegher F, Bourguignon JP, Heinrichs C, et al. Prevalence and demographic features of childhood growth hormone deficiency in Belgium during the period 1986-2001. Eur J Endocrinol 2004 Jul;151(1):67-72. [doi: 10.1530/eje.0.1510067] [Medline: 15248824]
8. Vimpani GV, Vimpani AF, Lidgard GP, Cameron EH, Farquhar JW. Prevalence of severe growth hormone deficiency. Br Med J 1977 Aug 13;2(6084):427-430 [FREE Full text] [doi: 10.1136/bmj.2.6084.427] [Medline: 890325]
9. Michie S, van SMM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
10. Jackson C, Eliasson L, Barber N, Weinman J. Applying COM-B to medication adherence: a suggested framework for research and interventions. Eur Health Psychol 2014;16(1):7-17 [FREE Full text]
11. Haverkamp F, Gasteyger C. A review of biopsychosocial strategies to prevent and overcome early-recognized poor adherence in growth hormone therapy of children. J Med Econ 2011;14(4):448-457. [doi: 10.3111/13696998.2011.590829] [Medline: 21651428]
12. Cockle-Hearne J, Faithfull S. Self-management for men surviving prostate cancer: a review of behavioural and psychosocial interventions to understand what strategies can work, for whom and in what circumstances. Psychooncology 2010 Sep;19(9):909-922. [doi: 10.1002/pon.1657] [Medline: 20119934]
13. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [Medline: 12867348]

XSL•FO
RenderX

14. Radhakrishnan K. The efficacy of tailored interventions for self-management outcomes of type 2 diabetes, hypertension or heart disease: a systematic review. J Adv Nurs 2012 Mar;68(3):496-510. [doi: 10.1111/j.1365-2648.2011.05860.x] [Medline: 22010971]
15. Child J, Davies C, Frost K, McDermid E, Pidcock R, Weinman J, et al. Managing paediatric growth disorders: integrating technology into a personalised approach. J Clin Res Pediatr Endocrinol 2019 Nov 20 [FREE Full text] [doi: 10.4274/jcrpe.galenos.2019.2019.0153] [Medline: 31744271]
16. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
17. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
18. Graham S, Weinman J, Auyeung V. Identifying potentially modifiable factors associated with treatment non-adherence in paediatric growth hormone deficiency: a systematic review. Horm Res Paediatr 2018;90(4):221-227 [FREE Full text] [doi: 10.1159/000493211] [Medline: 30522126]
19. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Intervention. London: Silverback Publishing; 2014:1-329.
20. Aydın BK, Aycan Z, Sıklar Z, Berberoğlu M, Ocal G, Cetinkaya S, et al. Adherence to growth hormone therapy: results of a multicenter study. Endocr Pract 2014 Jan;20(1):46-51. [doi: 10.4158/EP13194.OR] [Medline: 24013997]
21. Cutfield WS, Derraik JGB, Gunn AJ, Reid K, Delany T, Robinson E, et al. Non-compliance with growth hormone treatment in children is common and impairs linear growth. PLoS One 2011 Jan 31;6(1):e16223 [FREE Full text] [doi: 10.1371/journal.pone.0016223] [Medline: 21305004]
22. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
23. Boksmati N, Butler-Henderson K, Anderson K, Sahama T. The effectiveness of SMS reminders on appointment attendance: a meta-analysis. J Med Syst 2016 Apr;40(4):90. [doi: 10.1007/s10916-016-0452-2] [Medline: 26852337]
24. Lau PWC, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
25. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
26. Madson MB, Loignon AC, Lane C. Training in motivational interviewing: a systematic review. J Subst Abuse Treat 2009 Jan;36(1):101-109. [doi: 10.1016/j.jsat.2008.05.005] [Medline: 18657936]
27. Mannix KA, Blackburn IM, Garland A, Gracie J, Moorey S, Reid B, et al. Effectiveness of brief training in cognitive behaviour therapy techniques for palliative care practitioners. Palliat Med 2006 Sep;20(6):579-584. [doi: 10.1177/0269216306071058] [Medline: 17060250]
28. Cook PF, Emiliozzi S, El-Hajj D, McCabe MM. Telephone nurse counseling for medication adherence in ulcerative colitis: a preliminary study. Patient Educ Couns 2010 Nov;81(2):182-186. [doi: 10.1016/j.pec.2009.12.010] [Medline: 20079598]
29. McBride CM, Rimer BK. Using the telephone to improve health behavior and health service delivery. Patient Educ Couns 1999 May;37(1):3-18. [Medline: 10640115]
30. Bricker JB, Mann SL, Marek PM, Liu J, Peterson AV. Telephone-delivered acceptance and commitment therapy for adult smoking cessation: a feasibility study. Nicotine Tob Res 2010 Apr;12(4):454-458. [doi: 10.1093/ntr/ntq002] [Medline: 20142417]
31. Mohr DC, Likosky W, Bertagnolli A, Goodkin DE, Van DWJ, Dwyer P, et al. Telephone-administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. J Consult Clin Psychol 2000 Apr;68(2):356-361. [Medline: 10780138]
32. Turner AP, Sloan AP, Kivlahan DR, Haselkorn JK. Telephone counseling and home telehealth monitoring to improve medication adherence: results of a pilot trial among individuals with multiple sclerosis. Rehabil Psychol 2014 May;59(2):136-146. [doi: 10.1037/a0036322] [Medline: 24708232]
33. Del Pino-Sedeño T, Peñate W, de Las Cuevas C, Valcarcel-Nazco C, Fumero A, Serrano-Pérez PG, et al. Effectiveness and cost-effectiveness of a multicomponent intervention to improve medication adherence in people with depressive disorders - MAPDep: a study protocol for a cluster randomized controlled trial. Patient Prefer Adherence 2019;13:309-319 [FREE Full text] [doi: 10.2147/PPA.S172963] [Medline: 30863020]
34. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

XSL•FO
RenderX

Abbreviations

COM-B: capability, opportunity, motivation, and behavior model
MeSH: medical subject heading (index term)
r-hGH: recombinant human growth hormone

Edited by G Eysenbach; submitted 07.02.20; peer-reviewed by S Graham, K Curtis; comments to author 02.04.20; accepted 14.05.20; published 29.07.20.

Please cite as:
Malik S, Moloney C, Koledova E, Reston J, Weinman J
Designing a Personalized Digital Patient Support Program for Patients Treated With Growth Hormone: Key Design Considerations
J Med Internet Res 2020;22(7):e18157
URL: http://www.jmir.org/2020/7/e18157/
doi:10.2196/18157
PMID:32723712

©Sumaira Malik, Clare Moloney, Ekaterina Koledova, Jonathan Reston, John Weinman. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mobile Health Approach for Improving Outcomes in Suicide Prevention (SafePlan)

Conor O'Grady[1], MSc; Ruth Melia[2], PhD; John Bogue[3], PhD; Mary O'Sullivan[4], MA; Karen Young[1], MA; Jim Duggan[1], PhD

[1]School of Computer Science, National University of Ireland Galway, Galway, Ireland

[2]Health Service Executive, Ennis, Ireland

[3]School of Psychology, National University of Ireland Galway, Galway, Ireland

[4]HSE Community Healthcare West, Galway, Ireland

Corresponding Author:
Jim Duggan, PhD
School of Computer Science
National University of Ireland Galway
University Road
Galway, H91 TK33
Ireland
Phone: 353 91493336
Email: james.duggan@nuigalway.ie

Abstract

Background: Suicide is a prominent cause of death worldwide, particularly among young people. It was the second leading cause of death among those aged 15-29 years globally in 2016. Treatment for patients with suicidal thoughts or behaviors often includes face-to-face psychological therapy with a mental health professional. These forms of interventions may involve maintaining and updating paper-based reports or worksheets in between sessions. Mobile technology can offer a way to support the implementation of evidence-based psychological techniques and the acquisition of protective coping skills.

Objective: This study aims to develop a mobile app to facilitate service users' access to mental health support and safety planning. This process involved eliciting expert input from clinicians who are actively engaged in the provision of mental health care.

Methods: A survey was distributed to targeted health care professionals to determine what features should be prioritized in a new mobile app relating to suicide prevention. On the basis of the survey results, a clinical design group, comprising 6 members with experience in fields such as mobile health (mHealth), clinical psychology, and suicide prevention, was established. This group was supplemented with further input from additional clinicians who provided feedback over three focus group sessions. The sessions were centered on refining existing app components and evaluating new feature requests. This process was iterated through regular feedback until agreement was reached on the overall app design and functionality.

Results: A fully functional mobile app, known as the SafePlan app, was developed and tested with the input of clinicians through an iterative design process. The app's core function is to provide an interactive safety plan to support users with suicidal thoughts or behaviors as an adjunct to face-to-face therapy. A diary component that facilitates the generalization of skills learned through dialectical behavior therapy was also implemented. Usability testing was carried out on the final prototype by students from a local secondary school, who are representative of the target user population in both age and technology experience. The students were asked to complete a system usability survey (SUS) at the end of this session. The mean overall SUS rating was 71.85 (SD 1.38).

Conclusions: The participatory process involving key stakeholders (clinicians, psychologists, and information technology specialists) has resulted in the creation of an mHealth intervention technology that has the potential to increase accessibility to this type of mental health service for the target population. The app has gone through the initial testing phase, and the relevant recommendations have been implemented, and it is now ready for trialing with both clinicians and their patients.

(J Med Internet Res 2020;22(7):e17481) doi:10.2196/17481

XSL•FO
RenderX

KEYWORDS

mobile apps; suicide; mHealth

Introduction

According to the Global Observatory for eHealth, mobile health (mHealth) is defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [1]. The World Health Organization (WHO) states that "mHealth involves the use and capitalization on a mobile phone's core utility of voice and short messaging service as well as more complex functionalities and applications including general packet radio service, third and fourth generation mobile telecommunications (3G and 4G systems), global positioning system, and Bluetooth technology" [1].

Mobile devices have the potential to deliver evidence-based interventions with greater customization to the individual and at the time when the intervention is required. mHealth programs use mobile technology for a variety of functions, ranging from data collection tools for health care professionals and clinical decision support systems to support health behavior change by patients in the community [2].

The WHO reported that an estimated 800,000 people die by suicide each year globally, constituting a major public health concern. Suicide is the second leading cause of death in young people aged 15 to 24 years [3]. In Ireland, provisional figures indicate that 352 individuals died by suicide in 2018 [4]. Approximately one-third of people who died by suicide had been in contact with mental health care in the year before their death, and approximately 1 in 5 people had contact with a professional in the month before death [5]. Therefore, self-harm and suicide attempts (SAs) represent a pivotal opportunity for intervention. Suicide prevention guidelines recommend safety planning alongside treatment as usual for patients at risk for mental health care [6]. However, safety planning is often not implemented in practice [7], and the feasibility of carrying paper-based plans may be limited, given the transient nature of suicide risk states [8]. Stigma and geographical isolation are two major barriers to help seeking for individuals at risk of suicide [9]. Advancements in mHealth technology could lower these main barriers by directing individuals at risk of suicide, who would not otherwise seek help, to appropriate evidence-based, web-based apps or traditional mental health interventions [10]. Research has indicated the probable benefits of internet-based suicide prevention methods [11]. Furthermore, a survey completed by psychiatric outpatients outlined that 69 of 98 participants (70%) expressed an interest in using a mobile app to track their mental health on a daily basis [12]. Given the established challenge of suicide in young people [3], there was an identified need to design and develop an app for use by 17- to 35-year-old secondary-level mental health service users (adult mental health or child and adolescent mental health). Specifically, those who present with suicidal ideation, nonsuicidal self-injury, or a history of a previous SA. A new mobile app, which is titled SafePlan and developed by the National University of Ireland Galway, was designed to be used as an adjunct to therapy for this at-risk group.

Before the design and development process of SafePlan, a review of existing mobile apps (N=5) providing support in the area of mental health and suicide prevention was conducted (Table 1). These apps were selected based on their subject and overall app store ranking. The review identified a functionality gap whereby none of these apps were combining the capabilities of safety plans, diaries, and other therapeutic intervention worksheets in one. The reviewed apps tended to focus on only one of these support methods, and the majority of them failed to provide any reporting tools for sharing user data.

Table 1. Review of existing apps in the area of mental health and suicide prevention.

App names	Developers	Platforms (iOS/Android)	Safety plan	Diary	Dialectical behavior therapy	Sharing
Suicide Safety Plan [13]	MoodTools	Both	✓[a]	x[b]	x	x
Safety Plan [14]	Blue Bird Technologies	Android	✓	x	x	x
MYPLAN-Your safety plan [15]	Minplan, Denmark	Android	✓	x	x	x
Dbt112 [16]	annadroiddev@gmail.com	Android	x	✓	✓	✓
DBT[c] Travel Guide [17]	dialexisadvies.nl	Both	x	x	✓	✓

[a]✓: contained this functionality.

[b]x: did not contain this functionality.

[c]DBT: dialectical behavior therapy.

Dialectical behavior therapy (DBT) is an evidence-based program aimed at helping people with ongoing difficulties in managing intense emotions. It is used in the treatment of mood disorders, suicidal ideation, and changing behavioral patterns, such as self-harm and substance abuse [18]. In clinical practice, both safety planning and DBT interventions may be combined to provide a tailored approach to suicide prevention treatment. There are many apps currently on the market that specialize in one of these methods, which require a user to download and maintain two unrelated apps if they wish to support these treatments electronically. The lack of a centralized resource for safety planning and DBT supports was a highly motivational

XSL•FO
RenderX

factor in the development of a new app that would combine these prevention techniques. Early design meetings also identified that this approach should support additional intervention methods in future versions of the app by using a modular and flexible design architecture. The user could then enable or disable these options within the app's settings menu. This idea of intervention flexibility was well received by all members of the design team as they felt it had the potential to open the app to a much wider user base.

The proposed new app should have the added advantage of being co-designed with professional clinicians to support best practice patient-professional interaction based on valid data. Traditionally, both safety planning and DBT treatments involve patients maintaining and updating paper-based reports or worksheets between sessions. Mobile technology can provide an opportunity to streamline this process by eliminating the need for these paper-based resources and hence reduce human error and potential loss of material. It also helps to facilitate the recording of real-time data in a timely manner. The migration of these records to a mobile app could have a major impact on the accuracy and completeness of the data being collected [19].

Finally, the lack of any sharing functionality within existing safety planning apps summarized in Table 1, highlighted a further key requirement for the SafePlan app. This was particularly relevant because of the app's intended use envisaged as being an adjunct to therapy, that is, the information within the app should play a key role in patient-professional sessions. The ability to share data with a clinician or trusted helper was a core focus for the early stages of design. Users taking control of their own health information is a growing trend in digital and participatory health care [20,21]. A number of methods were analyzed for the extraction of app data in a user-friendly format, and a consensus was reached whereby a user could wirelessly print a safety plan or DBT report directly from their phone through the app's *My Reports* screen. This option was chosen primarily for its ease of use; however, it also plays an important role in the overall security and privacy of the app. Passing data to web-based portals or cloud accounts was considered too risky an approach based on the confidential nature of the app's data. This is also the reason why it was decided to store all the app's user-entered data locally on the user's device as opposed to an external server. In summary, the aim of this study was to investigate whether mobile technology can offer a way to support the implementation safety planning approaches and therefore support the acquisition of protective coping skills for patients at risk.

Methods

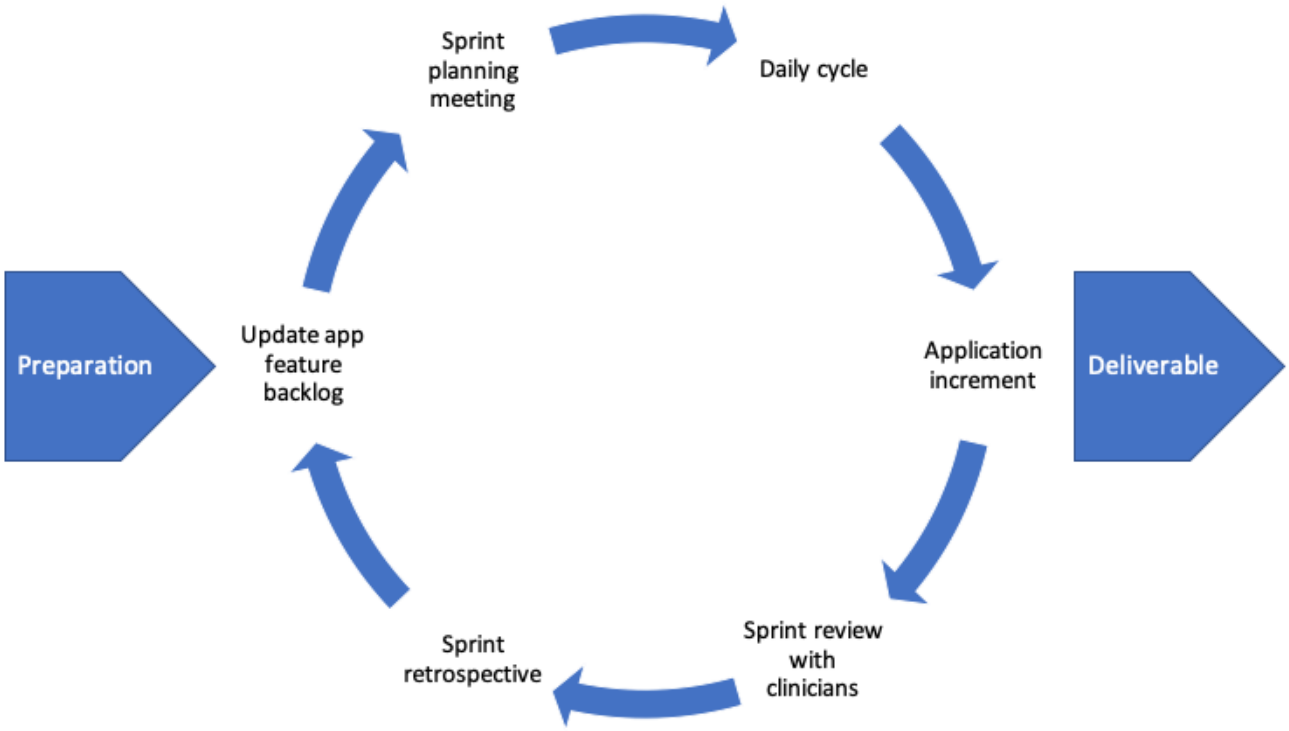
Design Process

The SafePlan app was built using an agile software development approach [22]. The agile methodology is an incremental model of app development where requirements are planned and delivered in short, fixed-length time slots, also known as sprints or iterations [23]. Application requirements are established and refined through active user involvement within the software development process. In this case, external clinicians were recruited to fulfill the role of the user because of the sensitive nature of SafePlan's target audience.

Once the initial app requirements were gathered, they were prioritized into a feature backlog based on their importance and completeness, that is, a proposed app feature was placed toward the bottom of the backlog if it was not yet fully defined. The sprints were typically 2-3 weeks in length and contained a number of stages (Figure 1). A full team review took place at the end of each sprint where the development team and external clinicians provided feedback on the tasks that had been carried out and any possible improvements that could be made. This effectively provided a form of continuous acceptance testing throughout the app's development lifecycle.

XSL•FO
RenderX

Figure 1. Design workshop process.



Requirements Elicitation

A survey (Multimedia Appendix 1) was circulated to frontline health care professionals to assess what features they would prioritize in a new mobile app relating to mental health and suicide prevention. The survey comprised a list of questions relating to the overall usefulness of mobile app technology to their service needs and consisted of 18 potential components that could be incorporated in the new app. There were 15 respondents to the survey, all health care professionals involved in the treatment of adults or children with mental health difficulties.

The majority of respondents were confident that mobile apps could be used in their service, for example, in session, between sessions, as an intervention itself, as an adjunct to other interventions, for younger clients, or as a tool for staff. All but 4 of the respondents used apps in their work or signposted/recommended particular apps to their clients. The most popular apps used and recommended were mindfulness-based apps, including Headspace and Smiling Mind [24,25].

There were concerns raised around using mobile technology to treat patients relating to confidentiality, accessibility, usability, and reliability. These concerns were noted and significantly informed about the app design process, as will be discussed in detail later. The respondents were also asked to independently rank the 18 potential app components in order of usefulness. The top 5 ranked features are set out below:

1. Interactive safety plan that can be recorded and potentially shared
2. Coping skills/training tools (eg, mindfulness exercises, DBT worksheets)
3. Tracking of symptoms/behaviors (eg, behavioral activation, urge to self-harm, suicidal thoughts, and coping skills used)—guided by DBT cards
4. Opportunity to link warning signs with specific coping skills to aid problem solving in times of distress (with a therapist)
5. Details of crisis support services/contacts

These features formed the basis for the initial app design. A number of features were deemed less important but still desirable, and as a result, they were placed on the *nice to have* list, that is, they would be implemented, time and resource permitting. These included components for tracking sleep, mood, and exercise, a capability to visually represent user data using graphs and allowing for a privacy function, that is, the app requires a log-in or passcode to gain access.

Design Process

A design group was established to design and develop the app. The group consisted of 6 members and was multidisciplinary in nature with expertise in a variety of fields such as mHealth, computer science, clinical psychology, and suicide prevention.

In addition to this core group, the design process was supplemented with the inclusion of 5 additional health care workers who took part in the clinician survey. It was agreed that their input would be captured in the form of design workshops over the course of the app's development lifecycle. This iterative method of having regular checkpoints with external clinicians proved to be a valuable way to ensure that the app's construction was aligned with the initial requirements. The design workshops took place at the National University of Ireland, Galway (NUIG), and were held every 2 months during the SafePlan's design and development phases.

XSL•FO
RenderX

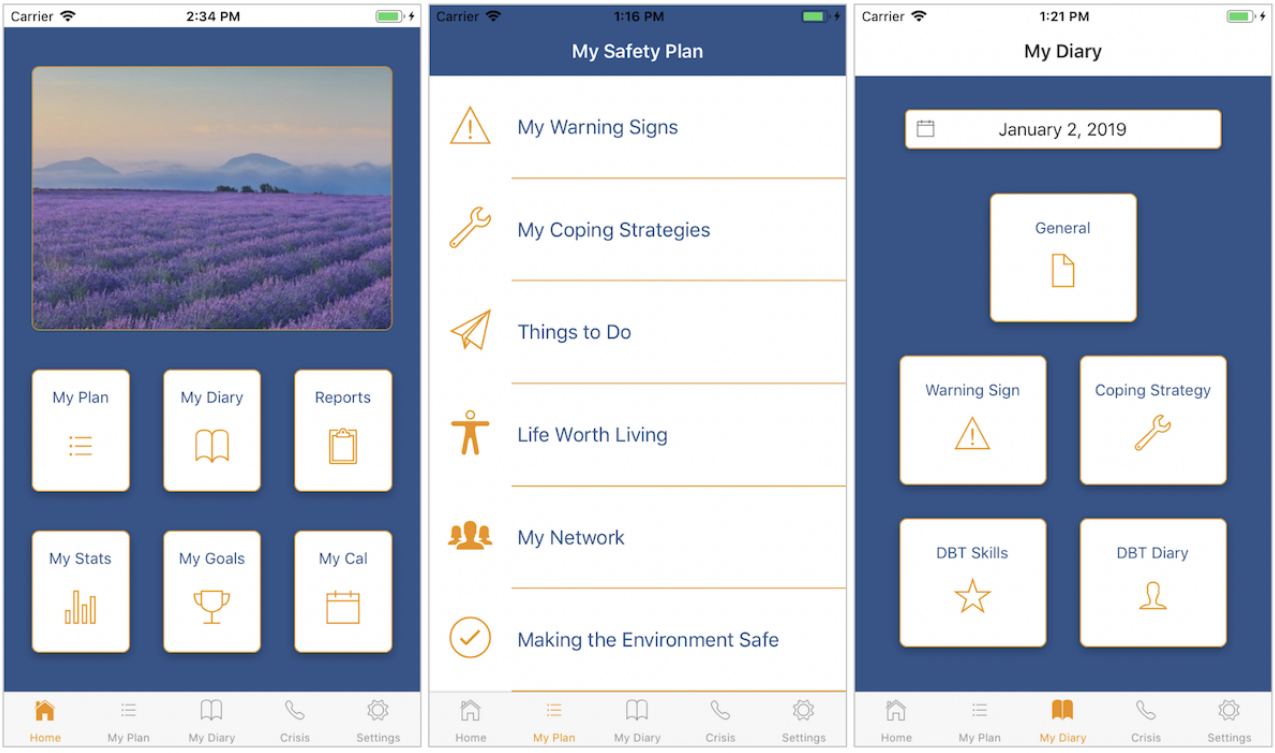
Input and evaluation were provided through the design workshops, and this informed which app features were worked on for that particular development cycle. In an agile approach to software development, these features are formed through the process of refining *User Stories*. A user story is a short, simple description of a requirement told from the perspective of the person who desires the new capability, usually a user or customer of the system [26]. In the context of SafePlan, the user stories originated from the initial requirements survey and were refined over the course of the design workshops.

User Interface Design

Before any development work was carried out, a significant amount of time was spent on creating a user interface (UI) design for SafePlan. Early prototype designs were developed based on the functionality gathered during the requirements elicitation process. These prototypes were designed in accordance with industry-standard design guidelines for mobile app development as well as being informed by a human-computer interaction expert on the design team (with extensive experience of UI design in both industry and academia). Using JustInMind [27], a tool for creating interactive wireframes, a variety of screen mockups were presented during the first design workshop where a number of adjustments were suggested by the attending clinicians based on their experience of working with patients in the area of suicide prevention. These suggestions were mainly focused on usability and resulted in minor updates to the app's overall UI design. For instance, the initial mockups used individual buttons to capture a user's feelings or urges within the DBT diary section of the app. The clinicians were all in agreement that this would be better presented in the form of a slider scale, as this approach should be more familiar to potential users based on their experience with similar apps.

Owing to the vulnerable nature of SafePlan's intended users, it was of extreme importance that the app's content was presented in a simple and well-structured manner. The design team felt it was crucial that navigation to the app's 2 core components—Safety Plan (*My Plan*) and Diary (*My Diary*)—was always evident to the user no matter where they were located in the app's screen hierarchy. This is why both components can be accessed through shortcut buttons on the home screen and via the bottom tab bar, which remains visible on every screen within the app. This can be seen in the screenshots that display the app's Home, Plan, and Diary UI (Figure 2). Care was also taken to ensure that the pictorial imagery available was of a sensitive nature.

Figure 2. Screenshots from Home, Plan and Diary tabs within the SafePlan app.



The final step in completing SafePlan's visual design was devising a color scheme for the app's UI. Following a presentation of alternate themes, the design team decided on a complementary color scheme for SafePlan. A complementary color scheme uses one base color and its complement, the color on the exact opposite side of the color wheel [28] as an accent color to draw attention to important elements on the screen. A base color of deep blue was chosen by the design team, as blue is a calming color, with the associated accent color being orange. Other color schemes, such as monochromatic and analogous, were explored in earlier design workshops; however, a consensus was reached to proceed with a complementary approach.

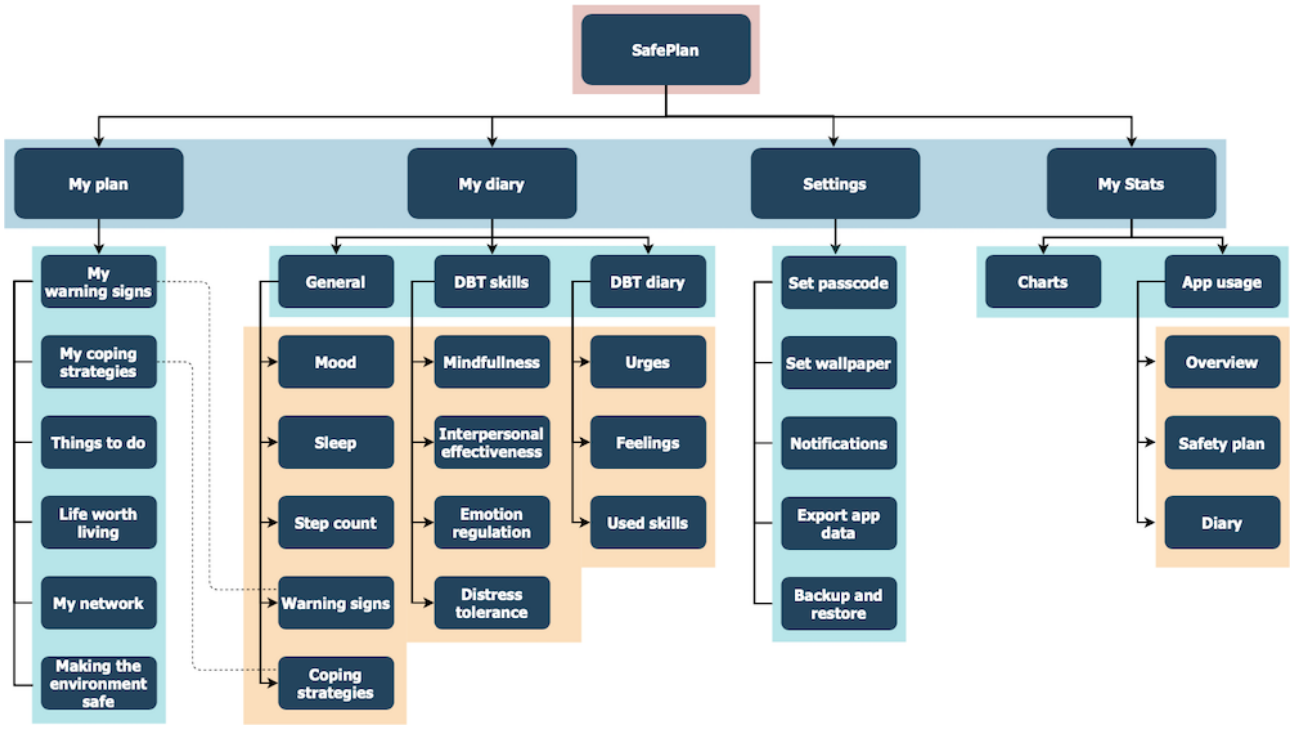
Use Cases

A functional design structure chart (Figure 3) was created based on the requirements feedback obtained from the clinician survey. This outlines the overall depth of the app's wide variety of use

XSL•FO
RenderX

cases that are spread across its core components. These use cases were approved at the first design workshop, and a development schedule was prepared to identify the priority of each app component and the timeline expected to complete each one.

Figure 3. Functional design of SafePlan use cases.



Safety Plan

As discussed, the highest ranked and most requested feature for the app was a shareable interactive safety plan. The safety plan was based on the text-based Safety Plan Template by Brown and Stanley [29], and was digitally enhanced to enable attaching of relevant media to the various sections within the plan, that is, images and videos could be associated with a user's particular coping strategy.

It was also suggested that a user should be able to link their warning signs with specific coping strategies to help with problem solving in times of distress. This feature was added via the *New Warning Sign* screen, where the user can select appropriate strategies from a list of their saved coping strategies within the app. This is a crucial attribute in any safety plan; the user can address a particular warning sign before it fully manifests by following the applicable linked coping strategies. This highlights another key benefit in maintaining a user's safety plan in digital format, as this type of automatic linking could not be achieved using a paper-based approach.

Diary

The diary element of the app is divided into 2 distinct parts. The first is a general diary where a user tracks their overall mood, sleep, or any thoughts they want to note for that day. The second centers around DBT, a well-researched treatment program that aims to help people who have difficulties managing intense emotions [30]. All health care workers participating in the focus group were in favor of including an interactive version of a DBT diary card within the app. This involves the user entering a rating, generally between 0 and 5, for specific feelings and urges they experience throughout the day. It would also enable the user to track what DBT skills they have used over a certain time frame.

One of the key objectives behind the development of the SafePlan was to ensure that it provided access to mental health support for as many people as possible. Appropriately, the DBT option was added to the app's settings menu, whereby a user could activate/deactivate the DBT material within the app. This was to safeguard against users with no experience of DBT treatment feeling overwhelmed or that the app would not be suitable for them. When the DBT option is switched off, the diary component is still a valuable tool as the user can track mood, sleep, and save any warning signs or coping strategies they have encountered during that day. The DBT option is switched off as a default setting when initially downloaded from the respective app stores.

It was also felt that this customization could provide a unique feature for future versions of the SafePlan. The ability to enable or disable the DBT material allows for the potential to incorporate further treatments within the app. For instance, following the same format, other therapeutic interventions could be integrated into a newer version with the same activation functionality, for example, cognitive behavioral therapy (CBT) for suicide prevention. Facilitating this type of tailored approach to mental health treatments could possibly expand the app's user base even further.

Clinician Reports

Eliminating the need for patients to maintain paper plans and worksheets was an important motivation for developing the app.

XSL•FO
RenderX

This was a common challenge reported by clinicians involved throughout the focus group sessions. They explained how easy it is for paper reports to get lost or damaged between weekly sessions and that this can lead to a lack of consistency in patient data. Informed with this knowledge, the design group invested significant effort in establishing the correct format in which to summarize the safety plan and DBT data so that they could be used by clinicians and be available as a replacement for paper reports.

It was agreed that the report structures would remain similar to their paper counterparts. The safety plan report is a point-in-time PDF document that compiles all the entries in the various plan sections. The DBT diary card report is slightly different in that the user can select the specific seven-day period in which they would like to view their entries. This was to ensure that consistency was maintained between traditional DBT diary cards, which are generally updated on a weekly basis.

The aggregation of a user's digital data and its presentation proved challenging for the design team. The collected data relates both to the user's *diary*, stored in the app's local database, and their overall app usage. It was agreed that this information would be most effective in chart form. The user can access the diary charts through the *My Stats* section of the app. The charts are dynamic in that the following options can be selected:

1. Diary item to analyze
2. Period (eg, past 7 days, past 30 days)
3. Diary item to compare with (not applicable with DBT skills)

Tracking a user's overall usage of the app was identified as a potential key tool for clinicians during the second focus group session. Being able to know when a user has viewed a certain safety plan item or even how many times an item has been viewed could provide key insights to not only the therapist or clinician but also the user themselves. Similar to the diary data, these insights can be viewed through the *My Stats* section of the app. They are predominantly centered around the safety plan inputs and are divided into 3 types (last entered, last viewed, and most viewed) for each plan section, that is, warning signs, coping strategies, and contact.

One final method for data analysis that was incorporated into the app is the facility to export diary and usage data as a comma-separated values (CSV) file. This was requested by the clinicians, as it enables further analysis of the CSV data, using statistical software, into long-term trends that are captured through the various diary inputs and usage patterns of their patients. It could also support cohort analysis of patients to identify associations between variables. From these data, a clinician might, for example, observe that a patient's overall mood is significantly affected by either their quality of sleep or one of the many DBT diary card urges/feelings, such as suicide, self-hand, and anger. Patients could then work with their clinicians to identify the necessary steps to take to help manage their mood more effectively.

In collaboration with their clinician, the user is invited to set up regular notifications to remind them to (1) record agreed-upon behaviors and (2) share their data. This feature is included to enhance adherence. Importantly, and following considerable stakeholder input with clinicians, the user will share their data in person with their clinician. User data will not automatically be shared and will not be monitored outside of these agreed-upon times, and this will be made clear to users upon app download. In addition, a copy of the Safety Plan created in-session (to include who will support the person in times of crisis and how) will be shared with the named individuals identified in the plan to ensure that the safety plan is adhered to and supported. To ensure that data are not lost if a user changes the device, for example, users are reminded to back up their data regularly.

The method for sharing CSV files is discussed in the following section.

Technologies Used

The app was built using React Native, a framework for building native mobile apps for iOS and Android devices [31]. It enables the development of native apps using JavaScript and hence greatly reduces implementation time, as two codebases do not have to be maintained when building an app for both platforms.

Owing to the confidentiality and security concerns that were highlighted as part of the clinician survey, it was decided that the app's data should be stored on a device's local database rather than being uploaded to external servers. SQLite was used for the local database because it is lightweight, fast, and reliable [32].

Sharing and exporting data from the app also impacts privacy and security-related concerns raised by the clinicians. In order for the app to be effective, it is essential that a user shares their safety plan or weekly diary card with their clinician in a secure manner. This requires a thorough analysis, and many sharing strategies are considered. A detailed review (Table 2) was carried out to establish what approaches similar apps (mobile apps used as an adjunct to therapy) had used to extract user data for an external party, that is, a clinician. The review examined a wide variety of methods that have been used previously to share user data between mHealth apps and invested third parties, that is, clinicians. The majority of apps reviewed were using some form of an external server or cloud portal where the user could upload their data for a clinician to access. This sharing technique requires additional security resources because of the handling of sensitive user data and, consequently, may result in further legal and ethical considerations. A different approach, adopted by CBT-I Coach, provided functionality, whereby the user could email their sleep data to themselves so that they could print it and share with their clinician.

XSL•FO
RenderX

Table 2. Review of existing apps' sharing strategies.

References	App names	Sharing methods
Kuhn et al [33]	CBT-I Coach	Users can email data to themselves
Matthews and Doherty [34]	Mobile Mood Diary	Users asked if they would like to upload their data from the app to an external server. A web-based tool allows users to visualize their recorded data. Only users have log-in credentials for this tool, and the idea was that they would log on at the beginning of their session with their therapist so they could review
Rivzi et al [35]	DBT Coach	Button within app's settings menu that a user could press to send data. Data would be uploaded to a cloud file that could be accessed by research staff
Reger et al [36]	PE Coach	App did not facilitate electronic sharing of any data. This was noted as a limitation of the app by the researchers
Pramana et al [37]	SmartCAT	Therapist can access user's data through a web-based portal

Having conducted the review and discussed the potential implications of each strategy with the app stakeholders, a safety-first approach was adopted to extract user data. The design group concluded that all the control should remain with the user and decided against the facility to upload data to any cloud file or external server. In the current version of the app, a user can generate a report of either their safety plan or DBT diary card for a specified period. These reports can then be printed directly from the user's device or can be saved as a PDF file and distributed as they themselves might require. This solution satisfied the criteria for ensuring that the user is fully responsible for their data and that it will stay within the user's environment unless they decide otherwise.

Implementation and Testing

Following the design workshops and development sprints, the first version was implemented for beta testing. The app was tested internally by the core design team, after which final changes were made to remove bugs and prepare the app for external interaction.

Ethical approval was granted by the School of Psychology Ethics Committee at the NUIG on May 05, 2016, to evaluate the Corofin Abbeyknockmoy, Athenry, Turloughmore Community Health (CAATCH) app. This app was also developed by the design group to raise awareness and provide information about local support services. Ethical approval was granted by the School of Psychology Ethics Committee NUI Galway on May 28, 2019, with respect to the SafePlan study.

On completion of beta testing, students (N=18) were recruited from a local secondary school for an evaluation session. The group was representative of age (14-16 years) and technology experience (they used their own mobile phones, both Android and iOS, with the devices and user technology experience being representative of this demographic) of one of the target user populations. Their role was to test the usability and functionality of the app, rather than the actual content. The students were carefully screened in advance of participating in the testing process, via a consent process involving parents (Multimedia Appendix 2), and clinicians were available onsite during the evaluation session to answer any questions. The need for representative users was required due to the sensitive and vulnerable nature of SafePlan's intended audience. In addition, any interaction with this audience would require intervention by clinicians over a longer period. There are plans for this type of testing in the next stage of the SafePlan project through the form of a randomized controlled trial (RCT).

The students were asked for their opinion on the overall functionality and usability of the app. A proposed task list was created (Multimedia Appendix 3) to ensure that coverage of the app's features mirrored its expected long-term normal usage. This is a typical approach adopted for the usability testing of mHealth apps [1]. After the students had carried out these tasks, they were asked to complete a system usability survey (SUS) [38], a well-known tool in usability practice and research, and an open-ended evaluation questionnaire. The SUS is an industry-standard 10-item scale that explores the perceived usability of a technological application. The scale has excellent psychometric properties with measures of reliability scoring over 0.90, good indicators of validity, and sensitivity [39]; and a Cronbach alpha of .92 [38]. Norming studies [39] have presented normative data providing an empirical basis for the interpretation of SUS scores. According to such normative data, an SUS score of 65 can be interpreted as a marginally acceptable result. More recently, Finstad [40] compared the original five-point Likert scale instrument with a modified seven-point Likert scale version. The findings supported the conclusion that seven-point Likert items provide a more accurate measure of a participant's true evaluation, and therefore, this was the approach adopted in this study.

Results

Quantitative Analysis

The quantitative data determined from the SUS are presented in Multimedia Appendix 4. Participants recorded higher scores for ease of use and confidence in using the app. There was a relatively high mean score (3.17) in the last question on the SUS with regard to users needing to learn many things before commencing using the app. However, this question also resulted in the highest SD (1.95), suggesting that a greater spread of ratings occurred for this question.

The question relating to inconsistency within the app received the lowest mean score (2.28) and the lowest SD (0.89). These markers indicate that the participants came to a consensus that the app provides users with an overall cohesive experience. The mean overall SUS rating was 71.85 (SD 1.38). An SUS score above 68 is considered above average based on research in this area [41]. Although the SafePlan scored well in this regard, the

XSL•FO
RenderX

usability testing was carried out by representative users, and therefore, the app's functionality was not entirely relevant to them. This score is likely to increase when tested with real users.

Qualitative Analysis

The feedback collected through the evaluation questionnaire is categorized and summarized below. This was carried out following a thematic analysis approach. Thematic analysis is a method for identifying, analyzing, and reporting patterns within data [42].

Theme 1: UI Design

The SafePlan design and UI were well received by the majority of participants. The intuitive structure and overall simplicity of the visual appearance were frequently noted in the questionnaire responses:

> *I like the set up. It's visually pleasing and easy to read. Planning was somewhat fun!* [P4]

> *Simple easy UI. Many different uses for all users and cases.* [P5]

The shade of orange used in the color scheme was disliked by two of the participants. They noted that this color was quite sharp and may potentially unsettle users. However, this particular participant also remarked that the app's primary color of blue worked well because of its calming nature:

> *I didn't like the colour orange on the apps cover. The blue was a good idea because it's a relaxing colour but the orange is intense.* [P10]

Two more participants commented on the switch interface used to change the app's home screen wallpaper. During the session, they suggested that this would be more intuitive if it was implemented with a button instead:

> *...the fact that the set wallpaper has a slider.* [P5]

Theme 2: SafePlan Tutorial

Although the overall UI was rated highly by the participants, it was suggested that adding some form of tutorial for first-time users would be helpful. This could possibly be in the form of a video directly linked within the app or, alternatively, a number of nonintrusive dialog boxes could appear on initial downloads containing instructions for key components within the app:

> *...quick tutorial the first time you enter each screen (speech bubbles of important points).* [P5]

> *Some features were initially a bit difficult to locate and understand but other than that it was excellent (tutorial would be helpful).* [P9]

Similarly, the high number of functions available within the app proved to be overwhelming and difficult to locate for some participants during the evaluation session:

> *Not everything is completely obvious with how to use.* [P17]

However, there was a sense that once they got the required directions and after some time had elapsed, the app's value became more apparent:

> *It is a bit confusing to get into the flow of the app, but you can understand it quickly.* [P16]

Theme 3: Confidentiality

At the beginning of the evaluation session, the participants attended a presentation that detailed the background and development of the app. The storage of user data created within the app was explained, and this was positively noted by one participant in the questionnaire:

> *I liked how the information was strictly confidential. I feel it will allow the user to be more at ease using the app.* [P8]

The ability to secure the app with a passcode was also remarked upon favorably by the participants. They felt this was a fundamental feature for the app because of the nature of the information being recorded and stored:

> *I liked how organised and accessible the app was. I also thought that the way you could put a passcode on the app was important.* [P11]

Theme 4: Additional Functionality

The final point in the evaluation questionnaire asked the participants if they would like to see any additional features incorporated into the app. This is a valuable open-ended question, as some of these suggestions may influence future versions of the app.

The idea of a food tracker diary was mentioned by more than one participant. A place where a user could record the food they have eaten for a particular day and obtain insights based on the data collected over a specified period of time. This would be a natural supplement to the app's diary component, where the user already has the ability to record their mood, sleep, and step count on a given day:

> *...maybe a food tracker because some people may not believe they are eating too much / too little when they are.* [P9]

Another recommendation from one of the participants focused on providing the app user with motivational messages each morning through the form of push notifications:

> *Maybe some relaxing videos or a notification each morning with something motivational.* [P1]

Finally, adding calming music or sounds to the app was also quite a popular suggestion. The following participant brought this idea up in connection with aiding the user to fall asleep. This feature could potentially be linked with the app's sleep diary in future versions of the app:

> *You should add a feature where you can add in calming music or noises to fall asleep.* [P10]

Discussion

Principal Findings

The principal findings of this project were obtained from acceptance and usability tests, which were based on feedback from health care workers and representative users. Health care workers were recruited to provide iterative feedback on the

XSL•FO
RenderX

app's design and functionality throughout its development lifecycle. This feedback was key to finalizing the app from a clinical perspective, as it ensured that no unnecessary or inappropriate features were added to the app's portfolio. Usability testing was undertaken with representative users in terms of age (14-16 years) and technology experience. Representative users were necessary at this early stage of testing because of the sensitive nature of SafePlan's intended audience. Carefully screened students from a local secondary school were invited to attend an evaluation session, where they were presented with a number of specific tasks to carry out on the app to determine its overall usability.

The feedback received from the usability testing day was largely positive. The participants perceived the main benefits of the SafePlan app were its overall UI design and emphasis on user confidentiality. In particular, the acknowledgment of the app's privacy features was a strong vindication for the design team, as this was an area of significant focus during the initial planning meetings. The most common issues in mHealth are privacy and data security [43]. These concerns were addressed from the outset of SafePlan's development by ensuring that all user data are stored locally on their own device and providing the user with an option for securing the app with a passcode.

Although the majority of evaluations were positive, a small number of potential improvements were identified for future versions of the app. The amount of information content and functionality within the app was somewhat overwhelming for some participants on initial use. There were suggestions that an introductory tutorial video for first-time users could be extremely beneficial to give them a general overview of the app's features and use cases. On this point, it is also worth noting that the participants' experience with SafePlan was not representative of real, ongoing use. It is expected that the majority of the app's users will be guided through its content by a clinician, and this may mitigate any confusion they may experience on initial use. Another participant proposed that the app provides users with motivational messages or relaxing videos each morning through the form of push notifications.

Limitations

Although the usability and acceptance testing proved to be beneficial in designing and developing a clinically informed, user-friendly mobile app, further ethically conducted evaluations involving patients with suicidal thoughts or behavior will provide much greater insight into whether participants interact with the app in the way it was intended. Previous researchers have highlighted the need for more robust testing of mHealth apps within clinical settings [44]. The research team has established experience in co-design, iterative design processes, and the selection and use of measures to evaluate user engagement and digital intervention usability [45-47]. A pilot RCT investigating the feasibility of using SafePlan as an adjunct to therapy in mental health services is currently being planned (HRB CSF 2020–010). The aims of this trial are to (1) assess the feasibility and acceptability of the SafePlan intervention as an adjunct to therapy for individuals at risk of suicide accessing Irish mental health services, (2) examine the feasibility of a definitive RCT of the SafePlan intervention, and (3) assess the feasibility and acceptability of the SafePlan data collection methods (including in-built ecological momentary assessment) for clinicians and patients.

This is particularly relevant when it comes to observing a user's long-term interaction with the app's various diary and safety plan functions. It is anticipated that a user will maintain and update their data within the app on a regular basis, and this is where the most value will be gained, both for the user and their clinician. The long-term usage of health apps is highly significant, as some studies have reported log-ins dropping sharply to nearly zero after 1 month from download, even for the highest rated apps [18]. With this in mind, further recording of user feedback during the pilot stage is required to determine if the app has the desired lasting effect on its users, and this would be a key element of a follow-on study with end users. Currently, there is no warning on the app in terms of usage, and a cautionary preamble section will be added to the app stating that it is only to be used after undergoing an orientation session. Furthermore, currently, there are no automated warning notifications built into the system, and this has been identified as a future requirement, whereby patients will be informed when they should consider seeking professional help.

Comparison With Prior Work

Many members of SafePlan's core design team were previously involved in the development of another app relating to improving outcomes in suicide prevention. The CAATCH app [44] was released in 2014, and its main goal was to raise awareness, provide information, and signpost to education and training programs available at the community level within the local area. In terms of the involvement of end users, CAATCH involved members of the public and patients in the design and testing phases, which guided the overall intervention and subsequently shaped the design of the current SafePlan app. It was the experience of young people accessing mental health services and actively seeking a digital means of storing and using their SafetyPlan/DBT code cards, which ultimately led to the development of the SafePlan.

Although SafePlan also aims to raise awareness and provide information, it operates at a patient-clinician level where it can be used as an adjunct to treatment. It is a significantly more interactive app in that it accepts various forms of user input and provides tools for visualizing this input in the form of charts and reports, so that users can analyze their data over a specified period.

Conclusions

In conclusion, through the process of iterative design with the input of targeted clinicians and key design personnel, a fully functional version of the SafePlan mobile app was developed. This app has the potential to offer an accessible adjunct to face-to-face therapy and to support the generalization of skills acquired in-session to beyond the clinical setting. It also provides useful functionality that may lead to improvements in the interaction between a patient and their clinician. This can primarily be demonstrated through the process of transferring previously recorded paper-based reports and worksheets to specific sections within the app. This facilitates more accurate

XSL•FO
RenderX

and timely data recording practices and subsequent analysis, with a correspondingly positive impact on treatment outcomes.

Following on from our review of existing mobile apps providing support in the area of mental health and suicide prevention (as presented in Table 1), further apps have been added to the app stores [48]. With approximately 20,000 mental health apps now reported in the app stores by the American Psychological Association [49], this number will continue to grow, whereas the core challenge of maintaining privacy, providing reporting tools for controlled sharing of user data, and combining safety planning with other therapeutic interventions remains. Safety planning intervention is an evidence-based brief intervention for suicidality, which is increasingly considered to be the best practice in terms of health care practice [6]. SafePlan potentially operationalizes a more interactive and individualized means of implementing this intervention while also offering the potential to combine this safety planning intervention with other prevention techniques (eg, DBT or CBT) in one flexible app. This intervention flexibility was welcomed by all participating clinicians, as it provides the potential to open the app to a wider user cohort.

SafePlan will require further evaluation from users within the target population. This was a topic of consideration for the design team, and it was agreed that because of the nature of the app's intended users, it would not be appropriate to involve patients at this early stage of development. The combination of input from clinicians and groups that were partially representative of the end users was seen as a valid substitute for patients' evaluation until sufficient evidence was gathered to carry out a pilot implementation with the app's intended audience, for example, a pilot study with patients and clinicians to gain further insight.

Acknowledgments

This work was funded by the Health Service Executive (HSE) National Office for Suicide Prevention and carried out with the support of the NUIG and HSE Community Healthcare West and Midwest.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of eighteen potential app components used in the initial app survey:.
[DOCX File , 14 KB - jmir_v22i7e17481_app1.docx]

Multimedia Appendix 2
Evaluating the SafePlan App Parent/Guardian Consent Form.
[DOCX File , 35 KB - jmir_v22i7e17481_app2.docx]

Multimedia Appendix 3
Task list for usability testing day.
[DOCX File , 32 KB - jmir_v22i7e17481_app3.docx]

Multimedia Appendix 4
Responses were scored on a 7-point Likert scale ranging from 1=strongly disagree to 7=strongly agree.
[DOCX File , 14 KB - jmir_v22i7e17481_app4.docx]

References

1. Kay M, Santos J, Takane M. Mhealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
2. Melia R, Francis K, Duggan J, Bogue J, O'Sullivan M, Chambers D, et al. Mobile health technology interventions for suicide prevention: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2018 Jan 26;7(1):e28 [FREE Full text] [doi: 10.2196/resprot.8635] [Medline: 29374003]
3. Suicide. World Health Organization. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2020-07-06]
4. Suicide. National Suicide Research Foundation. URL: https://www.nsrf.ie/statistics/suicide/ [accessed 2019-07-28]
5. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 2002 Jun;159(6):909-916 [FREE Full text] [doi: 10.1176/appi.ajp.159.6.909] [Medline: 12042175]
6. Stanley B, Brown GK, Brenner LA, Galfalvy HC, Currier GW, Knox KL, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. JAMA Psychiatry 2018 Sep 1;75(9):894-900 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1776] [Medline: 29998307]

XSL•FO
RenderX

7. Higgins A, Morrissey J, Doyle L, Bailey J, Gill A. Best Practice Guidance for Mental Health Services: Supporting You to Meet Regulatory Requirements and Towards Continuous Quality Improvement. Ireland's Health Services. 2015. URL: https://www.hse.ie/eng/services/list/4/mental-health-services/mental-health-guidance/best-practice-guidance/documents/best-practice-guidance-for-mental-health-services.pdf [accessed 2020-07-10]
8. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Lee SJ, Rial KV, et al. Developing a brief suicide prevention intervention and mobile phone application: a qualitative report. J Technol Hum Serv 2015 Oct 1;33(4):345-357 [FREE Full text] [doi: 10.1080/15228835.2015.1106384] [Medline: 26977137]
9. Christensen H, Batterham PJ, O'Dea B. E-health interventions for suicide prevention. Int J Environ Res Public Health 2014 Aug 12;11(8):8193-8212 [FREE Full text] [doi: 10.3390/ijerph110808193] [Medline: 25119698]
10. Ybarra ML, Eaton WW. Internet-based mental health interventions. Ment Health Serv Res 2005 Jun;7(2):75-87. [doi: 10.1007/s11020-005-3779-8] [Medline: 15974154]
11. Lai MH, Maniam T, Chan LF, Ravindran AV. Caught in the web: a review of web-based suicide prevention. J Med Internet Res 2014 Jan 28;16(1):e30 [FREE Full text] [doi: 10.2196/jmir.2973] [Medline: 24472876]
12. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
13. MoodTools: Feeling Sad or Depressed? Lift Your Mood With Free, Evidence-Based Tools. 2019. URL: http://www.moodtools.org/ [accessed 2020-07-10]
14. BlueBird IT Solutions Inc. 2019. URL: http://www.bluebirdtechnologies.com [accessed 2020-07-10]
15. Minplan. 2019. URL: http://minplan.org/ [accessed 2020-07-10]
16. Dbt112. Google Play. 2019. URL: https://play.google.com/store/apps/details?id=se.annadroid.Dbt112&hl=en_IE [accessed 2020-07-10]
17. Dialexis. 2019. URL: https://www.dialexisadvies.nl/en/ [accessed 2020-07-10]
18. Flannery RB. Review of cognitive-behavioral treatment of borderline personality disorder. Psychiatr Rehabil J 1993 Oct;17(2):201-202. [doi: 10.1037/h0095583]
19. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. The Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
20. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 5;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
21. Hood L, Auffray C. Participatory medicine: a driving force for revolutionizing healthcare. Genome Med 2013;5(12):110 [FREE Full text] [doi: 10.1186/gm514] [Medline: 24360023]
22. Williams L, Tomayko J. Agile software development. Comput Sci Educ 2010 Aug 9;12(3):167-168. [doi: 10.1076/csed.12.3.167.8613]
23. Agile Methodologies. Blueprint: Enterprise Software for Intelligent Process Automation. 2019. URL: https://www.blueprintsys.com/agile-development-101/agile-methodologies [accessed 2020-07-10]
24. Headspace: Meditation and Sleep Made Simple. 2019. URL: https://www.headspace.com/ [accessed 2020-07-10]
25. Smiling Mind. 2019. URL: https://www.smilingmind.com.au/ [accessed 2020-07-10]
26. Cohn M. User Stories. Mountain Goat Software. 2019. URL: https://www.mountaingoatsoftware.com/agile/user-stories [accessed 2020-07-10]
27. All-in-One Prototyping Tool for Web and Mobile Apps: From Wireframes to Highly Interactive Prototypes. Justinmind. 2019. URL: https://www.justinmind.com/ [accessed 2020-07-10]
28. Complementary Color Scheme. Paletton - The Color Scheme Designer. 2019. URL: https://www.paletton.com/wiki/index.php?title=Complementary_color_scheme [accessed 2020-07-10]
29. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012 May;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
30. Linehan M. Cognitive Behavioral Treatment of Borderline Personality Disorder. New York, USA: Guilford; 1993.
31. React Native: A Framework for Building Native Apps Using React. 2019. URL: https://facebook.github.io/react-native/ [accessed 2020-07-10]
32. What Is SQLite? SQLite. 2019. URL: https://www.sqlite.org/index.html [accessed 2020-07-10]
33. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: 10.5664/jcsm.5700] [Medline: 26888586]
34. Matthews M, Doherty G. In the Mood: Engaging Teenagers in Psychotherapy Using Mobile Phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada URL: https://dl.acm.org/doi/10.1145/1978942.1979379 [doi: 10.1145/1978942.1979379]
35. Rizvi SL, Hughes CD, Thomas MC. The DBT coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]

XSL•FO
RenderX

36. Reger GM, Hoffman J, Riggs D, Rothbaum BO, Ruzek J, Holloway KM, et al. The 'PE coach' smartphone application: an innovative approach to improving implementation, fidelity, and homework adherence during prolonged exposure. Psychol Serv 2013 Aug;10(3):342-349. [doi: 10.1037/a0032774] [Medline: 23937084]
37. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
38. Sauro J. Measuring Usability With the System Usability Scale (SUS). MeasuringU. 1996. URL: https://measuringu.com/sus/ [accessed 2020-07-10]
39. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
40. Finstad K. Response interpolation and scale sensitivity: evidence against 5-point scales. J Usability Stud 2010;5(3):104-110 [FREE Full text]
41. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
42. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012 Oct 2;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
43. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
44. CAATCH. AppAdvice LLC. 2019. URL: https://appadvice.com/app/caatch/939375449 [accessed 2020-07-10]
45. Safety Planning Guide: A Quick Guide for Clinicians. Suicide Prevention Resource Center. 2019. URL: http://www.sprc.org/sites/default/files/SafetyPlanningGuide%20Quick%20Guide%20for%20Clinicians.pdf [accessed 2020-07-10]
46. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile health technology interventions for suicide prevention: systematic review. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e12516 [FREE Full text] [doi: 10.2196/12516] [Medline: 31939744]
47. Melia R, Bogue J, Duggan J, Walton J, Brown G. Development of a Suicide Prevention Mobile App: A Case Study. In: Poster Presented at the International Congress of Female Mental Health. 2017 Presented at: FMH'17; November 29-December 1, 2017; Dublin, UK URL: https://www.iawmh2017.org
48. DBT Daily. Swasth. URL: https://www.swasth.co/dbt-daily/ [accessed 2020-07-10]
49. Pappas S. Providing Care in Innovative Ways. American Psychological Association (APA). 2020. URL: https://www.apa.org/monitor/2020/01/cover-trends-innovative-ways [accessed 2020-07-10]

Abbreviations

CAATCH: Corofin, Abbeyknockmoy, Athenry, Turloughmore Community Health
CBT: cognitive behavioral therapy
CSV: comma-separated values
DBT: dialectical behavior therapy
HSE: Health Service Executive
mHealth: mobile health
NUIG: National University of Ireland, Galway
RCT: randomized controlled trial
SA: suicide attempt
SUS: system usability scale
UI: user interface
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by V Menon, KS Pheh; comments to author 10.02.20; revised version received 27.05.20; accepted 11.06.20; published 30.07.20.

Please cite as:
O'Grady C, Melia R, Bogue J, O'Sullivan M, Young K, Duggan J
A Mobile Health Approach for Improving Outcomes in Suicide Prevention (SafePlan)
J Med Internet Res 2020;22(7):e17481
URL: http://www.jmir.org/2020/7/e17481/
doi:10.2196/17481
PMID:32729845

XSL•FO
RenderX

©Conor O'Grady, Ruth Melia, John Bogue, Mary O'Sullivan, Karen Young, Jim Duggan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Participatory Interventions for Sexual Health Promotion for Adolescents and Young Adults on the Internet: Systematic Review

Philippe Martin[1,2,3,4*], MSc; Lorraine Cousin[1*], MSc; Serge Gottot[4*], PhD, MD; Aurelie Bourmaud[1*], MD, PhD; Elise de La Rochebrochard[2,3*], PhD; Corinne Alberti[1*], MD, PhD

[1]Université de Paris, ECEVE, INSERM, Paris, France

[2]Institut National d'Etudes Démographiques, UR14 – Sexual and Reproductive Health and Rights, Paris, France

[3]Université Paris-Saclay, Université Paris-Sud, UVSQ, CESP, INSERM, Le Kremlin Bicetre, France

[4]GDID Santé, Paris, France

[*]all authors contributed equally

Corresponding Author:
Philippe Martin, MSc
Université de Paris
ECEVE
INSERM
10 Avenue de Verdun
Paris, 75010
France
Phone: 33 676606491
Email: philippe.martin@inserm.fr

Abstract

Background: The World Health Organization recommends the development of participatory sexuality education. In health promotion, web-based participatory interventions have great potential in view of the internet's popularity among young people.

Objective: The aim of this review is to describe existing published studies on online participatory intervention methods used to promote the sexual health of adolescents and young adults.

Methods: We conducted a systematic review based on international scientific and grey literature. We used the PubMed search engine and Aurore database for the search. Articles were included if they reported studies on participatory intervention, included the theme of sexual health, were conducted on the internet (website, social media, online gaming system), targeted populations aged between 10 and 24 years, and had design, implementation, and evaluation methods available. We analyzed the intervention content, study implementation, and evaluation methods for all selected articles.

Results: A total of 60 articles were included, which described 37 interventions; several articles were published about the same intervention. Process results were published in many articles (n=40), in contrast to effectiveness results (n=23). Many of the 37 interventions were developed on websites (n=20). The second most used medium is online social networks (n=13), with Facebook dominating this group (n=8). Online peer interaction is the most common participatory component promoted by interventions (n=23), followed by interaction with a professional (n=16). Another participatory component is game-type activity (n=10). Videos were broadcast for more than half of the interventions (n=20). In total, 43% (n=16) of the interventions were based on a theoretical model, with many using the Information-Motivation-Behavioral Skills model (n=7). Less than half of the interventions have been evaluated for effectiveness (n=17), while one-third (n=12) reported plans to do so and one-fifth (n=8) did not indicate any plan for effectiveness evaluation. The randomized controlled trial is the most widely used study design (n=16). Among the outcomes (evaluated or planned for evaluation), sexual behaviors are the most evaluated (n=14), followed by condom use (n=11), and sexual health knowledge (n=8).

Conclusions: Participatory online interventions for young people's sexual health have shown their feasibility, practical interest, and attractiveness, but their effectiveness has not yet been sufficiently evaluated. Online peer interaction, the major participatory component, is not sufficiently conceptualized and defined as a determinant of change or theoretical model component. One potential development would be to build a conceptual model integrating online peer interaction and support as a component.

(J Med Internet Res 2020;22(7):e15378) doi:10.2196/15378

XSL•FO
RenderX

KEYWORDS

sexual health; health promotion; internet; participatory interventions; adolescents and young adults; methods

Introduction

Adolescent sexual exposure is of concern due to the risk of contracting sexually transmitted infections (STIs), experiencing an unwanted pregnancy, and unexpected paternity/maternity [1]. Among the 333 million new cases of STIs each year, the highest rates occur among those aged 20 to 24 years, followed by those aged 15 to 19 years [2]. Among a group of 21 countries, the pregnancy rate among those aged 15 to 19 years is highest in the United States (57 pregnancies per 1000 females) [3]. The proportion of teenage pregnancies that result in abortion varies by country, but in half of those for which recent information is available (mainly in Europe, North America, and Oceania), 35%-55% of pregnancies ended in abortion [3]. In 2014, in the United States, females aged <15 years and 15 to 19 years accounted for 0.3% and 10.4% of all reported abortions in the country, respectively [4].

Adolescence and the transition to adulthood marks the entry into sexuality. Sexual health requires a positive and respectful approach to sexuality and sexual relations, and the ability to have enjoyable and safe sexual experiences that are free from coercion, discrimination, and violence [5]. Adolescents and young adults (AYA) represent a priority population for sexual health promotion and education [6]. The associated fields of intervention encompass the development of knowledge and level of information, the development of attitudes to sexual health (attitudes toward safe sex practice, including attitudes to condom use or voluntary testing for STIs), and the development of personal competencies and supportive relational skills (critical thinking, consent, negotiation, open-mindedness, respect, self-esteem).

For example, as stated by the Information Motivation Behavioral Skills (IMB) model (applied and validated for HIV risk reduction), behavioral competencies and therefore health behaviors may be influenced by the level of information, but also by motivation, namely beliefs and attitudes toward a particular health behavior and the perceived social support (or social norm) to engage in this behavior [7]. In addition, health literacy is the ability of individuals to obtain, process, and understand the information and services necessary to make appropriate health decisions [8]. Increase health literacy would enable the improvement of appropriate health decision making with regard to sexual health, promoting equity and achieving the United Nations' Sustainable Development Goals 2030 [9].

The recommendations of the World Health Organization are clearly stated [10]: sexuality education must be participatory (young people should not be mere passive receivers), interactive (with educators and program designers), and continuous. This education must be adapted to the language of the young people, while also teaching appropriate terminology to strengthen their communication skills.

In health promotion, digital media interventions for sexual health have great potential because of the scope and popularity of technologies such as the internet and mobile phones, especially among young people [11,12]. Interactive online interventions for sexual health promotion can also lead to better knowledge, self-efficacy, and positive sexual behavior, and have demonstrated a reduction in STIs [12].

The internet is a major health information resource, and online health information research is an important prerequisite for health empowerment and literacy [13,14]. Moreover, research on information flows and attitudes within social networks suggests that links between people can promote the exchange of relevant information between peers, and affect their attitude toward this information, as individuals are more receptive to information shared by others who are like them [15]. For example, the popularity of social networking sites and their interactive features have great potential to reach young people, and offer a new way to engage and communicate with AYAs, including the provision of appropriate education [16]. Nevertheless, their uses are for the most part "passive," and social networking sites are not yet used as tools for multidimensional communication and networking [17].

Our research question is whether interventions for the promotion of young people's sexual health include participatory components, and if so, how they are integrated and how the interventions are evaluated. Some publications and literature reviews have investigated sexual health interventions on the internet, social media [12,18], online serious games [19], or in digital media [12,20,21]. However, no publication has focused on the participatory aspects of this type of intervention in sexual health specifically aimed at young people (participation in an activity such as online games, quizzes), particularly interactive features such as the exchange of information and experiences between peers (persons of the same age, social context, function, education, or experience) or with professionals. The aim of this review is to identify and describe existing studies and the methods used to assess online participatory interventions aimed at promoting AYA's sexual health.

Methods

Overview

This systematic review was based on international scientific literature and grey literature. The review is structured in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement [22] and follows the associated guidelines (Multimedia Appendix 1). The systematic review protocol has previously been published on the PROSPERO International Prospective Registry of Technical Reviews (ID CRD42018088240).

Inclusion Criteria

Articles were included without time restriction according to the 5 following criteria: (1) Study of an intervention including a sexual health theme; (2) Population aged between 10 and 24 years (with an average age or an interval comprising all or at least part of this age group), because the WHO defines adolescents as aged 10 to 19 years and young people as aged

XSL•FO
RenderX

15 to 24 years [23,24]; (3) Study of a participatory intervention; (4) Study of an intervention conducted on the internet (website, social media, online gaming); (5) Design, implementation, and evaluation methods must be available via the article.

Strategy Search

The electronic search strategies are described in Multimedia Appendix 2. We used the PubMed search engine for our main search. For complementary research, we used the Aurore database of Institut National d'Études Démographiques (INED; a French public research institute), which includes scientific databases and grey literature, allowing access to a range of databases and electronic journals (see Multimedia Appendix 2 for selected international search engines). The last update was on January 28, 2019.

Study Selection

Reports were assessed by two reviewers (PM and LC), who screened the titles and abstracts to identify relevant studies. Full texts were read when abstracts met inclusion criteria, and when abstracts were not clear enough to determine eligibility. Disagreements between reviewers were resolved by discussion. When the full text was not available, authors were contacted by email; all the contacted authors responded favorably and shared their articles with us.

Data Collection

A standardized data collection form was developed, and two reviewers independently extracted data from studies. Our extraction grid was developed using the PICOTS (populations, interventions, comparators, outcomes, timing, and setting) elements [22], and was completed using Michie's taxonomy [25] to collect information on the behavior change techniques (BCT) used by interventions. The studies were classified according to different types: research protocol only, effectiveness evaluation, and process evaluation. Protocol articles are planned studies containing only the conceptual and evaluative methods intended for intervention research. An effectiveness study is defined as a demonstration of an intervention's efficacy in natural situations. It provides evidence of the intervention's effect on determinants or health outcomes. A process study provides evidence on the implementation and feasibility of an intervention, and also rates the intervention for attractiveness and acceptability. It helps to assess the reliability and quality of implementation, to clarify causal mechanisms, and to identify contextual factors associated with variations in outcomes [26].

Analysis

For the final studies selection phase, the degree of interreader agreement was assessed for both readers through the calculation of the κ coefficient.

We conducted descriptive analyses on data collected from studies on the following points: description of the population; characteristics of study methodology; description of the intervention; description of the media used; description of methods used for effectiveness, and process evaluation. We used Michie's [25] taxonomy to analyze the BCT used by interventions, depending on the information available in the intervention.

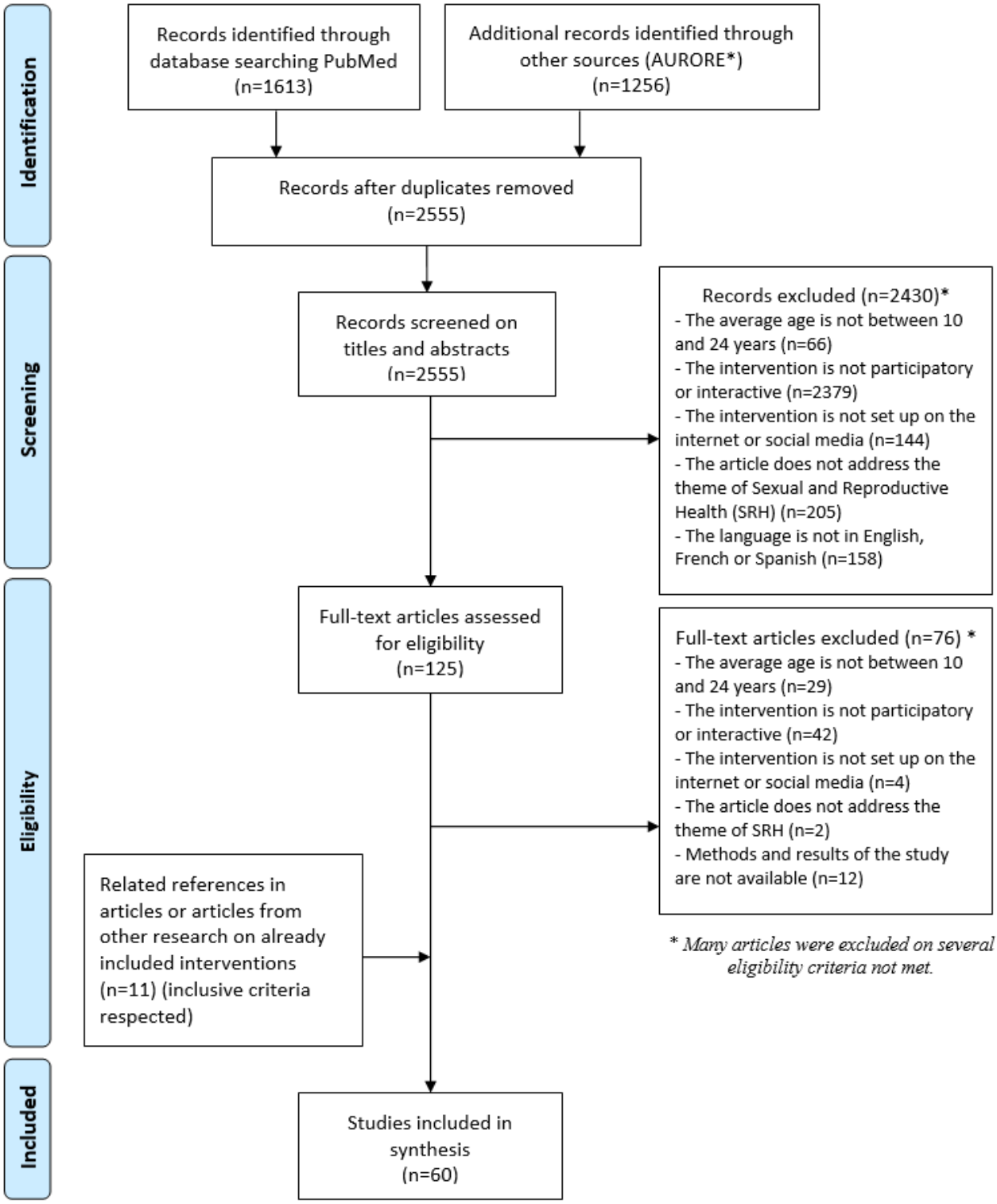
Results

The electronic search strategies used identified a total of 2555 references after removing duplicates. After selection based on title and abstract screening, the full text of 125 references was evaluated. After this inclusion phase, 49 articles describing 37 interventions were included. For each intervention included, we searched for other publications concerning it, and 11 additional studies were included, based on the references cited in the included articles. A total of 60 articles describing 37 interventions were included; several articles were published for the same intervention (Figure 1). The degree of interreader agreement for the final selection of the 60 articles was calculated with the κ coefficient and it was equal to 0.98. All the studies included in this systematic review are available in Multimedia Appendix 3. Descriptive data for the included studies and interventions are available in Table 1. Of the 60 articles included, 52% (n=31/60) were published in the last 5 years (Table 2).

Overall, 62% of the studies (n=36/58) were conducted in the United States. Of the types of studies, 45% (n=27/60) exclusively concerned process results, 22% (n=13/60) included process results and effectiveness results, 17% (n=10/60) exclusively had effectiveness results, and 17% (n=10/60) were exclusively protocol publications. Of the 37 interventions, 51% (n=19/37) addressed sexual health holistically. Overall, 51% (n=19/37) targeted a general population. In cases where specific populations were targeted (49%, n=18/37), 44% (n=8/18) were identified by their sexual orientation. In total, 65% (n=24/37) of all interventions were for both sexes, 22% (n=8/37) were for males only, and 11% (n=4/37) were for women only. The targeted population in terms of age was mainly individuals aged 10 to 24, strictly defined in 35% of the interventions (n=13/37). However, other studies had a less specific or different range of age targeted: aged 10 to 17 years, aged 10 to >24 years, aged 18 to 24 years, or aged 18 to >24 years; some studies simply referred to "students" or "youth." In total, 43% (n=16/37) used multiple recruitment methods.

XSL•FO
RenderX

Figure 1. Flow chart of the literature reviewing process. Aurore is a database of Institut National d'Études Démographiques (a French public research institute) that combines scientific databases and grey literature, allowing access to a range of databases and electronic journals.



XSL•FO
RenderX

Table 1. Description of the characteristics of the 60 articles and the 37 interventions.

Characteristics	Studies, n (%)
Characteristics of articles	
Year of publication (n=60)	
2006-2009	2 (3)
2010-2014	27 (45)
2015-2019	31 (52)
Study country (n=58; NI[a]=2)	
United States	36 (62)
Canada	1 (2)
United Kingdom	4 (7)
Netherlands	1 (2)
Europe (other)	2 (3)
Australia	3 (5)
Uganda	4 (7)
Brazil	2 (3)
Chile	2 (3)
Asia	3 (5)
Study objective (n=60)	
Process evaluation only	27 (45)
Process and effects evaluation in one article	13 (22)
Effects evaluation only	10 (17)
Protocol study only	10 (17)
Measure for evaluation[b] (n=60)	
Process evaluation (quantitative questionnaire)	21 (35)
Process evaluation (qualitative measure)	21 (35)
Effectiveness evaluation (quantitative questionnaire)	19 (32)
Effectiveness evaluation (qualitative measure)	3 (5)
Characteristics of interventions	
Target population[b] (n=37)	
General	19 (51)
Specific	18 (49)
Sexual orientation	8 (22)
Ethnic minorities	4 (11)
Others	7 (19)
Sex (n=37)	
Males and females	24 (65)
Males only	8 (22)
Females only	5 (14)
Age group (years; n=37)	
10 to 17	2 (5)
10 to 24	13 (35)

XSL•FO
RenderX

Characteristics	Studies, n (%)
10 to >24	8 (22)
18 to 24	4 (11)
18 to >24	8 (22)
Age not specified but considered as “students” or “youth”	2 (5)
Recruitment[b] (n=34; NI[a]=3)	
Social networking sites	12 (35)
Internet	11 (32)
Secondary schools	9 (26)
Community or youth organizations	8 (24)
Clinics	7 (21)
Universities	5 (15)
Email	4 (12)
Peers and word of mouth	3 (9)
Phone	2 (6)
Registers	1 (3)
Smartphone apps	1 (3)
Health educators	1 (3)
Incentives[b] (n=23; NI[a]=14)	
Yes	21 (91)
Direct remuneration	12 (52)
Gift card	10 (43)
Book or movie voucher	1 (4)
Points for lot	1 (4)
Raffle for remuneration	1 (4)
No	2 (8)
Theme (n=37)	
Sexual health promotion	19 (51)
HIV/sexually transmitted infection prevention specifically	12 (32)
Sexual violence prevention	3 (8)
Hepatitis B virus and hepatitis C virus testing promotion	1 (3)
Improve HIV care linkage	1 (3)
Observe peer influence in sexual situations only	1 (3)

[a]NI: no information in the article.

[b]For a given article (N=60) or an intervention (N=37), several entries are possible. Totals do not always equal 100%.

XSL•FO
RenderX

Table 2. Number of publications over time.

Year	Studies published, n
2006	1
2009	1
2010	3
2011	2
2012	8
2013	10
2014	4
2015	5
2016	9
2017	10
2018	6
2019 (January)	1

Descriptive data on the intervention types, online supports, and features are shown in Table 3 (for a description of each intervention, see Multimedia Appendix 4). Concerning intervention types, 41% (n=15/37) involve a dissemination of information with participatory components (game, quizzes, discussions). The medium used is a website in 54% (n=20/37) of cases, followed by online social networks (35%, n=13/37), with Facebook used in 22% (n=8/37) of cases. Furthermore, 14% (n=5/37) use several different online supports for the implementation of the intervention. To protect the identity of participants, 49% (n=18/37) of the interventions provide anonymity. Of these, 72% (n=13/18) allow participants to use personal identifiers, and 67% (n=12/18) use private websites. The interventions based on social networking sites do not mention anonymity because this is not possible on such sites. However, on Facebook, one (n=1) intervention used a secret group for greater confidentiality, another (n=1) used a private SMS text messaging system, and another (n=1) used a private page that only registered participants can access. Concerning participatory features, 68% allow interaction, either between peers (62%, n=23/37) or with a professional (43%, n=16/37). This interaction is mainly through online social networks (22%, n=8/37) and discussion forums (19%, n=7/37). Overall, 5% (n=2/37) use multiple supports for interaction. Involvement in a game-type activity was possible in 27% (n=10/37) of cases. Videos were broadcast in 54% (n=20/37) of cases. Finally, 43% (n=16/37) of the interventions were constructed from a theoretical model, with 19% (n=7/37) using the Information-Motivation-Behavioral Skills model.

XSL•FO
RenderX

Table 3. Intervention type, online support, and features description (N=37).

Variables	Studies, n (%)
Intervention type	
Information dissemination with participatory components (games, quizzes, discussions)	15 (41)
Online community/discussion only	11 (30)
Participation in activities only (including games)	6 (16)
Participatory educational session only	3 (8)
Personalized assistance	2 (5)
Online support for implementation[a]	
Website	20 (54)
Social networking sites	13 (35)
Online game only	5 (14)
Apps	4 (11)
Social networking sites used[a]	
Facebook	8 (22)
YouTube	3 (8)
MySpace	2 (5)
Twitter	1 (3)
Flickr	1 (3)
Tumblr	1 (3)
Instagram	1 (3)
WeChat	1 (3)
Not specified	1 (3)
Participatory features (1) - interactive part[a]	25 (68)
Interaction between peers and with professionals	14 (38)
Interaction between peers only	9 (24)
Interaction with professionals only	2 (5)
Peer leaders formation and implication	5 (14)
Section to ask a professional	5 (14)
Support for interaction (peers and professionals)[a]	
Social networking sites	8 (22)
Forum discussion	7 (19)
Blog	3 (8)
On website without more information	3 (8)
Chat	2 (5)
In the online game	2 (5)
Video comment section	1 (3)
On application	1 (3)
“Ask the expert” section	1 (3)
Participatory features (2) - involvement in an activity[a]	16 (43)
Online video game system	10 (27)
Interactive quiz	4 (11)
Personal goals	2 (5)

XSL•FO
RenderX

Variables	Studies, n (%)
Other features (3) - receipt of information[a]	
Video system	20 (54)
Transmission or link of existing websites	4 (11)
Theory model used for intervention conception[a]	
No	21 (57)
Yes	16 (43)
Information-Motivation-Behavioral Skills Model	7 (19)
Social Identity Theory	2 (5)
Social Cognitive Theory	2 (5)
Social Learning Theory	1 (3)
Others	9 (24)
Two or more theories used	5 (14)
Community-based participatory research	
Yes	21 (57)
Unspecified	16 (43)

[a]An intervention can use several theories or several supports and contain different functionalities. Totals are not always equal to 100%.

The five most commonly used behavior change techniques are as follows (Multimedia Appendix 5). First, 78% (n=29/37) of interventions introduce or define an environmental or social stimulus to encourage or guide behavior. Second, 78% (n=29/37) provide information on the health consequences of performing the behavior. Third, 73% (n=27/37) present information from a credible source in favor of or against the behavior. Fourth, 70% (n=26/37) organize and provide some form of social support within the intervention. Fifth, 65% (n=24/37) provide information on what others think about the behavior. No intervention provides punitive measures or remuneration for the conduct of the behavior sought.

Of the 37 interventions, 57% (n=21/37) indicate that they called on young people for community-based participatory research (collective construction). This takes various forms: 38% (n=14) of the interventions conducted focus groups to discuss the proposed intervention, 27% (n=10) directly included youth in the development of content, 8% (n=3) adapted their content based on feedback from young people in pretest studies, 5% (n=2) involved youth in the evaluation, and 3% (n=1) formed a youth advisory committee.

Data on the design and evaluation methods are available in Table 4. For a description of the methods of each intervention, see Multimedia Appendix 6. In total, 43% (n=16/37) were evaluated according to a randomized controlled trial (RCT) design. Overall, 22% (n=8/37) provided a follow-up between 1 and 2 years, while the remainder reported a follow-up shorter than 1 year (59%, n=22/37) or did not specify a follow-up time (19%, n=7/37). For process evaluation, 35% (n=13/37) did an acceptability study, 30% (n=11/37) did an attractiveness study, and 27% (n=10/37) assessed feasibility. Regarding effectiveness, 46% (n=17/37) of the interventions were subject to an outcome evaluation and 32% (n=12/37) had a planned outcome evaluation. Among the outcomes evaluated (conducted or planned evaluation), sexual behaviors were the most evaluated (38%, n=14/37), followed by condom use (29%, n=11/37) and sexual health knowledge (22%, n=8/37).

XSL•FO
RenderX

Table 4. Intervention design and evaluation methodology (N=37).

Study information	Studies, n (%)
Design study	
Randomized controlled trial (RCT)	16 (43)
Control group (NI=2)[a,b]	15 (41)
Information-only control website[b]	4 (11)
Before-after study (no RCT)	7 (19)
Cross-sectional study	3 (8)
Other design	8 (22)
Unspecified	3 (8)
Follow-up	
No follow-up	3 (8)
0.5-2 months	3 (8)
3-5 months	9 (24)
6-11 months	7 (19)
12-24 months	8 (22)
Unspecified	7 (19)
Process outcomes evaluated[c]	
Acceptability	13 (35)
Attractiveness	11 (30)
Feasibility	10 (27)
Satisfaction	3 (8)
Implementation	3 (8)
Outcomes evaluation conducted[c]	17 (46)
Behaviors	10 (27)
Condom use, condom use intention, self-efficacy toward condom use, and attitude toward condom use	9 (24)
Attitudes	4 (11)
Communication	3 (8)
Knowledge	3 (8)
Behavioral skills	2 (5)
Self-efficacy	2 (5)
Contraception use	1 (3)
History of sexually transmitted infections	1 (3)
HIV stigma	1 (3)
HIV test history (date and result of the last test)	1 (3)
Incidence of sexually transmitted infections	1 (3)
Intentions related to risky sexual activity	1 (3)
Internalized homophobia	1 (3)
Intimate partner violence	1 (3)
Motivation	1 (3)
Pubertal development	1 (3)
Sexual abstinence	1 (3)
Waiting before having sex	1 (3)

XSL•FO
RenderX

Study information	Studies, n (%)
Other outcomes evaluated only once	17 (46)
Outcomes evaluation planned[c]	12 (32)
Knowledge	5 (14)
Behaviors	4 (11)
Condom use	2 (5)
Intentions	2 (5)
Self-efficacy	2 (5)
Occurrence of pregnancy	1 (3)
Occurrence of sexually transmitted infections	1 (3)
Self-reported pregnancy	1 (3)
Self-reported sexually transmitted infections	1 (3)
Fertility distress	1 (3)
Repeat HIV/sexually transmitted infection screening	1 (3)
Number of tests for *Chlamydia trachomatis*	1 (3)
HBsAg and anti–hepatitis C virus IgG test uptake	1 (3)
HIV-related care engagement	1 (3)
Motivation	1 (3)
Number of partners	1 (3)
Sexual communication self-efficacy	1 (3)
Use of safety strategies	1 (3)
Viral suppression	1 (3)
Other outcomes planned for evaluation only once	7 (19)
Unspecified outcomes evaluation	8 (22)

[a]NI: no information in the article.

[b]Since a control group can also be a group receiving an informational website only, the total exceeds the number of RCTs.

[c]An intervention can evaluate several outcomes or process components. Totals are not always equal to 100%.

Discussion

Principal Results

Our review identified 37 different interventions, which were the subjects of 60 articles. The number of online participatory interventions for the promotion of young people's sexual health has increased significantly over the past 5 years, especially in the United States. Three key points drew our attention: (1) Several different online supports are used by interventions and we would recommend adapting these to young people's preferences; (2) Online peer interaction is the participatory element most often used in interventions and is a promising health promotion approach; (3) In view of the limited number of effectiveness evaluations, it is necessary to define a conceptual model of interventions to enable comprehensive and rigorous evaluation and to understand the effect of peer interaction and participatory components.

How to Adapt to the Favorite Media of Young People?

Concerning the online support used, interventions are mainly first developed on websites. The second most popular medium is social networks, with Facebook dominating, as already shown in a previous review of social networking sites [18].

Surprisingly, young people's favorite social networks [27] are rarely used. Only one intervention was on Instagram [28], three were on YouTube, and none were on Snapchat. However, these three media have been described as the new preferred ones of youth, whereas the popularity of Facebook is declining [27]. The future challenge for researchers will be to develop interventions that can evolve with young people’s preferences, keeping up with rapid generational changes. In our review, few interventions use more than one online medium. One option would be to use a multichannel approach for interventions. Such an approach already exists to some extent in the American intervention “weCare,” which allows young people to choose how they connect with educators, with three possible contact modalities: Facebook Messenger, SMS text messaging, and app-based instant messages [29].

Our findings also highlight the need to design interventions adapted to the uses, languages, interests, and realities of young people, particularly through interactive and playful components. One way to remain close to the interests of young people is to

XSL•FO
RenderX

integrate promising new media in interventions, such as videos and games. It is also possible to allow users to insert their own content or to customize websites. Integrating attractive components that are correctly implemented will ensure better group retention. To know what is preferred by young people, it is therefore necessary to have measures of attractiveness. This review has cited different measures: online media usage, process data (number of visits, time spent, and interaction rate), technical recommendations, content adapted to the target audience (specificity and age), satisfaction, points of view, and involvement of participants (especially sexual minorities).

Web-based interventions also raise the challenges of security, privacy, and anonymity. For example, the lower use of social networking sites for research compared to websites may also be due to the fact that the ownership of the data from youth participation belongs to these media. This data would be less easy to protect in terms of security, confidentiality, and privacy, especially against cyberstalking, requiring moderation at all times. In the studies reviewed here, authors provided little information on how they protected participants' data. On social networking sites, some researchers use closed groups to control the exchange of participants' data. Others host the data through a secure external website. Technical partners, such as social networking sites, are bound by specific laws and contractual data protection clauses, and there is a clear regulatory framework for many countries [30]. As noted by some authors [11,31,32], ethical and data security frameworks need to be strengthened. For example, the importance of blocking public access to online interventions and developing powerful security features is underlined [33]. Concerning anonymity, protection of the identity of participants is possible mainly on private websites, which is especially important in the context of sexual health, where the internet is used to avoid embarrassment and overcome privacy issues [34].

How to Implement Peer Dynamics in Interventions?

All media can be used to disseminate information among young people, either top-down (from an educator to a young person) or cross-functionally (between peers). The interest of the 37 interventions assessed here rests on their participatory activities, of which peer interaction is the most frequent component.

Peer exchanges were described in different ways: counselling, experience-sharing, community involvement, personal stories, self-help, and peer support. Peers were considered not only as participants, but also as peer educators (opinion leaders) previously trained by professionals [35-37]. In one study, the potential for sharing and comparing real experiences was supported [38], with an expressed need for sharing experiences among peers. Participants also expressed the desire for social interaction online with other young people [39].

More personalized approaches better target the concerns of each individual, as seen in the Media Aware [40] and Queer Sex Ed [41] interventions (individuals' goals). Participants could also disseminate their own content, as seen in the HealthMpowerment intervention [42-44]. Peer dynamics also occur when young people are directly involved in the community-based participatory research process, especially in sexuality education programs [10]. This process can validate the role of community members and academics as equitable partners [45]. In our review, we determined that this process is widely used at the design stage. Peer interaction is thus enabled by most interventions and is described as strengthening an intervention's capacity to change behaviors, even if professionals are involved. The dynamics between peers, and the feeling of being "between young people," are seen as potentialities. Surprisingly, the term "peer education" is not a term used in the reviewed articles. "Peer education" is actually an exchange of experiences and information between peers in "real life," integrating the notion of "shared education" [46], and is thus well suited to these interventions. One intervention did use the term "peer-led" [35]. Peer dynamics are little conceptualized by the authors, and a model for designing and evaluating interventions is lacking.

How to Evaluate Interventions?

The objective of interventions is to change sexual health outcomes positively. For the moment, although experimental plans are defined, publications focus more on intervention processes than effectiveness in terms of health outcomes. This probably reflects the need to identify implementation problems beforehand, as a lack of effect may reflect a failure in implementation rather than the ineffectiveness of the intervention [26]. Implementing an intervention correctly will ensure better group retention. To evaluate effectiveness, the randomized controlled trial remains the most widely used or planned design. It does not preclude assessing the effect of an intervention on a range of outcome measures [47].

In interventions dealing with evaluation, behaviors were most often the main outcome, followed by knowledge, self-efficacy, and attitudes. A majority of follow-up interventions lasted less than 1 year. Nonetheless, it would be interesting to have a long-term follow-up to determine whether short-term changes persist [21]. Behavior measures are based on self-reported data, and many authors have highlighted the issue of social desirability bias as a limitation [36,40,41,48-51].

Our review found few plans to observe a robust indicator, such as STI incidence [52], HIV-related care engagement and viral suppression [29], or pregnancy [53,54]. These indicators can measure the real impact of an intervention on sexual health. Nevertheless, this requires a large sample size in order to have sufficient power to detect the effects of the intervention, especially when the expected outcomes have a low baseline rate of incidence (eg, HIV incidence), unless these studies are conducted on high-risk groups.

In this context of complex intervention, mechanisms of action should be identified and interventions should rely on a theoretical, conceptual, and operational model. This will enable all the participatory, social, and collective variables involved in the process to be measured and validated. Based on a literature review, Borek and Abraham developed a conceptual model of mechanisms of change in small groups [55]. For peer interventions, Simoni et al [56] argue for a strong theoretical framework to support behavior promotion, link to outcomes, and justify peer inclusion. In addition, strategies combining several theories and concepts may have a greater effect [57], as seen in the TeensTalkHealth intervention [58], which used the IMB model [7] combined with communication theory [59].

XSL•FO
RenderX

Several interactive processes (group development, group dynamics, social change) have been highlighted and could be used for the constitution and animation of social groups [55]. Finally, applying a comprehensive model of internet-based peer education (or peer-led behavior change) for sexual health is a promising approach, as long as a proliferation of concept and theoretical models does not occur. Rigorous methods, such as the 5 steps of the Intervention Mapping protocol, can contribute to the development of more effective behavior change interventions and methods of evaluation, assessing all stages of adoption, implementation, and sustainability of the intervention [60,61].

Limitations

Our review was conducted with a cross-validation methodology based on two search tools (PubMed and Aurore), but we cannot rule out that some interventions escaped our research. Participatory or interactive interventions may exist but may not be evaluated and published (for example, the website Sex, Etc [62]). Finally, wide variations in interventions made it inappropriate to synthesize the results using a meta-analysis.

Conclusions

This review describes existing interventions in participatory sexuality education for young people on the internet. It aims to provide guidance for interventions that meet the expectations of national and international strategies on youth sexuality education. Identified interventions are deployed on many internet media and have shown their feasibility, practical interest, and attractiveness. However, they are still in the early stages of design and evaluation, particularly as regards the effect of peer interaction, and do not always adhere to existing theoretical models. We recommend building a conceptual, theoretical, and evaluation model for community-based interventions involving peer interaction and participation in activities, providing the necessary operational and evaluative tools. Interventions must be designed with regard to media multiplicity, youth populations (orientations, gender identities), and a holistic sexual health approach. To improve these interventions, we recommend having a more participatory approach, involving young people in the whole process, including the design phase.

Acknowledgments

First, we would like to thank the National Association for Research and Technology, which funded this project. We would like to thank all members of the Institut National de la Santé Et de la Recherche Médicale (Inserm) Unit 1123 and Ined-Inserm-Univ. Paris-Sud-UVSQ (Université Versailles Saint-Quentin) Unit 14 teams for their support of our work. We also thank the following librarians for their research articles and methodological support: Catherine Sluse (Ined), Catherine Le Huu Nho (Paris Diderot University), and Fabienne Warin (Paris Diderot University). We also thank all Ined members for their feedback in workshops. We would like to thank our translator Duncan Fulton for proofreading the article. Finally, we thank all the authors included in our review who sent us their publications. and the joint authors of this article, according to their respective roles: PM for all stages of the project, EDLR and CA for participation in the design of the project, LC for his reading of the journals, and all authors for analyzing results and writing the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist items pertaining to the content of a systematic review and meta-analysis.
[DOCX File , 17 KB - jmir_v22i7e15378_app1.docx]

Multimedia Appendix 2
Research strategies used for PubMed research and Aurore complementary research.
[DOCX File , 13 KB - jmir_v22i7e15378_app2.docx]

Multimedia Appendix 3
All the studies included in the systematic review.
[DOCX File , 25 KB - jmir_v22i7e15378_app3.docx]

Multimedia Appendix 4
Description of the interventions included and their participatory components.
[DOCX File , 92 KB - jmir_v22i7e15378_app4.docx]

Multimedia Appendix 5
Coding of Michie's taxonomy on Behaviour Change Techniques.
[DOCX File , 24 KB - jmir_v22i7e15378_app5.docx]

XSL•FO
RenderX

Multimedia Appendix 6
Description of intervention studies, designs, and evaluation methods.
[DOCX File , 30 KB - jmir_v22i7e15378_app6.docx]

References

1. Kar SK, Choudhury A, Singh AP. Understanding normal development of adolescent sexuality: A bumpy ride. J Hum Reprod Sci 2015;8(2):70-74 [FREE Full text] [doi: 10.4103/0974-1208.158594] [Medline: 26157296]
2. Dehne K, Riedner G. Sexually Transmitted Infections among adolescents - The need for adequate health services. Switzerland: World Health Organization and Deutsche Gesellschaft fuer Technische Zusammenarbeit (GTZ); 2005:1-90.
3. Sedgh G, Finer LB, Bankole A, Eilers MA, Singh S. Adolescent pregnancy, birth, and abortion rates across countries: levels and recent trends. J Adolesc Health 2015 Feb;56(2):223-230. [doi: 10.1016/j.jadohealth.2014.09.007] [Medline: 25620306]
4. Jatlaoui TC, Shah J, Mandel MG, Krashin JW, Suchdev DB, Jamieson DJ, et al. Abortion Surveillance - United States, 2014. MMWR Surveill Summ 2018 Nov 23;66(25):1-44 [FREE Full text] [doi: 10.15585/mmwr.ss6625a1] [Medline: 30462631]
5. World Health Organization. Introducing WHO's reproductive health guidelines and tools into national programmes: Principles and processes of adaptation and implementation. Word Health Organization 2007:1-36 [FREE Full text]
6. Haut Conseil de la Santé Publique. Stratégie nationale de santé sexuelle: Agenda 2017-2030. In: HCSP, coll. Avis et Rapports. Paris: HCSP; 2017:1-75.
7. Fisher JD, Fisher WA, Shuper PA. The information-motivation-behavioral skills model of HIV preventive behavior. In: DiClemente RJ, Crosby RA, Kegler MC, editors. Emerging theories in health promotion practice and research. San Francisco, CA, USA: Jossey Bass; 2002:40-70.
8. Institute of Medicine (US) Committee on Health Literacy. Health Literacy: A Prescription to End Confusion. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Institute of Medicine (US). Washington, DC, USA: National Academies Press (US); 2004.
9. WHO Europe. Implementing the action plan for sexual and reproductive health – how policies can make a real difference. WHO regional office for Europe. Denmark: WHO regional office for Europe; 2018 Sep 10. URL: https://tinyurl.com/y2lh7uy2 [accessed 2018-10-17]
10. WHO Regional Office for Europe, Federal Centre for Health Education, BZgA. Standards for Sexuality Education in Europe - A framework for policy makers, educational and health authorities and specialists. Federal Centre for Health Education, BZgA. Cologne: WHO Collaborative center for Sexual and Reproductive Health; 2010. URL: https://www.bzga-whocc.de/fileadmin/user_upload/WHO_BZgA_Standards_English.pdf [accessed 2019-04-12]
11. Levine D. Using Technology, New Media, and Mobile for Sexual and Reproductive Health. Sex Res Soc Policy 2011 Feb 25;8(1):18-26. [doi: 10.1007/s13178-011-0040-7]
12. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C. Sexual health promotion for young people delivered via digital media: a scoping review. In: Public Health Research. Southampton (UK): NIHR Journals Library; Nov 2015.
13. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
14. Higgins O, Sixsmith J, Barry M, Domegan C. A literature review on health information-seeking behaviour on the web: a health consumer and health professional perspective. Insights into health communication. In: European Centre for Disease Prevention and Control. Sweden: European Centre for Disease Prevention and Control; 2011.
15. Hayat TZ, Brainin E, Neter E. With Some Help From My Network: Supplementing eHealth Literacy With Social Ties. J Med Internet Res 2017 Mar 30;19(3):e98 [FREE Full text] [doi: 10.2196/jmir.6472] [Medline: 28360024]
16. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. "Friending" teens: systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015 Jan 05;17(1):e4 [FREE Full text] [doi: 10.2196/jmir.3692] [Medline: 25560751]
17. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
18. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016 Sep 19;9(1):32193 [FREE Full text] [doi: 10.3402/gha.v9.32193] [Medline: 27649758]
19. DeSmet A, Shegog R, Van Ryckeghem D, Crombez G, De Bourdeaudhuij I. A Systematic Review and Meta-analysis of Interventions for Sexual Health Promotion Involving Serious Digital Games. Games Health J 2015 Apr;4(2):78-90. [doi: 10.1089/g4h.2014.0110] [Medline: 26181801]
20. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019 Apr;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
21. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C. Evidence on effectiveness of digital interventions for sexual health for young people Internet. Public Health Research 2015 Nov;3:1-154 [FREE Full text] [doi: 10.3310/phr03130]

XSL•FO
RenderX

22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. World Health Organization. Adolescent development. World Health Organization. 2017. URL: http://www.who.int/maternal_child_adolescent/topics/adolescence/dev/en/ [accessed 2017-12-12]
24. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478. [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
25. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
26. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
27. Statista. Most popular social networks of teenagers in the United States from fall 2012 to fall 2019. statista.com. 2019. URL: https://www.statista.com/statistics/250172/social-network-usage-of-us-teens-and-young-adults/ [accessed 2019-02-04]
28. O'Donnell NH, Willoughby JF. Photo-sharing social media for eHealth: analysing perceived message effectiveness of sexual health information on Instagram. J Vis Commun Med 2017 Oct;40(4):149-159. [doi: 10.1080/17453054.2017.1384995] [Medline: 29022412]
29. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: A Social Media-Based Intervention Designed to Increase HIV Care Linkage, Retention, and Health Outcomes for Racially and Ethnically Diverse Young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]
30. Greenleaf G. Global Data Privacy Laws: 89 Countries, and Accelerating. Privacy Laws & Business International Report, Special Supplement 2012 Feb(115):1-14 [FREE Full text]
31. Bull SS, Breslin LT, Wright EE, Black SR, Levine D, Santelli JS. Case study: An ethics case study of HIV prevention research on Facebook: the Just/Us study. J Pediatr Psychol 2011;36(10):1082-1092 [FREE Full text] [doi: 10.1093/jpepsy/jsq126] [Medline: 21292724]
32. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
33. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
34. Park E, Kwon M. Health-Related Internet Use by Children and Adolescents: Systematic Review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
35. Sun WH, Wong CKH, Wong WCW. A Peer-Led, Social Media-Delivered, Safer Sex Intervention for Chinese College Students: Randomized Controlled Trial. J Med Internet Res 2017 Dec 09;19(8):e284 [FREE Full text] [doi: 10.2196/jmir.7403] [Medline: 28793980]
36. Rice E, Tulbert E, Cederbaum J, Barman AA, Milburn NG. Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention. Health Educ Res 2012 Apr;27(2):226-236 [FREE Full text] [doi: 10.1093/her/cyr113] [Medline: 22247453]
37. Danielson CK, McCauley JL, Jones AM, Borkman AL, Miller S, Ruggiero KJ. Feasibility of delivering evidence-based HIV/STI prevention programming to a community sample of African American teen girls via the internet. AIDS Educ Prev 2013 Oct;25(5):394-404 [FREE Full text] [doi: 10.1521/aeap.2013.25.5.394] [Medline: 24059877]
38. Franck LS, Noble G. Here's an idea: ask the users! Young people's views on navigation, design and content of a health information website. J Child Health Care 2007 Dec;11(4):287-297. [doi: 10.1177/1367493507083941] [Medline: 18039731]
39. McCarthy O, Carswell K, Murray E, Free C, Stevenson F, Bailey JV. What young people want from a sexual health website: design and development of Sexunzipped. J Med Internet Res 2012;14(5):e127 [FREE Full text] [doi: 10.2196/jmir.2116] [Medline: 23060424]
40. Scull TM, Kupersmidt JB, Malik CV, Keefe EM. Examining the efficacy of an mHealth media literacy education program for sexual health promotion in older adolescents attending community college. J Am Coll Health 2018 Apr;66(3):165-177. [doi: 10.1080/07448481.2017.1393822] [Medline: 29068772]
41. Mustanski B, Greene GJ, Ryan D, Whitton SW. Feasibility, acceptability, and initial efficacy of an online sexual health promotion program for LGBT youth: the Queer Sex Ed intervention. J Sex Res 2015;52(2):220-230. [doi: 10.1080/00224499.2013.867924] [Medline: 24588408]
42. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011 Feb;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]

XSL•FO
RenderX

43. Bauermeister JA, Muessig KE, LeGrand S, Flores DD, Choi SK, Dong W, et al. HIV and Sexuality Stigma Reduction Through Engagement in Online Forums: Results from the HealthMPowerment Intervention. AIDS Behav 2019 Mar;23(3):742-752 [FREE Full text] [doi: 10.1007/s10461-018-2256-5] [Medline: 30121727]
44. Barry MC, Threats M, Blackburn NA, LeGrand S, Dong W, Pulley DV, et al. "Stay strong! keep ya head up! move on! it gets better!!!!": resilience processes in the healthMpowerment online intervention of young black gay, bisexual and other men who have sex with men. AIDS Care 2018 Aug;30(sup5):S27-S38 [FREE Full text] [doi: 10.1080/09540121.2018.1510106] [Medline: 30632775]
45. Israel BA, Schulz AJ, Parker EA, Becker AB, Community-Campus Partnerships for Health. Community-based participatory research: policy recommendations for promoting a partnership approach in health research. Educ Health (Abingdon) 2001;14(2):182-197. [doi: 10.1080/13576280110051055] [Medline: 14742017]
46. Svenson GR. European guidelines for youth AIDS peer education. Department of Community Medicine (Samhällsmedicinska institutionen, Lund University) 1998:1-56 [FREE Full text]
47. Rosen L, Manor O, Engelhard D, Zucker D. In defense of the randomized controlled trial for health promotion research. Am J Public Health 2006 Jul;96(7):1181-1186. [doi: 10.2105/AJPH.2004.061713] [Medline: 16735622]
48. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
49. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A Randomized Trial of an Online Risk Reduction Intervention for Young Black MSM. AIDS Behav 2019 May;23(5):1166-1177. [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
50. Bull S, Nabembezi D, Birungi R, Kiwanuka J, Ybarra M. Cyber-Senga: Ugandan youth preferences for content in an internet-delivered comprehensive sexuality education programme. East Afr J Public Health 2010 Mar;7(1):58-63 [FREE Full text] [Medline: 21413574]
51. Greene GJ, Madkins K, Andrews K, Dispenza J, Mustanski B. Implementation and Evaluation of the Keep It Up! Online HIV Prevention Intervention in a Community-Based Setting. AIDS Educ Prev 2016 Jun;28(3):231-245. [doi: 10.1521/aeap.2016.28.3.231] [Medline: 27244191]
52. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and Behavioral Outcomes of Keep It Up!: An eHealth HIV Prevention Program RCT. Am J Prev Med 2018 Aug 14;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]
53. Nicholas A, Bailey JV, Stevenson F, Murray E. The Sexunzipped trial: young people's views of participating in an online randomized controlled trial. J Med Internet Res 2013;15(12):e276 [FREE Full text] [doi: 10.2196/jmir.2647] [Medline: 24334198]
54. Nielsen A, De Costa A, Bågenholm A, Danielsson KG, Marrone G, Boman J, et al. Trial protocol: a parallel group, individually randomized clinical trial to evaluate the effect of a mobile phone application to improve sexual health among youth in Stockholm County. BMC Public Health 2018 Dec 05;18(1):216 [FREE Full text] [doi: 10.1186/s12889-018-5110-9] [Medline: 29402241]
55. Borek AJ, Abraham C. How do Small Groups Promote Behaviour Change? An Integrative Conceptual Review of Explanatory Mechanisms. Appl Psychol Health Well Being 2018 Mar;10(1):30-61. [doi: 10.1111/aphw.12120] [Medline: 29446250]
56. Simoni JM, Franks JC, Lehavot K, Yard SS. Peer interventions to promote health: conceptual considerations. Am J Orthopsychiatry 2011 Jul;81(3):351-359 [FREE Full text] [doi: 10.1111/j.1939-0025.2011.01103.x] [Medline: 21729015]
57. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
58. Brady SS, Sieving RE, Terveen LG, Rosser BRS, Kodet AJ, Rothberg VD. An Interactive Website to Reduce Sexual Risk Behavior: Process Evaluation of TeensTalkHealth. JMIR Res Protoc 2015 Sep 02;4(3):e106 [FREE Full text] [doi: 10.2196/resprot.3440] [Medline: 26336157]
59. Fishbein M, Cappella J. The Role of Theory in Developing Effective Health Communications. J Commun 2006;56(s1):S1-S17. [doi: 10.1111/j.1460-2466.2006.00280.x]
60. Kok G, Schaalma H, Ruiter RAC, van EP, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
61. Kok G, Mesters I. Getting inside the black box of health promotion programmes using intervention Mapping. Chronic Illn 2011 Sep;7(3):176-180. [doi: 10.1177/1742395311403013] [Medline: 21900338]
62. Sex, etc. URL: https://sexetc.org [accessed 2020-07-17]

Abbreviations

AYA: adolescents and young adults
BCT: behavior change techniques
IMB: Information-Motivation-Behavioral Skills
INED: Institut National d'Études Démographiques

XSL•FO
RenderX

RCT: randomized controlled trial
STI: sexually transmitted infection
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.07.19; peer-reviewed by S Brady, E Neter; comments to author 02.09.19; revised version received 06.11.19; accepted 22.02.20; published 31.07.20.

Please cite as:
Martin P, Cousin L, Gottot S, Bourmaud A, de La Rochebrochard E, Alberti C
Participatory Interventions for Sexual Health Promotion for Adolescents and Young Adults on the Internet: Systematic Review
J Med Internet Res 2020;22(7):e15378
URL: http://www.jmir.org/2020/7/e15378/
doi:10.2196/15378
PMID:32735217

©Philippe Martin, Lorraine Cousin, Serge Gottot, Aurelie Bourmaud, Elise de La Rochebrochard, Corinne Alberti. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceived Effectiveness and Motivations for the Use of Web-Based Mental Health Programs: Qualitative Study

Heidi Eccles[1], MSc; Molly Nannarone[1], MPH; Bonnie Lashewicz[2], PhD; Mark Attridge[3], PhD; Alain Marchand[4,5], PhD; Alice Aiken[6], PhD; Kendall Ho[7], MD; Jianli Wang[1,8,9,10], PhD

[1]Work & Mental Health Research Unit, The Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada
[2]Department of Community Health Sciences,, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[3]Attridge Consulting, Inc, Minneapolis, MN, United States
[4]School of Industrial Relations, University of Montreal, Montreal, QC, Canada
[5]Public Health Research Institute, University of Montreal, Montreal, QC, Canada
[6]Office of Research and Innovation, Dalhousie University, Halifax, NS, Canada
[7]Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada
[8]Shandong Key Laboratory of Behavioral Medicine, School of Mental Health, Jining Medical University, Jining, China
[9]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
[10]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Jianli Wang, PhD
Work & Mental Health Research Unit
The Institute of Mental Health Research
University of Ottawa
Room 5404
1145 Carling Ave
Ottawa, ON, K1Z7K4
Canada
Phone: 1 6137226521
Email: jianli.wang@theroyal.ca

Abstract

Background: The prevalence of depression is high and has been stable despite increased treatment, research, and dissemination. People encounter barriers to seeking traditional mental health services, which could be mitigated by using web-based prevention methods.

Objective: This study aims to understand what people at high risk for depression perceive as effective aspects of web-based mental health programs and what motivates people at high risk for depression to use web-based mental health programs.

Methods: We conducted an inductive content analysis using telephone interview data from 77 participants at high risk for depression who were recruited from 2 randomized controlled trials (RCTs). Participants from the first RCT were working men who had been randomly assigned to 1 of the following 3 groups: control group, who had access to general depression information from a website called BroMatters; intervention group 1, who had access to the BroMatters website along with the associated BroHealth web-based mental health program; and intervention group 2, who had access to the BroMatters website, the BroHealth web-based mental health program, and telephone sessions with a life coach. Participants from the second RCT were men and women who had been assigned to the intervention group, who received access to the HardHat web-based mental health program, or the control group, who only received access to the HardHat web-based mental health program following completion of the RCT. Participants for this inductive content analysis study were recruited from the intervention groups in both RCTs. Two groups of participants (n=41 and n=20) were recruited from the BroHealth RCT, and a third group comprised 16 participants that were recruited from the HardHat RCT.

Results: We generated four categories regarding the perceived effectiveness of web-based programs and five categories related to what motivates the use of web-based programs. Participants identified awareness, program medium and functionality, program content, and coaches as categories related to the effectiveness of the programs. Categories of motivators to use web-based programs included providing reminders or incentives, promotion of the programs, providing appropriate medium and functionality, appropriate

XSL•FO
RenderX

content, and perceived need. The final category related to motivators reflects perceptions of participants who were either unsure about what motivates them or believed that there is no way to motivate use.

Conclusions: Conflicting evidence was obtained regarding the perceived effectiveness of aspects of the content and functionality of web-based programs. In general, web-based mental health programs were perceived to help increase mental health awareness, especially when it includes live access to a coach. However, the results also revealed that it is difficult to motivate people to begin using web-based mental health programs. Strategies that may motivate the use of such programs include perceived personal need, effective promotion, providing incentives and reminders, and improving functionality.

(J Med Internet Res 2020;22(7):e16961) doi:10.2196/16961

KEYWORDS

prevention; mental health; depression; cognitive behavioral therapy; motivators; perceived effectiveness; internet-based intervention; interview

Introduction

Background

Depression is a prevalent mental health disorder [1], is associated with an increased risk of suicide [2], and has comorbidities with multiple physical health conditions such as cardiovascular diseases [3], diabetes [4], and fibromyalgia [5]. Depression also impacts the economy, costing employers US $44 billion every year in the United States [6]. Globally, workplace stress can cost between US $220 million and US $200 billion per country [7]. Despite the growth in treatments for depression [8-10], the rate of depression has remained stable for more than 10 years [11,12]. Thus, treatment alone is unlikely to substantially reduce the prevalence and incidence of depression [13]; instead, a focus on prevention is necessary. Indeed, in a recent Alberta-based research project examining views from a range of mental health stakeholders, preventative measures were identified as a top priority for future depression research [14,15]. Preventing the occurrence or delaying the onset of depression can decrease symptoms [16] and reduce the burden for employers [17].

Barriers to seeking help related to treating and preventing depression are numerous and include stigma and availability of and access to services [18,19]. Web-based methods for treating and preventing depression have the potential to mitigate these barriers; in fact, web-based programs have been shown to be as effective as traditional methods [20-22] and web-based education tools [23]. Facilitators reported in the literature about web-based mental health programs include ease of website navigation and user characteristics such as curiosity about web-based mental health programs [24]. Web-based methods may be especially beneficial in rural communities where self-sufficiency is highly valued and access to therapeutic treatment may be particularly limited [25]. However, the usage of web-based mental health programs remains low. A study comparing face-to-face and web-based program usage reported completion rates of 84% and 65%, respectively [26]. Other studies of web-based mental health programs offered participants access to the program beyond the study time frame and reported that approximately 51% of participants wished to have continued access. Of those given continued access, only 38% went on to log on to the program [27].

People are often concerned about privacy and confidentiality for fear of being stigmatized for accessing mental health services [28]. These concerns may leave people reluctant to use web-based mental health programs. Furthermore, given the relative novelty of web-based mental health programs, many people are skeptical of program effectiveness for reducing or preventing depression compared with face-to-face therapies and, correspondingly, are less likely to use such programs [28,29]. However, few studies have investigated how users perceive web-based mental health programs for depression prevention [30].

Objectives

In view of user skepticism and the need for a fuller understanding of user perceptions of web-based mental health programs, our purpose is to identify factors that motivate use (initiation and adherence) of such programs and the perceived effectiveness of the programs. We achieve our purpose by analyzing data drawn from qualitative interviews conducted with a sample of web-based mental health program users who are at high risk of developing depression.

Methods

Research Design

This qualitative study complements 2 randomized controlled trials (RCTs). The RCTs were aimed at understanding strategies for preventing depression and evaluating user perceptions of the effectiveness of web-based mental health programs. This qualitative component elaborates on the understanding derived through the RCT studies by illuminating what users perceive as effective about web-based mental health programs and what motivates the use of such programs. In using a qualitative approach, we align ourselves with researchers who highlight the power of qualitative research to support contextual understanding of complex mental health phenomena [31] and to engage users in determining how to implement findings with respect to mental health in real-life settings [32].

We studied the perceived effectiveness of web-based mental health programming and what motivates the use of web-based mental health programs, and we treated these topics as emerging and yet to be coherently understood. Accordingly, we employed an inductive content analysis approach [33] of data collected through telephone interviews. Although we had some

XSL•FO
RenderX

predetermined goals, our study was inductive in 3 important ways: (1) although we had always planned to study user perceptions of web-based mental health program effectiveness, our decision to study what motivates the use of web-based mental health programs emerged as we noted trends of nonuse among our RCT sample; (2) correspondingly, we selected participants because they were nonusers; and (3) the findings we present were derived from the data and not in accordance with predetermined categories.

Below, we outline the broader RCT studies from which participants for this inductive content analysis study were drawn. We follow this with a description of our qualitative sample, data collection, and data analysis, including trustworthiness measures.

Broader RCT Study Context: BroHealth and HardHat

BroHealth RCT

The BroHealth RCT was about participants' perceptions of the effectiveness of the BroHealth web-based program, which was designed to prevent depression in working men who are at high risk of having a major depressive episode (MDE) in the next 4 years. BroHealth includes modules that provide information about stress and depression, web-based self-checks that help monitor symptoms, and a variety of self-help components such as cognitive behavioral therapy (CBT) and goal setting and tracking. Qualified coaches were also available to participants for 40-min sessions to a maximum of once per week. Coaches help participants set goals and help them work through problems. The study protocol from which BroHealth was developed is available elsewhere [34].

BroHealth RCT participants were recruited across Canada using the random digit dialing method. Participants were men who were (1) older than 18 years, (2) working for pay, and (3) at high risk for an MDE determined by the depression risk calculator developed and validated by Wang et al [35] but who had not had an MDE in at least two months. Participants were also required to be able to communicate in English or French and have access to a telephone and the internet. BroHealth RCT participants were interviewed at baseline, at 6 months, and at 12 months, and it was after the baseline interview that participants were randomized into the control group or into 1 of the 2 intervention groups:

1. Control group: participants received general information about depression through the BroMatters website
2. Intervention group 1: participants received general information about depression through the BroMatters website as well as access to the associated BroHealth web-based program
3. Intervention group 2: participants received general information about depression through the BroMatters website, access to the BroHealth program, and access to a qualified coach with whom they could schedule telephone sessions.

Every other week, participants received reminders via email to encourage their engagement with the BroHealth program. Participants received incentives (about Can $25 [US $19] gift cards) for completing each interview, but they did not receive any incentive for completing the web-based program.

HardHat RCT

The HardHat RCT was about participants' perceptions of the effectiveness of the HardHat web-based program designed to prevent depression in both men and women who are at high risk of having an MDE in the next 4 years. The HardHat program is coach assisted and comprises 5 compulsory and 4 supplementary sessions, which include short homework and in-class assignments. The 5 compulsory sessions had to be completed sequentially, and finishing one would unlock the next session. The 4 supplementary sessions could be accessed at any point in time. The sessions entail problem-solving therapy and components of CBT. Coaches had a minimum of a Bachelor of Science degree and had attended an intensive 2-day training session provided by a psychiatrist with CBT expertise. Coaches answered questions, guided users through the program, and provided feedback on homework and assignments. Coaches met weekly to discuss the participants' cases and any issues encountered during coaching. Coaches also met with the psychiatrist on a quarterly basis.

HardHat RCT participants were recruited through a partnership with Green Shield Canada, a not-for-profit Canadian health and dental insurance company. An advertisement and link were posted on the Green Shield Canada health portal. Prospective participants could click the link and complete the eligibility questionnaire, which included a depression risk assessment using the same risk calculator used with participants in the BroHealth RCT [34]. Participants were men and women who were aged 18 years or older and at high risk of having an MDE in the next 4 years or had experienced an MDE in the past but who had not experienced symptoms for at least two months. Participants also needed to be able to communicate in English or French and have access to a telephone and the internet. HardHat RCT participants were interviewed at baseline, at 3 months, and at 6 months. Following the baseline interview, participants were randomized to the wait-list control group or the intervention group:

1. Wait-list control group: participants gained access to the HardHat program following their final (6 months) interview
2. Intervention group: participants were given immediate access to the HardHat program.

HardHat participants received incentives (about Can $25 [US $19] gift cards) for completing each of the baseline, 3-month, and 6-month interviews. Every other week, participants in the intervention group received email reminders to engage in the HardHat program. Unlike the BroHealth RCT participants, HardHat participants were also incentivized to complete program sessions. For each session completed, participants earned 100 Green Shield Canada points and were given a ballot for a monthly draw to win a US $100 gift card. The Green Shield Canada points could be used to buy additional ballots for the chance to win rewards such as gift cards.

Sample Description

The original qualitative study design entailed a plan to interview a random sample of the BroHealth RCT intervention group

XSL•FO
RenderX

completers to ask them about their experience using the BroHealth program and what aspects of the program they found effective. However, after the initial month of both the BroHealth and HardHat RCTs, we noted that the use of the web-based programs was low among participants in both BroHealth intervention groups and the HardHat intervention group. The sampling plan was shifted to encompass a random sample of participants from these 3 intervention groups; we asked these participants about their perceptions of the effectiveness of their respective web-based programs as well as about what might motivate them to use web-based programs.

A total of 77 participants provided data for this study. Sixty-one participants were from the BroHealth intervention and were organized into 2 groups as follows: BroHealth qualitative group 1 comprised 41 participants who were randomly selected from either of the BroHealth intervention groups (BroHealth intervention group 1=370 participants and BroHealth intervention group 2=374 participants) because they had logged in to BroHealth a maximum of one time during the RCT and were interviewed 2 to 10 months after commencing the program; BroHealth qualitative group 2 comprised 20 participants who were also randomly selected from either of the BroHealth intervention groups, regardless of whether and to what extent they had used the BroHealth program. The BroHealth qualitative group 2 participants were interviewed following the completion of the RCT. Rounding out our sample, a third qualitative group comprising 16 HardHat participants was randomly selected from the HardHat intervention group (HardHat intervention group=103 participants) because they had logged in to HardHat a maximum of one time. HardHat participants were interviewed 1 month into the RCT.

Data Collection

Semistructured telephone interviews were conducted with participants, and interview duration ranged from 2 to 35 min and averaged approximately 9 min. Interviews were audio recorded and transcribed verbatim. All interviews included questions about aspects of the web-based programs that participants found effective, why they did not use the programs, and what did or would motivate program use. The interviews with the BroHealth group 2 participants tended to be lengthier, as these interviews included questions about experience with the RCT overall. The qualitative interview guides are included in the Multimedia Appendix 1.

Sample Size Justification

We sought to capture a full range of both aspects of participants' perceptions of the effectiveness of the web-based programs as well as factors that would motivate the use of such programs. As such, we were guided in our sample sizes by the principle of *data saturation*. Although aligning with Saunders et al [36] (among others) to recognize the multidimensional and contested nature of saturation, in our study, we aimed for data saturation, which we defined as "the degree to which new data repeat what was expressed in previous data." In their discussion and review of guidelines for determining sample sizes needed to achieve *saturation*, Sim et al [37] presented sample size recommendations from an array of qualitative methodologists. Although saturation is contingent on the goals of the study being conducted, Sim et al [37] note that for studies that entail interview data collection, methodologists tend to argue that saturation can be achieved through interviews with as few as 10 and as many as 40 participants. Guided by these numbers, and in light of the particular but novel nature of our study focus, we began with a relatively large sample of 41 participants in our first group (BroHealth group 1). Well before interview 41, interviews became repetitive and did not provide new information. Consistent with our inductive analysis approach, for our second group (BroHealth group 2), we conducted 20 interviews; before the 20th interview, interviews became repetitive both within our second group and between our first and second groups. Accordingly, we reduced our sample size slightly for group 3 (HardHat group) to 16 participants, and we again found data saturation both within and between groups.

Data Analysis

For this inductive content analysis, we began data analysis concurrent with interview data collection. After each interview, the recording was transcribed and reviewed by the research team. We then followed Elo and Kyngas' [33] steps of *open coding* and *creating categories*. These steps began with immersion in the data achieved as a researcher read and reread interview transcripts. This researcher then open coded by rereading transcripts and highlighting pieces of data—words, phrases, or sentences—that reflected meanings in response to our interview questions about aspects of the web-based program that participants considered effective and factors that would motivate participants' use of web-based programs. At this stage, this researcher created as many codes as necessary to reflect the meanings manifested in the interview content. Following this open coding, codes were grouped and collapsed into higher-order categories that captured the overall data set [38,39]. Transcript data and corresponding codes and categories were organized using NVivo 12 software.

Trustworthiness

We enhanced the trustworthiness of our findings by using the triangulation of investigators in our analysis. Specifically, author 2 (MN) conducted a brief analysis by reading select transcripts and providing a summary of key issues evident in the transcripts. All transcripts were then reviewed and coded by author 1, looking for important pieces of data. Authors 1 (HE) and 2 (MN) reviewed both analyses for the similarity of issues identified in the transcripts. A third researcher who was external to this project was brought in to ensure that the broader categories were a fair representation of the codes that were identified. The 3 researchers then discussed and reached consensus about any inconsistencies in codes and categories.

This project was approved by the Research Ethics Board of the Royal Mental Health Centre in Ottawa, Canada.

Results

Sample Demographics

Of the 77 participants, 70 were males and 7 were females. Response rates in the form of the number of prospective participants contacted versus the number who agreed to participate in the interviews are depicted in Table 1.

XSL•FO
RenderX

The average age of the participants was 40.6 years (SD 1.34) and ranged from 20 to 65 years. The average risk of depression in the next 4 years was 23.1% and ranged from 7% to 87%. We generated 4 categories pertaining to the perceived effectiveness of the web-based mental health programs: awareness of health, functionality and medium, content, and coaches. We generated 5 categories pertaining to what motivates the use of web-based mental health programs: reminders and incentives, promotion of the programs, appropriate medium and functionality, perceived need, and questioning how or if motivating use is possible. Table 2 shows a summary of all categories generated through the content analysis.

Table 1. Response rates of each group of qualitative interviews.

Interview groups	Population (n)	Contacted, n (%)	Interviews, n (%)
BroHealth low usage	744	81 (10.9%)	41 (5.5%)
BroHealth after randomized controlled trial	744	101 (13.6%)	20 (2.7%)
HardHat	103	61 (59.2%)	16 (15.5%)

Table 2. Summary of the categories that were found through content analysis divided into their respective topics.

Topics	Categories
Effectiveness	• Web-based programs increase awareness of mental health Perceived program functionality and the internet as a medium • Contrasting perceptions of content effectiveness • Coaches increase perceived effectiveness of programs
Motivators	• Providing reminders and incentives increases motivation • Promotion of web-based health programs increases motivation • Providing appropriate medium and functionality increases motivation • Perceived need for use increases motivation • Motivating the use of web-based programs may be impossible

Perceived Effectiveness of Web-Based Mental Health Programs

Web-Based Programs Increase Awareness of Mental Health

The most common benefit of web-based mental health programs identified by participants is that such programs improve mental health awareness. Some participants spoke positively about increased self-awareness resulting from the use of the programs. For instance, "...it forced me through some of the issues that I am having" and "I think it helped just be aware of some of the problems writing them down. You know, I just didn't really focus on them that much."

Participants appreciated being aware of a resource that they could count on to support their needs currently and in the future. One participant explained:

> *I'd like to have it as a resource I can get to because there are times where like work, life and what not is more stressful than normal...and I want to make sure that I have help when I need it.*

Participants believed that web-based programs are especially beneficial for individuals who are unwilling or financially unable to seek professional help.

Moreover, participants believe that web-based mental health programs can increase society's awareness of mental health. Participants viewed the web-based programs as a "step in the right direction" in terms of reducing stigma and increasing access to help, especially for men:

> *...it also makes it a nice tool for people who don't want to talk about mental health among men. It's still a big stigma right now. Not a lot of people that want to admit to it, but it's nice that it's there.*

Perceived Program Functionality and the Internet as a Medium

Functionality and medium were important considerations with respect to the perceived effectiveness of web-based health programs. Participants expressed conflicting views about using the internet. On the one hand, participants endorsed the idea of having 24/7 access and an increased level of privacy. For instance, one participant mentioned:

> *I would say that that makes it effective to me because it's available to me at 1 o' clock in the morning if I wake up and I can't sleep...or feeling anxious about something, that's an advantage to uh, to be able to look at something and at that time.*

Another participant said:

> *I assume a lot of people who are in my situation are in a crazy schedule so like you know the ability to do something on your own time and not always have appointments booked for you is good.*

On the other hand, participants mentioned having procrastinated because there was nobody to ensure their participation. Furthermore, many participants had limitations to their computer use for reasons such as long hours on the computer for work, not being comfortable with computer use, and not having internet access. For instance:

XSL•FO
RenderX

> *I work on computers all day in my job so to then after sit on a computer um you know, when I walk away from my work I cannot just spend a lot of time on computers.*

Although there was little reference to specific aspects of web-based program functionality, most participants were happy with the overall functionality and ease of use of both programs. However, some participants in both programs expressed finding it difficult to navigate the websites:

> *It was okay, for the first little bit it might have been a little bit confusing and not really sure where to go, what I'm looking for.*

Some participants also noted technical issues such as broken links that may inhibit the effectiveness of the program.

Contrasting Perceptions of Content Effectiveness

Participants had varying opinions regarding the effectiveness of the content of the web-based mental health programs. Some found the content to be appropriate, detailed and engaging, and effective at decreasing stress. One participant remarked:

> *What was helpful is that, I do need to get back to listing what my challenges and goals are.*

Conversely, some participants did not find the program content useful for the following reasons: the content was too similar to other resources, the content was not relevant to the user, the program had a poor flow, and the program lacked appropriate information on mental health. One participant stated:

> *after a half a dozen times I found that the content was redundant with what I had already gotten from my healthcare provider.*

Some participants believed that the amount and type of content were comprehensive and covered a wide area related to mental health, whereas other participants felt there was too much information and that "...there was lot of text so sometimes it became a bit heavy," discouraging them from using the program.

Participants also recommended having personalized profiles to allow them to more quickly and easily narrow in on appropriate resources. Participants who believed the content to be appropriate and helpful cast their feedback in broad terms such as "It was engaging. It was appropriate." The few people who gave specific positive feedback appreciated the elements that they could adopt and implement in their daily lives, thus increasing their personal benefits of program use. Participants also suggested that periodically updating the content would have encouraged ongoing use.

Coaches Increase the Perceived Effectiveness of Programs

Although 45 participants had access to a coach (intervention group 2 of BroHealth and HardHat intervention group participants), very few used their coach. Of the 29 participants in BroHealth intervention group 2 who had access to a coach, only 2 utilized the coach and they believed that it was a great feature of the program:

> *I think I called at least 6 to 8 times and each time I was really happy to talk to him...to bring me back and to give me good advice and to take me back to planet earth a little.*

None of the 16 HardHat intervention group participants used their coach. Most were not aware that there was a coach included in the program or were unsure of the role of the coach. A few participants expressed not feeling comfortable talking with a coach because they were unsure about what to say. Participants who were either not given access to a coach (intervention group 1 in the BroHealth RCT) or who did not use their coach speculated that having a counselor to talk to would be an added benefit of the program. One participant compared the availability of BroHealth coaches with the availability of counselors in another program:

> *...they utilize an interesting way of connecting live counsellors with people who are actively looking to chat right now. And that's the only thing that wasn't there.*

HardHat participants also believed that coaches would benefit them:

> *I think that getting that personal connection with the coach to be able to have that personal connection and possibly be able to relate to them.*

Participants from both programs recommended having counselors that they could speak to via text, phone, or in person by appointment or on a 24/7 basis. Participants believed this would improve the web-based programs in 2 ways: (1) by keeping participants accountable by scheduling appointments and (2) by giving them someone to talk to if they need extra help with their mental health.

What Motivates the Use of Web-Based Mental Health Programs

Providing Reminders and Incentives Increases Motivation

Most participants believed that the reminders were beneficial in increasing their program use:

> *…when I got an e-mail from you before that sort of was a trigger to take a look.*

At the same time, many participants did not recall receiving these reminders, thus limiting the perceived effectiveness of reminders. Many participants believed that text message reminders, rather than email, would be a more effective way to motivate program use:

> *I don't always check my emails, it's not something that I necessarily sit down and do on a daily basis. Everybody always has their phone with them.*

Participants also acknowledged that the follow-up interviews conducted as part of the RCTs served as reminders to use the programs:

> *A lot of it was after interviews that I've done previously. It was a good reminder that the website was there.*

Participants in the HardHat RCT received incentives to engage in the program, whereas participants in the BroHealth RCT did

XSL•FO
RenderX

not. Participants who received incentives and participants who did not receive incentives indicated that incentives were a good motivator for program use. For instance, one participant commented:

> *Something like points or rewards for progress as you work through it kind of thing might be a motivator to go back to it.*

Promotion of Web-Based Health Programs Increases Motivation

Overall promotion and marketing of web-based mental health programs are major motivators to use the program. Participants believed that better promotion, and thus awareness and knowledge of the program, is the first step in motivating program use. For instance, one participant said:

> *It really is just trying to motivate the person to click it once, I suppose, and go on there.*

Moreover, participants believed that having a good understanding of the program and its benefits before starting the program would motivate use. Various ways to promote use to people not already using the web-based programs were described by participants, including emails, text messages, word of mouth, and social media. One participant suggested using employers and workplaces to distribute information about the program, ensuring that a support system in the workplace is in place, and even having the program integrated within the workplace. Another participant said:

> *...it would be nice if things were within the workplace. In the workplace there's more structure. A lot of my stress derived from being self-employed and not having uh the workplace environment there to support me.*

Providing the Appropriate Medium and Functionality Increases Motivation

The type of medium and functionality of the web-based programs were significant motivators. Many participants mentioned the importance of personal preference, noting that younger generations tend to be drawn to web-based media. Participants conveyed the importance of the medium fitting into their everyday life to motivate them to use the program. For instance, having notifications on their social media accounts and/or being able to access the program as an app on their mobile device were motivators. One participant explained:

> *whatever kind of sites that a person connects through...so you know if I get, if I get Facebook updates like that.*

Many participants mentioned that being able to talk to a person (via text messaging, phone, or in person) in some capacity is beneficial for web-based mental health programs and would motivate use by increasing accountability. Ease of use (ie, “quick, easy, and convenient”) was also a motivator to use the programs, as these characteristics allow the programs to fit into the busy lives of the participants.

Perceived Need for Use Increases Motivation

Participants perceiving that they need the program and perceiving the program as beneficial are major motivations for program use. Many nonusers anticipated that they would use web-based mental health programs in the future if they perceived a need for it. These nonusers did not believe that they were currently going through a hard time or having any symptoms; thus, they did not believe they had a personal need for the program.

Participants who did perceive themselves as having a need for the program were more motivated to continue using the program when program content was relevant to their lives. For instance, one participant said:

> *That one time I did go into it, I liked what I saw so I'm going to be going back.*

Having content relevant to participants’ experiences makes participants more likely to achieve the benefits of the program, in turn, improving motivation to use the program on a continuous basis. Updating the content may also lead to continued use.

Motivating the Use of Web-Based Programs May Be Impossible

Although motivating the use of web-based mental health programs is an important endeavor to ensure the use of mental health programs, many participants were simply unsure of how to do so or felt that motivating use was not possible. For instance, one participant stated:

> *...that's a really good question and if it was an easy answer, I wish it was an easy answer.*

Others felt that the barriers to using such programs, such as lack of prioritization and stigma, are unsurmountable. For example, one participant explained:

> *when it comes to stacking priorities, how would you get around that, I really don't know.*

Another participant commented:

> *...in my opinion, probably can't...I just come from a different school of thought I guess. That's just not a man thing to do.*

Discussion

Principal Findings

Web-based mental health programs are valuable in preventing and treating depression [20-22]. To increase the use of prevention programs, it is important to understand what motivates the use and the perceived effectiveness of various components of the programs. Functionality and medium, awareness of mental health, and appropriate content were identified in this study as important factors for judging the perceived effectiveness of web-based mental health programs. Motivators to use such programs include providing reminders and incentives; promoting the programs (including endorsement by employers); providing appropriate medium, functionality, and content; and having a perceived need. At the same time,

XSL•FO
RenderX

some participants were unsure of how they could be motivated to use the programs or felt that there may be no way to do so.

Past literature describes many benefits of using web-based mental health programs, including convenience, accessibility, and cost-effectiveness [24]. Furthermore, having both increased awareness of one's mental health and good mental health literacy (ie, having knowledge about symptoms, treatments, and resources) is an important element of depression prevention [40,41]. Increased awareness reduces stigma and encourages individuals to access mental health services before the onset of depression. Our study echoes these findings, as mental health awareness was perceived as the largest benefit of the programs that we investigated.

Implications for Future Programs

Despite the benefits described earlier, adherence to web-based programs for mental health is low. Thus, the question is, how can we motivate people to use web-based health programs for mental health? Furthermore, how can we motivate people to use web-based programs before developing depression?

Promoting Web-Based Mental Health Programs

The first and, arguably, the hardest step is to convince people to use such programs for the first time. To convince people to use these programs, people need to (1) have the knowledge that such programs exist and (2) believe that they have a need for the services provided by the web-based mental health programs. Useful marketing of the programs was a very strong motivator uncovered in this study. Primary care providers are the number one source of mental health information [42], but many people are reluctant to see their physician about mental health problems, especially for prevention. As a result, a large proportion of people are not receiving the resources they need. The results of this study indicate that men who are reluctant to see mental health professionals and use mental health resources may be more open to discussing resources with people close to them. Although men are often afraid of stigmatization, recent studies have reported that many men have a small social circle with whom they feel they can share their concerns [43]. Men are also very willing to support the mental health of other people in their lives [43]. This close social circle/familial support adds an avenue through which to promote the initiation of web-based mental health programs.

Mental health programs within the workplace have been reported in the literature as decreasing stigmatization of poor mental health [44]; workplace mental health programs may contribute to an increase in web-based program use. This study uncovered a complementary finding that participants may be more likely to use web-based mental health programs if these programs are somehow incorporated into the workplace setting. In addition, having a program in the workplace was found in this study to increase feelings of mental health support. Thus, promoting web-based programs in the workplace may help to increase awareness about the program and more effectively encourage initial and ongoing participation.

Once a person has decided to initiate the use of a mental health program, they need to complete it to reap the full benefits [45]. Many methods have been noted in the literature to increase adherence to web-based programs.

Improving the Continued Use of Web-Based Mental Health Programs

Our study found that appropriate content was an important factor for increasing the perceived effectiveness and, consequently, the increased use of web-based mental health programs. However, as content needs often vary significantly from person to person, the participants in this study postulated that having individually tailored content may motivate the continued use of a web-based program. Moreover, meta-analyses and systematic reviews have shown that having tailored content can decrease symptoms of depression in study populations [46,47]. Having future programs tailored to a specific group or having the ability to tailor to each specific user would be beneficial to improve adherence and effectiveness of the program. In addition, the results of this study suggest that tailoring the web-based mental health program to people with lower mental health literacy would make the content less redundant. Lustria et al [46] found that computer assessments could be used to narrow down the content to that users' specific needs. Future research should focus on how to effectively tailor content to specific groups.

In addition, reminders not only increase adherence to [48] and perceived effectiveness of [49] web-based programs but also help foster a positive attitude toward them [50]. In this study, the reminder emails motivated program use. However, email reminders were often discarded immediately by participants without being read. Many participants speculated that reminders sent via text messages may be a better format. Having a direct link to the program to make it faster and easier for participants was endorsed by this study. Although studies have found results consistent with this finding [51], more research is needed to find the optimal medium for reminders to motivate program use.

The use of the internet as a medium was a polarizing topic, as web-based programs are not preferred by everyone. Although some participants believe using the internet as a medium is ideal because of privacy and convenience, others preferred alternate media as they had limited computer use. This discrepancy is consistent with other studies [52]. For instance, a study by Smail-Crevier et al [53] found that younger generations are more likely to engage with web-based programs compared with older generations. Another study found that high computer/internet use is negatively associated with the use of web-based mental health programs [54]. Moreover, having a live person to speak to is associated with increased use and completion of the program [54]. Agreeing with the literature, the results from this study demonstrate that having access to someone to speak to may make the programs more effective and meaningful to users and thus could increase motivation and adherence to future web-based programs. The results of this study indicate that although coaches or guides in programs could potentially be effective, unless participants are aware that they have access to a coach and understand the role of the coach, they are unlikely to use it.

XSL•FO
RenderX

Impact of Society on Web-Based Mental Health Programs

Although the methods to motivate use described above have been found to be effective, there are many individuals who are unsure of how to motivate their own use or who simply do not believe that it is possible to motivate people in general because depression and help-seeking remain highly stigmatized [55-58]. Specifically, at-risk individuals tend to have lower mental health literacy and are less likely to seek treatment [59]. The Health Belief Model suggests that a person will only take action to prevent disease if (1) they believe they are vulnerable to the disease, (2) the disease is severe enough to negatively affect them, and (3) there is a benefit to the prevention that outweighs the cost [60]. These findings have been echoed in previous research on depression [61-63] and in our study, as participants were more likely to use the programs if they believe that they needed it.

Limitations

This study is limited in several ways. First, given the specific and detailed nature of this study, we make no claims that our findings can be extrapolated to larger populations or to other web-based mental health programs. Second, there is a chance that a different research team with different backgrounds would generate different categories from our data. To offset this limitation, multiple researchers collaborated and achieved a consensus on the analysis. We note further that some participants were interviewed a few months after they had used the program and thus may have had difficulty with recall. Our study did not examine how demographics, such as age, income level, and occupation, could influence participants' perceptions about web-based mental health programs. Finally, very few participants engaged with their coaches. As a result, the views that they expressed about coaches were often not based on personal experience and should therefore be interpreted with caution.

Conclusions

In conclusion, awareness of one's own mental health is a benefit of using web-based mental health programs. The content and functionality of the programs are important factors when assessing perceived effectiveness. However, perceived effectiveness varies from person to person, and what may be effective for one person may not be effective for another. Access via chat to a real person and being able to customize content to fit the users' personal needs are improvements that could increase the perceived effectiveness and use of similar programs. Motivating people to use such programs is difficult; even current program users are unsure of the best ways to motivate use. Motivators include providing reminders or incentives, promoting the programs, providing appropriate medium and functionality, and perceived need, which, if adequately incorporated into web-based program design and implementation, may lead to increased use.

Acknowledgments

This study was supported by grants from the Movember Foundation and Green Shield Canada. The Movember Foundation and Green Shield Canada play no role in the study design, analysis, and interpretation.

Authors' Contributions

All authors contributed to study design, interpretation, review, and final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative interview guide.
[PDF File (Adobe PDF File), 178 KB - jmir_v22i7e16961_app1.pdf]

References

1. Steensma C, Loukine L, Orpana H, McRae L, Vachon J, Mo F, et al. Describing the population health burden of depression: health-adjusted life expectancy by depression status in Canada. Health Promot Chronic Dis Prev Can 2016 Oct;36(10):205-213 [FREE Full text] [doi: 10.24095/hpcdp.36.10.01] [Medline: 27768557]
2. Hawton K, Comabella CC, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord 2013 May;147(1-3):17-28. [doi: 10.1016/j.jad.2013.01.004] [Medline: 23411024]
3. Dhar AK, Barton DA. Depression and the link with cardiovascular disease. Front Psychiatry 2016;7:33 [FREE Full text] [doi: 10.3389/fpsyt.2016.00033] [Medline: 27047396]
4. Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care 2008 Dec;31(12):2383-2390 [FREE Full text] [doi: 10.2337/dc08-0985] [Medline: 19033418]
5. Chang M, Hsu J, Huang K, Su T, Bai Y, Li C, et al. Bidirectional association between depression and fibromyalgia syndrome: a nationwide longitudinal study. J Pain 2015 Sep;16(9):895-902. [doi: 10.1016/j.jpain.2015.06.004] [Medline: 26117813]
6. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. J Am Med Assoc 2003 Jun 18;289(23):3135-3144. [doi: 10.1001/jama.289.23.3135] [Medline: 12813119]

XSL•FO
RenderX

7. Hassard J, Teoh KR, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
8. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5):e1001454 [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
9. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
10. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
11. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017 Feb;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
12. Simpson KR, Meadows GN, Frances AJ, Patten SB. Is mental health in the Canadian population changing over time? Can J Psychiatry 2012 May;57(5):324-331. [doi: 10.1177/070674371205700508] [Medline: 22546065]
13. Chisholm D, Sanderson K, Ayuso-Mateos JL, Saxena S. Reducing the global burden of depression: population-level analysis of intervention cost-effectiveness in 14 world regions. Br J Psychiatry 2004 May;184:393-403. [doi: 10.1192/bjp.184.5.393] [Medline: 15123502]
14. Alberta's Top 11 Priorities for Depression Research. Alberta Innovates. URL: https://albertainnovates.ca/wp-content/uploads/2018/02/Alberta-Depression-Research-Priority.pdf [accessed 2019-07-18]
15. Breault LJ, Rittenbach K, Hartle K, Babins-Wagner R, de Beaudrap C, Jasaui Y, et al. The top research questions asked by people with lived depression experience in Alberta: a survey. CMAJ Open 2018;6(3):E398-E405 [FREE Full text] [doi: 10.9778/cmajo.20180034] [Medline: 30266777]
16. Smit F, Willemse G, Koopmanschap M, Onrust S, Cuijpers P, Beekman A. Cost-effectiveness of preventing depression in primary care patients: randomised trial. Br J Psychiatry 2006 Apr;188:330-336. [doi: 10.1192/bjp.188.4.330] [Medline: 16582059]
17. Callander EJ, Lindsay DB, Scuffham PA. Employer benefits from an early intervention program for depression: a cost-benefit analysis. J Occup Environ Med 2017 Mar;59(3):246-249. [doi: 10.1097/JOM.0000000000000939] [Medline: 28267095]
18. Mohr DC, Ho J, Duffecy J, Baron KG, Lehman KA, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010 Apr;66(4):394-409 [FREE Full text] [doi: 10.1002/jclp.20659] [Medline: 20127795]
19. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: barriers to the use of apps for depression. Digit Health 2017;3:2055207617713827 [FREE Full text] [doi: 10.1177/2055207617713827] [Medline: 29942605]
20. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
21. Cowpertwait L, Clarke D. Effectiveness of web-based psychological interventions for depression: a meta-analysis. Int J Ment Health Addiction 2013 Jan 18;11(2):247-268. [doi: 10.1007/s11469-012-9416-z]
22. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Aug 17;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
23. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: a randomized clinical trial. J Am Med Assoc 2016 May 3;315(17):1854-1863. [doi: 10.1001/jama.2016.4326] [Medline: 27139058]
24. Compen FR, Bisseling EM, Schellekens MP, Jansen ET, van der Lee ML, Speckens AE. Mindfulness-based cognitive therapy for cancer patients delivered via internet: qualitative study of patient and therapist barriers and facilitators. J Med Internet Res 2017 Dec 18;19(12):e407 [FREE Full text] [doi: 10.2196/jmir.7783] [Medline: 29254912]
25. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
26. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
27. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Bouck Z, et al. Extending access to a web-based mental health intervention: who wants more, what happens to use over time, and is it helpful? Results of a concealed, randomized controlled extension study. BMC Psychiatry 2019 Jan 24;19(1):39 [FREE Full text] [doi: 10.1186/s12888-019-2030-x] [Medline: 30678676]

XSL•FO
RenderX

28. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
29. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
30. Boydell KM, Hodgins M, Pignatiello A, Teshima J, Edwards H, Willis D. Using technology to deliver mental health services to children and youth: a scoping review. J Can Acad Child Adolesc Psychiatry 2014 May;23(2):87-99 [FREE Full text] [Medline: 24872824]
31. Crowe M, Inder M, Porter R. Conducting qualitative research in mental health: thematic and content analyses. Aust N Z J Psychiatry 2015 Jul;49(7):616-623. [doi: 10.1177/0004867415582053] [Medline: 25900973]
32. Peters S. Qualitative research methods in mental health. Evid Based Ment Health 2010 May;13(2):35-40. [doi: 10.1136/ebmh.13.2.35] [Medline: 21856603]
33. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
34. Wang JL, Patten SB, Lam RW, Attridge M, Ho K, Schmitz N, et al. The effects of an e-mental health program and job coaching on the risk of major depression and productivity in Canadian male workers: protocol for a randomized controlled trial. JMIR Res Protoc 2016 Nov 15;5(4):e218 [FREE Full text] [doi: 10.2196/resprot.6350] [Medline: 27847352]
35. Wang JL, Manuel D, Williams J, Schmitz N, Gilmour H, Patten S, et al. Development and validation of prediction algorithms for major depressive episode in the general population. J Affect Disord 2013 Oct;151(1):39-45 [FREE Full text] [doi: 10.1016/j.jad.2013.05.045] [Medline: 23790813]
36. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
37. Sim J, Saunders B, Waterfield J, Kingstone T. Can sample size in qualitative research be determined a priori? Int J Soc 2018 Mar 27;21(5):619-634. [doi: 10.1080/13645579.2018.1454643]
38. Burnard P. A method of analysing interview transcripts in qualitative research. Nurse Educ Today 1991 Dec;11(6):461-466. [doi: 10.1016/0260-6917(91)90009-y] [Medline: 1775125]
39. McCain GC. Content analysis: a method for studying clinical nursing problems. Appl Nurs Res 1988 Nov;1(3):146-147. [doi: 10.1016/s0897-1897(88)80029-6] [Medline: 3239999]
40. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
41. Finkelstein J, Lapshin O. Reducing depression stigma using a web-based program. Int J Med Inform 2007 Oct;76(10):726-734. [doi: 10.1016/j.ijmedinf.2006.07.004] [Medline: 16996299]
42. Whitton AE, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
43. Wellstead P, Norriss H. Information-seeking to support wellbeing: a pilot study of New Zealand men. Int J Wellbeing 2014 Dec 30;4(2):32-44. [doi: 10.5502/ijw.v4i2.3]
44. Nigatu YT, Huang J, Rao S, Gillis K, Merali Z, Wang J. Indicated prevention interventions in the workplace for depressive symptoms: a systematic review and meta-analysis. Am J Prev Med 2019 Jan;56(1):e23-e33. [doi: 10.1016/j.amepre.2018.08.027] [Medline: 30573152]
45. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
46. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
47. Wahle F, Bollhalder L, Kowatsch T, Fleisch E. Toward the design of evidence-based mental health information systems for people with depression: a systematic literature review and meta-analysis. J Med Internet Res 2017 May 31;19(5):e191 [FREE Full text] [doi: 10.2196/jmir.7381] [Medline: 28566267]
48. Jorm AF, Christensen H, Griffiths KM. The impact of beyondblue: the national depression initiative on the Australian public's recognition of depression and beliefs about treatments. Aust N Z J Psychiatry 2005 Apr;39(4):248-254. [doi: 10.1080/j.1440-1614.2005.01561.x] [Medline: 15777361]
49. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
50. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

XSL•FO
RenderX

51. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
52. Jackson DN, Peterson EB, Blake KD, Coa K, Chou WS. Americans' trust in health information sources: trends and sociodemographic predictors. Am J Health Promot 2019 Nov;33(8):1187-1193. [doi: 10.1177/0890117119861280] [Medline: 31337226]
53. Hermes ED, Merrel J, Clayton A, Morris C, Rowe M. Computer-based self-help therapy: a qualitative analysis of attrition. Health Informatics J 2019 Mar;25(1):41-50. [doi: 10.1177/1460458216683536] [Medline: 28027676]
54. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
55. Kelders SM, Bohlmeijer ET, van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
56. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
57. Cook TM, Wang J. Descriptive epidemiology of stigma against depression in a general population sample in Alberta. BMC Psychiatry 2010 Apr 19;10:29 [FREE Full text] [doi: 10.1186/1471-244X-10-29] [Medline: 20398429]
58. Lannin DG, Vogel DL, Brenner RE, Abraham WT, Heath PJ. Does self-stigma reduce the probability of seeking mental health information? J Couns Psychol 2016 Apr;63(3):351-358. [doi: 10.1037/cou0000108] [Medline: 26323042]
59. Oliffe JL, Ogrodniczuk JS, Gordon SJ, Creighton G, Kelly MT, Black N, et al. Stigma in male depression and suicide: a Canadian sex comparison study. Community Ment Health J 2016 Apr;52(3):302-310 [FREE Full text] [doi: 10.1007/s10597-015-9986-x] [Medline: 26733336]
60. Stanton R, Rosenbaum S, Rebar A. Associations between ability to recognise a mental health disorder and lived experience of mental illness in an Australian sample. Psychiatry Res 2019 Feb;272:206-208. [doi: 10.1016/j.psychres.2018.12.098] [Medline: 30590273]
61. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974 Dec 1;2(4):354-386. [doi: 10.1177/109019817400200405]
62. Boerema AM, Kleiboer A, Beekman AT, van Zoonen K, Dijkshoorn H, Cuijpers P. Determinants of help-seeking behavior in depression: a cross-sectional study. BMC Psychiatry 2016 Mar 23;16:78 [FREE Full text] [doi: 10.1186/s12888-016-0790-0] [Medline: 27009062]
63. Magaard JL, Seeralan T, Schulz H, Brütt AL. Factors associated with help-seeking behaviour among individuals with major depression: a systematic review. PLoS One 2017;12(5):e0176730 [FREE Full text] [doi: 10.1371/journal.pone.0176730] [Medline: 28493904]

Abbreviations

CBT: cognitive behavioral theory
MDE: major depressive episode
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by K Boydell, L Boß; comments to author 03.02.20; revised version received 30.03.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Eccles H, Nannarone M, Lashewicz B, Attridge M, Marchand A, Aiken A, Ho K, Wang J
Perceived Effectiveness and Motivations for the Use of Web-Based Mental Health Programs: Qualitative Study
J Med Internet Res 2020;22(7):e16961
URL: http://www.jmir.org/2020/7/e16961/
doi:10.2196/16961
PMID:32735216

©Heidi Eccles, Molly Nannarone, Bonnie Lashewicz, Mark Attridge, Alain Marchand, Alice Aiken, Kendall Ho, Jianli Wang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal

XSL•FO
RenderX

of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Virtual Reality–Supported Intervention for Pulmonary Rehabilitation of Patients With Chronic Obstructive Pulmonary Disease: Mixed Methods Study

Timothy Jung[1*], MSc, PhD; Natasha Moorhouse[1], MSc, PhD; Xin Shi[1*], MSc, PhD; Muhammad Farhan Amin[2*], MD

[1]Manchester Metropolitan University, Business School, Manchester, United Kingdom

[2]Burnett Edgar Medical Centre, Central Drive, Walney, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Xin Shi, MSc, PhD
Manchester Metropolitan University
Business School
Oxford Road
Manchester, M15 6BH
United Kingdom
Phone: 44 7527423967
Email: x.shi@mmu.ac.uk

Abstract

Background: The uptake of traditional pulmonary rehabilitation classes by patients with chronic obstructive pulmonary disease (COPD) is poor because of personal factors that prevent accessibility to the venue. Therefore, there is a need for innovative methods of pulmonary rehabilitation, and virtual reality (VR) could be a promising technology for patients with COPD to access services remotely.

Objective: This study aimed to investigate whether VR improves compliance with pulmonary rehabilitation among patients with COPD, a particularly vulnerable patient group (Medical Research Council [MRC] 4 or 5), and whether VR provides a credible alternative to traditional pulmonary rehabilitation programs.

Methods: This was an 8-week patient trial using an innovative VR pulmonary rehabilitation program. A purposive sample of 10 patients with COPD graded MRC 4 or 5 and registered at a selected health care center and a hospital in Cumbria, United Kingdom, were included. Qualitative (focus groups and interviews) data were collected, and to further support the qualitative findings, quantitative data (self-report patient surveys) were gathered before and after the 8-week trial. The 5 self-reported surveys included the Patient Activation Measure, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, Short Physical Performance Battery, and the Edmonton Frail Scale.

Results: In a thematic analysis of the qualitative data, 11 themes emerged specific to delivering pulmonary rehabilitation using VR. The quantitative data further support the qualitative findings by revealing significant improvements in all physical measures.

Conclusions: Overall, this study demonstrates how remotely supervised VR-based pulmonary rehabilitation could help to overcome current issues and limitations associated with providing this service to patients with COPD at scale.

(J Med Internet Res 2020;22(7):e14178) doi:10.2196/14178

KEYWORDS

virtual reality; COPD; rehabilitation

Introduction

Background

Pulmonary rehabilitation is a nonpharmacological intervention designed for patients with chronic obstructive pulmonary disease (COPD), involving supervised exercise training, disease education, and behavioral interventions [1]. It is now one of the most effective treatments for significantly improving symptoms of dyspnea (ie, breathlessness) [2-4], exercise capacity, and improved quality of life [2,4-6] and reducing anxiety and depression in patients with COPD [1,7]. However, the uptake of pulmonary rehabilitation is poor, and completion rates are low [1,7,8]. Patients who qualify for pulmonary rehabilitation

XSL•FO
RenderX

are those graded 3 or above on the Medical Research Council (MRC) breathlessness scale, which is a validated measure of disease severity used in many scientific studies across the world, given that this particular group is at high risk of exacerbation and hospital admission [9-11]. However, patients have difficulty attending classes because of lack of transport and geographic distance to a program, fatigue, lack of motivation, inconvenience, disruption caused to daily activities, and the quality of conversation that health care professionals have with patients about pulmonary rehabilitation [12,13]. Depression, burden of illness, low awareness of rehabilitation, and knowledge of and disbelief in beneficial rehabilitation outcomes are also considered barriers to patient uptake [14-16]. Therefore, researchers and practitioners are searching for innovative methods to deliver more engaging rehabilitation for patients with a variety of long-term conditions that affect physical activity.

More recently, there has been an increase in research focusing on the effects of virtual reality (VR) for patients with chronic diseases that inhibit physical activity. For example, studies focusing on individuals with Parkinson disease have investigated the effects of gait training with VR [17], whereas others have compared home-based VR balance training with conventional home-based balance training [18]. The physical and psychological benefits of VR for patients who have had a stroke have also been explored [19,20]. However, despite the extant research in this area, studies investigating VR for physical training rehabilitation programs have demonstrated varying results, which means that researchers and practitioners are unsure about its true impact [21,22]. The need for innovative methods of pulmonary rehabilitation is evident, and VR could be a promising technology for providing a convenient and remotely accessible pulmonary rehabilitation program to patients with COPD. COPD places a significant financial burden on individuals and societies worldwide and is a global public health burden that needs to be addressed to reduce the substantial cost associated with the disease and improve patients' quality of life [23,24]. Remotely supervised VR could complement conventional therapy [25], which has demonstrated poor uptake because of patients' personal barriers [8]. However, to the best of our knowledge, limited research has investigated the benefits and outcomes of VR for pulmonary rehabilitation, which leaves an important area of research requiring further investigation. To address this gap, this study aimed to investigate whether VR improves the compliance of patients with COPD for pulmonary rehabilitation, with a particular focus on the vulnerable patient group (MRC 4 or 5), and whether this technology provides a credible alternative to traditional pulmonary rehabilitation programs. The results of this study provide significant theoretical contributions to the field of health care research and provide important implications for health practitioners.

What Is Already Known on This Topic?

- COPD is a global public health burden that needs to be addressed to reduce the substantial cost associated with the disease and improve the patient's quality of life.
- Remotely supervised VR could complement conventional therapy, which has demonstrated poor uptake due to patients' personal barriers.
- Limited research has investigated the benefits and outcomes of VR for pulmonary rehabilitation, which leaves an important area of research requiring further investigation.

What This Study Adds?

- VR provides the opportunity for an interactive and visually stimulating approach for use in clinical treatment and could provide pulmonary rehabilitation to patients with COPD who cannot easily access traditional rehabilitation methods.
- This study provides a significant contribution to health care research investigating the benefits of implementing digital technologies to improve rehabilitation, with particular contribution to pulmonary rehabilitation research.
- Important practical implications for the health care industry and medical practitioners on the benefits of VR for pulmonary rehabilitation are offered.

Literature Review

Technology-Facilitated Methods for Pulmonary Rehabilitation

COPD is characterized by progressive airflow obstruction and cannot be fully reversed [26]. Patients with COPD find it difficult to engage in physical activity and often have episodes of exacerbation, including breathlessness and coughing, which provoke anxiety and episodes of panic leading to an overall decline in everyday function and overall quality of life [26-29]. The psychological consequences of COPD include symptoms of depression; general anxiety [30-35]; and psychological repercussions, including fear, panic, loss of confidence, and social isolation [36,37], which have important implications for the adoption and maintenance of healthy behaviors [38]. In turn, this restricts the patients' ability to perform daily activities, and owing to the fear of dyspnea, they often avoid physical activity and social events, which may contribute to an increased likelihood of depression in elderly patients with COPD [30,36,37]. In addition, an increasingly sedentary lifestyle leads to muscle weakness, further reductions in physical activity, reduced exercise capacity, and even more severe symptoms [39,40]. The MRC breathlessness scale measures perceived respiratory disability and is widely used to describe patient cohorts and stratify them for interventions such as pulmonary rehabilitation in COPD [29,40]. More specifically, patients graded MRC ≥3 are considered eligible for pulmonary rehabilitation services [9].

Rehabilitation facilitated by information and communication technologies is known as telerehabilitation [41,42]. Over the past few years, there has been an increase in telerehabilitation apps owing to the development of new technologies [43]. Compared with traditional inpatient or person-to-person rehabilitation, telerehabilitation allows for remote communication and is, therefore, more cost-effective for both health care providers and patients [43]. Some examples include in-home videoconferencing [44,45], mobile phone–based exercise programs [46], web-based self-monitoring [47-49], and more recently, socially assistive robots [50] and VR. Studies have suggested that telerehabilitation is suited to older people as it can improve the quality of life, allows for more independent living, is more convenient in terms of not having to travel for

XSL•FO
RenderX

appointments, and enables cost-effective services [51-54]. However, despite growing evidence-based research demonstrating positive clinical outcomes and increasing telehealth utilization, patients are faced with equipment setup-related difficulties, limited scope of exercise, and connectivity [55].

Although the completion of pulmonary rehabilitation is associated with improved outcomes, including reduced subsequent hospital admission rates and better survival [10], the uptake and completion rates remain poor because of personal barriers associated with time, costs, fatigue, motivation, and inconvenience [8]. COPD is the third leading cause of death worldwide. In 2016, the Global Burden of Disease study reported a prevalence of 251 million global cases of COPD that year [56,57]. Research has indicated that COPD affects at least 320 million people and has a global economic cost of US $2.1 trillion [58,59]. Significant health care costs are associated with the treatment of exacerbations, including hospital visits and medication costs for maintenance therapy and outpatient treatment [24,29]. Over the past few years, general practice has been encouraged to explore new ways to improve COPD management and reduce hospital admissions [60]. A deeper understanding of the barriers associated with the acceptance and uptake of pulmonary rehabilitation by patients with COPD could reveal new methods to improve its uptake [10,60], which would reduce the considerable health care burden of COPD on the patients' quality of life and the health economy. In addition, self-management interventions could lead to a reduction in COPD-related health costs by reducing hospital admissions [61-63]. Therefore, researchers and practitioners have sought innovative methods to create more exciting rehabilitation programs for patients such as VR, which is said to increase their excitement and interest, and excitement often leads to increased motivation to complete the rehabilitation program [21,64-66].

The Use of Virtual Reality for Physical Training and Rehabilitation

With recent technological advancements in VR, innovative approaches to improve traditional physical therapy and rehabilitation programs can be explored [21,67,68]. VR can be defined as a computer-assisted technology that provides the user with an immersive, interactive, and multisensory experience in a 3D virtual environment. The most cutting-edge hardware to facilitate VR is the head mounted display. Patients can control the digital recreations of their physical bodies (ie, avatars) to perform practice behaviors in the virtual environment [21]. Several studies have demonstrated that VR rehabilitation programs may develop physical outcomes [69-72]; outperform traditional rehabilitation programs [17,73,74]; and provide greater benefits to improve walking speed, balance, and mobility [75]. Many VR programs are designed with tasks that challenge the user, and well-designed VR rehabilitation programs could lead to improvements in cognitive, motor, and social aspects [76,77]. Given that VR can be accessed without supervision, an increased dosage of therapy can be provided without increasing staffing levels [77]. Remotely supervised VR interventions could complement conventional therapy programs and would be particularly suited in situations where cost savings are mandatory and/or when transport to the clinic is difficult for the patient [25]. From the patients' perspective, VR environments for rehabilitation are naturally considered to be more exciting and enjoyable than traditional rehabilitation methods, which means that patients exert more effort and are more motivated to participate in VR rehabilitation compared with traditional rehabilitation methods [21,64,78,79]. As a result, patients benefit from greater autonomy and mobility in performing their daily activities [17]. However, few studies have investigated whether increased motivation explicitly leads to positive rehabilitation outcomes [21]. As VR is still a fairly recent addition to delivering physical therapy and rehabilitation, there is uncertainty on the benefits, despite the emerging research in this area [22,80]. However, as VR technology becomes more accessible and affordable, it will likely become more widely used in clinical rehabilitation settings [80-82]. Therefore, additional trials are required to evaluate the efficacy and determine the acceptability and feasibility of VR to guide future design and implementation of these systems in clinical practice [22,42].

Methods

Study Design

Patients participated in an 8-week trial using the Pulmonary Rehabilitation in Virtual Reality (*PR in VR*) program, and both qualitative and quantitative data were collected. During the trial, each patient was provided with a VR headset which was manufactured by Pico Interactive (Pico) Goblin and whose concept is developed by Concept Health Technologies. The headset was embedded with the PR in VR app and a small probe Nonin 3150 at home for an 8-week period. The patients wore the probe when exercising to measure their heart rate (HR) and oxygen saturation levels (displayed as a percentage). Pulse oximeter data were used by clinical staff to remotely measure the patients' oxygen saturation and HR when the patients were exercising. The clinical staff used a web-based dashboard to observe these data remotely during the rehabilitation session. PR in VR was designed to enable patients to perform various exercises in their homes as displayed in the VR environment and comprised 8 separate modules. It was anticipated that the patients would spend at least 20 min per day using the VR rehabilitation program and complete 1 module per day during the 8-week period. Specifically, the app was divided into 2 subgroups: (1) education and (2) rehabilitation. Each subgroup has additional modules that can be accessed from within the app. The education section contains high definition videos with cinematic effects (visuals and audio) to increase retention of patients. The rehabilitation section (modules 5-7) comprises physical exercises led by a virtual instructor in the form of a 3D avatar. The physical exercises were drawn from the traditional pulmonary rehabilitation program and tailored to be suitable for patients while wearing the VR headset (eg, seated exercises). The final module (module 8) is a summary of the PR in VR program. The app is designed to send these data in real time to the server, and patients can monitor these data while exercising as they are displayed in the corner of the VR display screen.

XSL•FO
RenderX

Participants

A total of 10 elderly patients with COPD (graded at MRC 4 or 5), aged between 63 and 75 years, were recruited in the United Kingdom. Overall, 6 patients from a health practice in South Cumbria and 4 patients from a general hospital in West Cumbria were recruited by the local physiotherapist at each location. Ethical approval was granted by the ethical committee of the Manchester Metropolitan University before data collection. Patients graded MRC 4 and 5 were provided the option to trial the PR in VR program or wait for 1 year to attend the traditional rehabilitation classes at the local rehabilitation center. All patients invited to take part in the trial agreed; however, because of the limited number of devices, only the patients who confirmed their interest first were recruited. There were 6 out of 10 male patients. Table 1 provides the patients' demographic profile and entails the coding (patient [P]; P1-P10) used for the qualitative analysis. The average age of male patients in this study was 70 (SD 4.80) years, and the female participants were aged 3 years younger than the male participants on average (mean 67 years, SD 5.2 years).

Table 1. Demographic profile of the patients.

Patient number	Age (years)	Gender
P1	75	Male
P2	68	Male
P3	69	Male
P4	73	Female
P5	62	Female
P6	71	Male
P7	76	Male
P8	69	Female
P9	63	Female
P10	63	Male

Data Collection

Before using PR in VR, the functional ability of the patients was assessed, and together with the physiotherapist, they were each asked to set a short-term goal (eg, increase walking distance without breathlessness) and a long-term goal (eg, feeling confident to leave the house on his/her own). The patients completed 5 self-reported surveys before and after using PR in VR.

In health and well-being research, qualitative methods are widely used as they provide insights into the perceptions and experiences of patients and health care professionals [83]. During June 2018, 2 focus groups and 6 one-to-one interviews were conducted in Cumbria. Each focus group was led by 2 researchers, 1 physiotherapist, and 1 health care assistant in each location. The first focus group, comprising 3 patients with COPD, was conducted at a health practice in South Cumbria and lasted approximately 75 min. Later, 4 one-to-one interviews were conducted by the physiotherapist in the same location and lasted between 11 min and 25 min each (interview 1: 12 min; interview 2: 11 min; interview 3: 15 min; and interview 4: 23 min). The second focus group, comprising 2 patients with COPD, took place at a general hospital in West Cumbria and lasted approximately 45 min. Then, the physiotherapist and health care assistant conducted 2 more one-to-one interviews at the same location, which lasted 16 min (interview 5) and 15 min (interview 6). The focus groups and interviews were conducted at the end of each patient's 8-week VR-based rehabilitation program. During the focus groups and interviews, the questions aimed to explore how the patients benefited from the PR in VR program, explore how satisfied they were with using it, and identify the areas requiring further development. Drawing on the integrated educational aspect, the questions explored the effectiveness and benefits of immersive training in this specific context and further explored the usability of the VR device and app as well as the patients' intention to use PR in VR in the future.

Data Analysis

The qualitative data were transcribed verbatim before being analyzed using thematic analysis, which is a method used to identify patterns (ie, themes) within a dataset through a rigorous process of identification, analysis, organization, description, and reporting of those themes [84]. Thematic analysis is widely used in health care research [55,85-87] and was considered the most appropriate for this study as it is suited to questions related to people's experiences or views and perceptions such as in this study.

In addition, the patients completed the 5 self-reported surveys before and after the trial to ensure that a comparison could be made between each patient's scores. This was important to determine whether the patient showed improvements after using the PR in VR program. An analysis of the survey results before and after using VR was conducted using Excel.

Results

Qualitative Analysis

The purpose of the focus groups and interviews was to explore how the patients benefited from the PR in VR program, explore how satisfied they were with using it, and identify the areas requiring further development. In addition, the effectiveness

XSL•FO
RenderX

and benefits of immersive learning in this specific context, the usability of the VR device and app, and the patients' intention to use PR in VR in the future were explored. From the analysis, 11 themes emerged and are presented in Table 2. The themes are discussed in further detail below and provide support with direct quotations from the qualitative data collection.

Table 2. Themes and description of each theme.

Themes	Description
Increased compliance	Significant increase in the patients' compliance with pulmonary rehabilitation (ie, doing their exercises)
Increased engagement	Increased engagement in pulmonary rehabilitation when using virtual reality because of enjoyment
Physical improvements	Significant improvements in patients' physical health (ie, strength, mobility, and flexibility)
Improved psychological well-being	Patients' psychological well-being has significantly improved
Improved health-related quality of life	Patients feel healthier and fitter (ie, they can confidently leave the house and socialize more than before)
Increased confidence	Significant improvements in confidence in terms of the patients managing their condition and performing daily activities and physical exercises
Patient satisfaction	Patients are satisfied with completing the Pulmonary Rehabilitation in Virtual Reality program and achieving their short-term and long-term goals
Increased feeling of security	Patients feel more secure, reassured, and confident to exercise, knowing their physiological data are being remotely supervised
Effective immersive learning	Effectiveness of immersive learning for patients with COPD[a] was demonstrated
Personalized programs	A recommendation to provide programs tailored to suit various levels of COPD
Need for technological improvements	Need for technological improvements of both the device (eg, a more lightweight headset) and the content of the app (eg, additional functions to control the pace and standing-up exercises) to improve the patients' experience

[a]COPD: chronic obstructive pulmonary disease.

Increased Compliance

The findings demonstrate that PR in VR significantly increased the patients' (P1-P10) compliance with the frequency and consistency of performing their exercises. For example, patients had difficulty attending the rehabilitation center because of barriers associated with time and other personal commitments taking priority:

> *I couldn't complete traditional rehabilitation. There was always something cropped up.* [P9]

In comparison, using PR in VR remotely was "easier" (P9) and "pleasurable rather than chore-like" (P9), given that it could be integrated into their daily routine in a more flexible basis. For patients with COPD, this is an important factor as several patients (P1-P10) mentioned that some days they do not feel well enough, physically and/or mentally, to travel to the rehabilitation center. Therefore, being able to access pulmonary rehabilitation anywhere at any time is highly beneficial and attractive for this target group:

> *VR is more akin to my needs. I did not feel like traditional classes were doing anything for me.* [P9]
>
> *I prefer to do it at home, partly because of getting to the venue.* [P2]

Increased Engagement

Compared with previous pulmonary rehabilitation methods, PR in VR increased engagement for all participants (P1-P10) as it was more enjoyable, and more importantly, having the 3D avatar gave the experience a social aspect. Therefore, patients were more disciplined with PR in VR than they had been with previous methods such as the booklet:

> *I was amazed how I actually looked forward to doing [PR in VR] and getting on it and seeing the chap and doing [the exercises] with him.* [P8]
>
> *You discipline yourself to use it. If it's not there, then I don't think I would. You just don't do them [exercises using the booklet]...but [with VR] the guy is there and he's talking to you, and you feel as though you are with the two people, it's just the feeling it gives you.* [P9]

Patients were more engaged with PR in VR as they felt more comfortable performing their exercises in their home environment at their own leisure:

> *Because it was at home, I think I did it more. Whereas I would have been ringing the class to tell them I cannot make it because I don't feel well enough.* [P8]

Physical Improvements

Participants (P1, P3-P5, and P7-P10) demonstrated physical improvements, including increased strength, mobility, and flexibility:

> *I know certainly my legs are stronger. I can feel that even just getting up off the chair.* [P8]
>
> *I can't walk to the car without having to stop usually. But when I am not doing the exercises, I could walk to the car without having to stop.* [P4]
>
> *I am a lot more flexible and without the pain!* [P1]

XSL•FO
RenderX

Several exercises in the PR in VR program focused on developing strength in the thigh muscles, which subsequently improved the patients' cardiovascular fitness, enabling them to walk further (P1-P10). In addition, patients reported increased strength in their upper body, which resulted in them feeling less breathless and having a much quicker recovery time when breathlessness occurred. As a result of the consistent physical activity, as opposed to minimal daily movement, patients felt healthier both physically and mentally:

> *After using VR, I am less breathless...my recovery time is much quicker.* [P9]
>
> *I do feel healthier in myself...I feel a lot better with myself, with the movement, physically. I used to just sit in my chair and do nothing.* [P1]

Improved Psychological Well-Being

In relation to the previous theme (physical improvements), the patients' psychological well-being significantly improved:

> *When I've finished the Thai Chi, I do feel physically and mentally relieved.* [P1]

All patients (P1-P10) confirmed that they would recommend the PR in VR program to others in their situation largely because of the psychological benefits, which are particularly important for patients with COPD, given the associated mental health problems (eg, depression) often experienced from the initial diagnosis to adapting to living with COPD. In line with this, motivation played a key role in the patients' enthusiasm to continue with PR in VR. Overall, this could assist in overcoming mental health problems and improving their general psychological well-being:

> *Motivation is my main reason to continue. Because I get depressed really easily.* [P4]
>
> *It gets me motivated and I'm in the [wheel]chair a lot because of my health so doing the exercises does motivate me but sometimes I don't like being honest about it, but I have to mentally build up to do it because of the way I am feeling. But yes, I would recommend it.* [P3]

Improved Health-Related Quality of Life

Improvements in physical and psychological well-being were positively associated with health-related quality of life (HRQoL), as demonstrated by all participants (P1-P10). Performing daily activities was more enjoyable than before, and patients were more sociable:

> *It has made my daily activities easier...I am more comfortable with what I do.* [P9]
>
> *I am getting out and about now.* [P7]

For example, P4 had not left the house for almost 2 months before starting PR in VR. In addition, patients felt happier as they could spend more quality time with friends and family:

> *From sitting on a sofa, I had been able to go with my husband for a coffee, to go out in the car, and I had not done any of that for 7 weeks. It had got me to the stage again where I felt I could go out and go shopping with the trolley. I have done none of that for about 2 months.* [P4]
>
> *Going out every day has made a difference; I have been seeing my friends.* [P7]

Increased Confidence

Most patients (P1-P5 and P7-P10) reported significant improvements in confidence in terms of managing their condition, conducting daily activities, participating in social activities, and managing their breathing:

> *I can get upstairs in one go now.* [P7]
>
> *It builds my confidence up because I can go out more.* [P4]
>
> *It has given me more confidence in breathing along with what I am doing.* [P9]

Reflecting on this, patients discussed how they had previously lost motivation and confidence in their ability to manage tasks themselves without the support of their partners and several had "given up" (P3, P4, and P7-P10). However, with the PR in VR program, the patients had gained back more control in their lives:

> *I think I had given up. I decided that the illness had gone so far that I was not going to get better, I was not going to get out again, or enjoy life. It has completely reassured me that way and given me the energy to get going again.* [P8]
>
> *Possibly it was a mental thing. I felt confident enough. I started filling my own flasks.* [P7]

Patient Satisfaction

Successful completion of their exercises within a given day using PR in VR was positively associated with patient satisfaction (P1-P10). Noticing improvements in physical wellness enabled patients to feel satisfied and confirmed the benefits associated with following the PR in VR program:

> *What I enjoyed most was knowing that I could get through the exercises. Knowing that my oxygen was not dropping. It was a real achievement for me. I was used to going to the toilet and coming back and having to catch my breath. All of a sudden, I could do these exercises and after a few times I was still breathing normal.* [P8]

Given that PR in VR is self-managed and remotely supervised, patients' self-motivation played a critical factor in their consistency with exercising and subsequent self-satisfaction:

> *I feel satisfied that I motivated myself to do it. It did help me to do it. I feel satisfied when I sit down and think that's me done for tonight.* [P4]

Increased Feeling of Security

A primary concern for patients was experiencing exasperation when exercising alone at home. In comparison to previous methods, PR in VR is advantageous as it allows for remote supervision, meaning that the patients' physiological data (ie, HR and oxygen level) are continuously tracked and monitored by health practitioners. Therefore, this function improved the

XSL•FO
RenderX

patients' experience as they felt more secure (P9), confident (P7 and P9), and reassured (P8), knowing that they were being monitored while exercising in real time should exacerbation occur:

> *The minute I am getting out of breath [at home] I go into panic.* [P8]
>
> *It has given me more confidence in breathing along with what I am doing.* [P9]
>
> *It reassured me being able to see my oxygen level and heart rate while I was exercising. I feel it has increased my confidence.* [P8]

Furthermore, this function assisted patients with better management of their condition and to avoid overexerting themselves into exasperation:

> *[PR in VR] has improved my breath...in the VR it shows the oxygen go up and back down again, it gave me that back again where I could settle myself down.* [P8]

Effective Immersive Learning

PR in VR provides patients with information on COPD with the aim of helping them better manage their condition. The findings demonstrated the effectiveness of immersive learning for patients with COPD as it educated them on the disease, which some participants were never informed of. For instance, P1 found the learning aspect useful and "very informative to find out what is actually happening." During the focus groups, P1 could reiterate what he had learned from the program 5 weeks prior:

> *That's the first time I have seen on screen what emphysema actually is.* [P1]

The PR in VR program had increased the patients' awareness of their breathing techniques (P1, P3, and P8) and the variety of exercises required to target various muscles and support their overall health and well-being. The findings further support the effectiveness of immersive learning, specifically using VR for patients with COPD:

> *He says breath out when you do the exertion, I was not doing that before...it has improved my breath...the voice over was helpful and reassuring.* [P8]
>
> *I think it's a good idea. It reminds you that you've got to do different exercises to target different muscles.* [P3]

Personalized Programs

Recommendations to improve the patients' experience of PR in VR were provided. Patients (P1-P6, P8, and P10) suggested various levels of exercises (ie, harder set of exercises and easier set of exercises) to suit the many levels of patients with COPD to improve the program. More specifically, patients (P2, P5, and P10) explained how the first 2 to 3 levels were "too easy" (P5), slow paced (P5), and too long (P10) and were eager to push onto more "severe" modules (P10). Therefore, these patients required more advanced exercises to ensure that they were constantly being challenged. On the contrary, the more advanced exercises were too challenging for other patients (P1, P3, P4, and P5) on the first attempt and required repetition, given that they might not have exercised for a prolonged period before using PR in VR. Indeed, those patients enjoyed the flexibility of being able to move back and forth throughout the modules depending on their well-being at that time:

> *It is beneficial to have the different levels so I can drop down depending how I feel each day.* [P4]

Furthermore, it was suggested that a variety of exercises at the same difficulty level would retain engagement:

> *I would like two or more routines in each module instead of doing the same thing all the time. For longer than six weeks it would become boring.* [P5]

Needs for Technological Improvements

Furthermore, to improve the patients' experience of PR in VR, technological improvements of both the device and app were required (P1-P10). For instance, the main technical issue within the app whereby the camera moved to the right and took several seconds to re-center was a recurrent topic of conversation. Hence, improved graphics could enhance the experience but was considered a minor factor in the overall rehabilitation experience that could easily be solved:

> *The whole thing was excellent it really was except for the glitches but I am sure that can be overcome.* [P5]

Suggestions for additional functions included a fast-forward and pause button (P1-P5) to allow patients to have more control over the program and pause it while they recovered:

> *If you could pause it when you know that you're going to get out of breath.* [P1]

All patients (P1-P10) found the headset easy to use; however, some required it to be more lightweight (P4, P5, and P9). With regard to the content, some patients (P1-P5) enjoyed the standing-up exercises.

Quantitative Data Analysis

In addition to the focus groups and interviews, quantitative data were collected, and the findings further support the qualitative findings. Table 3 presents the results of the outcome measures. The average of each of the 5 surveys was calculated based on the patient data gathered before and after the VR rehabilitation program. As can be seen, the findings demonstrate an improvement of all outcome measures after patients used the VR program. The Chronic Respiratory Disease Questionnaire (CRQ) measures both physical and emotional aspects of chronic respiratory disease. The results indicate that the patients had improved dyspnea, fatigue, and emotional function after completing VR-assisted treatment. The female participants collectively showed more improvements in both dyspnea and emotional function than the male participants. The Patient Health Questionnaire-9 and generalized anxiety disorder-7 demonstrated that a substantial proportion of patients had reduced feelings of depression and anxiety, respectively, which further supported the theme of improvements in the patients' psychological well-being. The Short Physical Performance Battery is used to assess lower extremity physical performance status by combining the results of the gait speed, chair stand, and balance tests. The results of the Edmonton Frail Scale and

XSL•FO
RenderX

Short Physical Performance Battery were positively associated with the patients' physical improvements, including strength, mobility, and flexibility. Finally, the Patient Activation Measure results demonstrated the patients' progression with knowledge, skills, and confidence in self-managing their condition after using VR.

Table 3. Outcome measure results before and after using virtual reality.

Outcome measures	Before using virtual reality, mean (SD)	After using virtual reality, mean (SD)
Short Physical Performance Battery	6.78 (2.95)	8.43 (1.72)
CRQ[a]–dyspnea	2.22 (1.09)	2.96 (1.15)
CRQ–fatigue	3.11 (1.43)	3.27 (1.15)
CRQ–emotional	3.85 (1.52)	4.36 (1.01)
CRQ–mastery	3.83 (1.19)	4.22 (0.74)
Patient Activation Measure	60.81 (10.83)	63.46 (15.89)
Edmonton Frail Scale	6.56 (2.07)	5.38 (2.00)
Patient Health Questionnaire-9	8.44 (5.90)	6.11 (4.76)
Generalized Anxiety Disorder-7	5.78 (5.89)	4.22 (3.11)

[a]CRQ: Chronic Respiratory Disease Questionnaire.

Discussion

Principal Findings

This study aimed to assess the benefits and outcomes of VR as an innovative method for pulmonary rehabilitation in 2 aspects. First, the study aimed to investigate whether VR improves compliance with pulmonary rehabilitation in patients with COPD, particularly the vulnerable patient group (MRC 4 or 5). Finding innovative pulmonary rehabilitation treatment methods that engage this particular patient group is vital, given the severity of the disease that affects the HRQoL and the barriers associated with attending traditional rehabilitation classes [12-15]. This study demonstrates that PR in VR significantly increases the patients' compliance with pulmonary rehabilitation compared with traditional methods mainly because of the flexibility to exercise at any time of the day and location and as it is not restricted to home- or clinic-based scenarios, which is more accommodating for their daily psychological and physical well-being [88]. The second research question was to investigate whether VR provides a credible alternative to traditional pulmonary rehabilitation. Thematic analysis revealed nine themes illustrating the benefits associated with PR in VR compared with traditional methods.

Conclusions

The patients demonstrated significant improvements in physical ability and psychological well-being because of their consistency with exercises, thus improving their HRQoL. These findings are consistent with both the quantitative results of this study and previous research investigating the benefits of VR for other chronic diseases [17,64,69-71,73,74,78] and provide evidence specific to patients with COPD. Owing to its remote accessibility, PR in VR provides a solution to overcome the associated barriers of attending traditional rehabilitation classes (ie, transport and travel, fatigue, motivation, inconvenience, burden of illness, disruption caused to daily activities, low awareness of rehabilitation, and knowledge of and disbelief in the beneficial outcomes of the rehabilitation) [12-15]. Indeed, patients feel secure and confident enough to exercise without face-to-face supervision as health practitioners can remotely monitor their physiological performance data when using PR in VR. Pulmonary rehabilitation is delivered to a low proportion of the population despite its demonstrated benefits for patients with COPD [8]. However, PR in VR can be deployed at scale, which provides a solution to the recurring issue of too long a waiting list for rehabilitation classes [10] and allows health practitioners to measure patient performance in real-time. Overall, this study demonstrates how remotely supervised VR-based pulmonary rehabilitation can help to overcome current issues and limitations associated with providing this service to patients with COPD at scale. Compared with previous methods, VR could provide a more cost-effective solution for the health treatment to deliver pulmonary rehabilitation to a vast majority of this patient group nationwide.

Contribution, Implications, and Future Research

Key Contributions

Previous studies have investigated the benefits of using VR for rehabilitation for long-term conditions [17,18,89]. However, its benefits for self-managed and remotely supervised pulmonary rehabilitation remain unexplored, and to the best of the researchers' knowledge, this study is one of the first to provide empirical evidence using qualitative data that are further supported by quantitative results. Therefore, this study contributes important findings demonstrating that PR in VR represents an effective form of self-managed and remotely supervised pulmonary rehabilitation that can be delivered at scale. Overall, this study provides a significant contribution to health care research investigating the benefits of implementing digital technologies to improve rehabilitation, with particular contribution to pulmonary rehabilitation research. Moreover, this study provides important practical implications for the health care industry and medical practitioners on the benefits of VR for pulmonary rehabilitation. PR in VR is an innovative app distinct from previous self-managed rehabilitation methods as it allows health practitioners to supervise patients with COPD

XSL•FO
RenderX

remotely and at scale and to immediately measure patient performance. This investigation offers proof of concept that PR in VR intervention can be used for elderly patients with COPD. With evidence, this alternative intervention platform provides an interactive and visually stimulating approach for use in clinical treatment and could provide pulmonary rehabilitation to patients with COPD who cannot easily access traditional rehabilitation methods.

Possible Future Research

This study has several limitations that could be mitigated by further research. The first limitation pertains to the limited sample size; however, this is common in other exploratory studies investigating VR for rehabilitation [18,89]. This study employed 10 patients with COPD, aged between 63 and 75 years, from the United Kingdom. Although small sample sizes can provide important initial inferences about a topic, they can provide few firm conclusions [21]. However, as this is the first study investigating VR for pulmonary rehabilitation, a small sample size offers promising pilot data as a first phase trial. Furthermore, COPD in the United Kingdom is common among people aged ≥40 years [90]; therefore, it is important that future studies employ a larger sample covering the majority of the respective age groups.

Conflicts of Interest

None declared.

References

1. Hayton C, Clark A, Olive S, Browne P, Galey P, Knights E, et al. Barriers to pulmonary rehabilitation: characteristics that predict patient attendance and adherence. Respir Med 2013 Mar;107(3):401-407 [FREE Full text] [doi: 10.1016/j.rmed.2012.11.016] [Medline: 23261311]
2. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015 Feb 23(2):CD003793. [doi: 10.1002/14651858.CD003793.pub3] [Medline: 25705944]
3. Sundararajan L, Balami J, Packham S. Effectiveness of outpatient pulmonary rehabilitation in elderly patients with chronic obstructive pulmonary disease. J Cardiopulm Rehabil Prev 2010;30(2):121-125. [doi: 10.1097/HCR.0b013e3181be7c56] [Medline: 19952774]
4. Norweg AM, Whiteson J, Malgady R, Mola A, Rey M. The effectiveness of different combinations of pulmonary rehabilitation program components: a randomized controlled trial. Chest 2005 Aug;128(2):663-672. [doi: 10.1378/chest.128.2.663] [Medline: 16100152]
5. Lacasse Y, Martin S, Lasserson TJ, Goldstein RS. Meta-analysis of respiratory rehabilitation in chronic obstructive pulmonary disease. A Cochrane systematic review. Eura Medicophys 2007 Dec;43(4):475-485 [FREE Full text] [Medline: 18084170]
6. Ries AL, Bauldoff GS, Carlin BW, Casaburi R, Emery CF, Mahler DA, et al. Pulmonary rehabilitation: joint ACCP/AACVPR evidence-based clinical practice guidelines. Chest 2007 May;131(5 Suppl):4S-42S. [doi: 10.1378/chest.06-2418] [Medline: 17494825]
7. Hogg L, Garrod R, Thornton H, McDonnell L, Bellas H, White P. Effectiveness, attendance, and completion of an integrated, system-wide pulmonary rehabilitation service for COPD: prospective observational study. COPD 2012 Aug;9(5):546-554. [doi: 10.3109/15412555.2012.707258] [Medline: 23030586]
8. Seidman Z, McNamara R, Wootton S, Leung R, Spencer L, Dale M, et al. People attending pulmonary rehabilitation demonstrate a substantial engagement with technology and willingness to use telerehabilitation: a survey. J Physiother 2017 Jul;63(3):175-181 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.010] [Medline: 28652080]
9. NHS Greater Glasgow and Clyde. 2018. MRC Grade: Functional Ability / Breathlessness Scale URL: http://www.nhsggc.org.uk/about-us/professional-support-sites/cdm-local-enhanced-services/copd/symptoms/mrc-grade/ [accessed 2020-05-11]
10. Royal College of Physicians London. 2017. Pulmonary Rehabilitation: Beyond Breathing Better URL: https://www.rcplondon.ac.uk/projects/outputs/pulmonary-rehabilitation-beyond-breathing-better [accessed 2020-05-25]
11. Williams N. The MRC breathlessness scale. Occup Med (Lond) 2017 Aug 1;67(6):496-497. [doi: 10.1093/occmed/kqx086] [Medline: 28898975]
12. Liu Y, Dickerson T, Early F, Fuld J, Clarkson PJ. Understanding influences on the uptake of pulmonary rehabilitation in the east of England: an inclusive design/mixed-methods study protocol. BMJ Open 2018 Apr 24;8(4):e020750 [FREE Full text] [doi: 10.1136/bmjopen-2017-020750] [Medline: 29691248]
13. Keating A, Lee A, Holland AE. What prevents people with chronic obstructive pulmonary disease from attending pulmonary rehabilitation? A systematic review. Chron Respir Dis 2011;8(2):89-99. [doi: 10.1177/1479972310393756] [Medline: 21596892]
14. Cox NS, Oliveira CC, Lahham A, Holland AE. Pulmonary rehabilitation referral and participation are commonly influenced by environment, knowledge, and beliefs about consequences: a systematic review using the theoretical domains framework. J Physiother 2017 Apr;63(2):84-93 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.002] [Medline: 28433238]

XSL•FO
RenderX

15. Foster F, Piggott R, Riley L, Beech R. Working with primary care clinicians and patients to introduce strategies for increasing referrals for pulmonary rehabilitation. Prim Health Care Res Dev 2016 May;17(3):226-237. [doi: 10.1017/S1463423615000286] [Medline: 26072909]
16. Johnston KN, Young M, Grimmer KA, Antic R, Frith PA. Barriers to, and facilitators for, referral to pulmonary rehabilitation in COPD patients from the perspective of Australian general practitioners: a qualitative study. Prim Care Respir J 2013 Sep;22(3):319-324 [FREE Full text] [doi: 10.4104/pcrj.2013.00062] [Medline: 23797679]
17. de Melo GE, Kleiner A, Lopes J, Dumont AJ, Lazzari RD, Galli M, et al. Effect of virtual reality training on walking distance and physical fitness in individuals with Parkinson's disease. NeuroRehabilitation 2018;42(4):473-480. [doi: 10.3233/NRE-172355] [Medline: 29660956]
18. Yang W, Wang H, Wu R, Lo C, Lin K. Home-based virtual reality balance training and conventional balance training in Parkinson's disease: a randomized controlled trial. J Formos Med Assoc 2016 Sep;115(9):734-743 [FREE Full text] [doi: 10.1016/j.jfma.2015.07.012] [Medline: 26279172]
19. Warland A, Paraskevopoulos I, Tsekleves E, Ryan J, Nowicky A, Griscti J, et al. The feasibility, acceptability and preliminary efficacy of a low-cost, virtual-reality based, upper-limb stroke rehabilitation device: a mixed methods study. Disabil Rehabil 2019 Sep;41(18):2119-2134. [doi: 10.1080/09638288.2018.1459881] [Medline: 29644897]
20. Shin J, Park SB, Ho Jang S. Effects of game-based virtual reality on health-related quality of life in chronic stroke patients: a randomized, controlled study. Comput Biol Med 2015 Aug;63:92-98. [doi: 10.1016/j.compbiomed.2015.03.011] [Medline: 26046499]
21. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Hum Behav 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
22. Yamato TP, Pompeu JE, Pompeu SM, Hassett L. Virtual reality for stroke rehabilitation. Phys Ther 2016 Oct;96(10):1508-1513. [doi: 10.2522/ptj.20150539] [Medline: 27081204]
23. Levack W, Watson J, Hay-Smith E, Davies C, Ingham T, Jones B, et al. Factors influencing referral to and uptake and attendance of pulmonary rehabilitation for chronic obstructive pulmonary disease: A qualitative evidence synthesis of the experiences of service users, their families, and healthcare providers (protocol). Cochrane Database Syst Rev 2018;2018(11):CD013195. [doi: 10.1002/14651858.cd013195]
24. May SM, Li JT. Burden of chronic obstructive pulmonary disease: healthcare costs and beyond. Allergy Asthma Proc 2015;36(1):4-10. [doi: 10.2500/aap.2015.36.3812] [Medline: 25562549]
25. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Mar;96(3):418-25.e2. [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
26. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS, GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO global initiative for chronic obstructive lung disease (GOLD) workshop summary. Am J Respir Crit Care Med 2001 Apr;163(5):1256-1276. [doi: 10.1164/ajrccm.163.5.2101039] [Medline: 11316667]
27. Rutten-van Mölken MP, Oostenbrink JB, Tashkin DP, Burkhart D, Monz BU. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? Chest 2006 Oct;130(4):1117-1128. [doi: 10.1378/chest.130.4.1117] [Medline: 17035446]
28. Williams V, Bruton A, Ellis-Hill C, McPherson K. The importance of movement for people living with chronic obstructive pulmonary disease. Qual Health Res 2011 Sep;21(9):1239-1248. [doi: 10.1177/1049732311405801] [Medline: 21511979]
29. Casey M, Mulkerns A, O?Donnell C, McDonnell T. IntechOpen. 2019. Pulmonary Rehabilitation in COPD: Current Practice and Future Directions URL: https://www.intechopen.com/books/copd-an-update-in-pathogenesis-and-clinical-management/pulmonary-rehabilitation-in-copd-current-practice-and-future-directions [accessed 2020-05-26]
30. Connolly MJ, Yohannes AM. The impact of depression in older patients with chronic obstructive pulmonary disease and asthma. Maturitas 2016 Oct;92:9-14. [doi: 10.1016/j.maturitas.2016.07.005] [Medline: 27621232]
31. Hill K, Geist R, Goldstein RS, Lacasse Y. Anxiety and depression in end-stage COPD. Eur Respir J 2008 Mar;31(3):667-677 [FREE Full text] [doi: 10.1183/09031936.00125707] [Medline: 18310400]
32. di Marco F, Verga M, Reggente M, Casanova FM, Santus P, Blasi F, et al. Anxiety and depression in COPD patients: the roles of gender and disease severity. Respir Med 2006 Oct;100(10):1767-1774 [FREE Full text] [doi: 10.1016/j.rmed.2006.01.026] [Medline: 16531031]
33. Kühl K, Schürmann W, Rief W. Mental disorders and quality of life in COPD patients and their spouses. Int J Chron Obstruct Pulmon Dis 2008;3(4):727-736 [FREE Full text] [doi: 10.2147/copd.s3375] [Medline: 19281087]
34. Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. Gen Hosp Psychiatry 2007;29(2):147-155. [doi: 10.1016/j.genhosppsych.2006.11.005] [Medline: 17336664]
35. Yohannes AM, Baldwin RC, Connolly MJ. Depression and anxiety in elderly outpatients with chronic obstructive pulmonary disease: prevalence, and validation of the BASDEC screening questionnaire. Int J Geriatr Psychiatry 2000 Dec;15(12):1090-1096. [doi: 10.1002/1099-1166(200012)15:12<1090::aid-gps249>3.0.co;2-l] [Medline: 11180464]

XSL•FO
RenderX

36. Agle DO, Baum GL. Psychological aspects of chronic obstructive pulmonary disease. Med Clin North Am 1977 Jul;61(4):749-758. [doi: 10.1016/s0025-7125(16)31294-9] [Medline: 559893]
37. Mikkelsen RL, Middelboe T, Pisinger C, Stage KB. Anxiety and depression in patients with chronic obstructive pulmonary disease (COPD). A review. Nord J Psychiatry 2004;58(1):65-70. [doi: 10.1080/08039480310000824] [Medline: 14985157]
38. Paine NJ, Bacon SL, Bourbeau J, Tan WC, Lavoie KL, CanCOLD Study Investigators, Canadian Respiratory Research Network and the CanCOLD Collaborative Research Group. Psychological distress is related to poor health behaviours in COPD and non-COPD patients: evidence from the CanCOLD study. Respir Med 2019 Jan;146:1-9 [FREE Full text] [doi: 10.1016/j.rmed.2018.11.006] [Medline: 30665505]
39. Robinson H, Williams V, Curtis F, Bridle C, Jones AW. Facilitators and barriers to physical activity following pulmonary rehabilitation in COPD: a systematic review of qualitative studies. NPJ Prim Care Respir Med 2018 Jun 4;28(1):19 [FREE Full text] [doi: 10.1038/s41533-018-0085-7] [Medline: 29867117]
40. Wedzicha JA, Bestall JC, Garrod R, Garnham R, Paul EA, Jones PW. Randomized controlled trial of pulmonary rehabilitation in severe chronic obstructive pulmonary disease patients, stratified with the MRC dyspnoea scale. Eur Respir J 1998 Aug;12(2):363-369 [FREE Full text] [doi: 10.1183/09031936.98.12020363] [Medline: 9727786]
41. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized telehealth in the future: a global research agenda. J Med Internet Res 2016 Mar 1;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
42. Laver K, Schoene D, Crotty M, George S, Lannin N, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2013 Dec 16(12):CD010255 [FREE Full text] [doi: 10.1002/14651858.CD010255.pub2] [Medline: 24338496]
43. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: review of the state-of-the-art and areas of application. JMIR Rehabil Assist Technol 2017 Jul 21;4(2):e7 [FREE Full text] [doi: 10.2196/rehab.7511] [Medline: 28733271]
44. Tsai LL, McNamara RJ, Moddel C, Alison JA, McKenzie DK, McKeough ZJ. Home-based telerehabilitation via real-time videoconferencing improves endurance exercise capacity in patients with COPD: the randomized controlled TeleR study. Respirology 2017 May;22(4):699-707 [FREE Full text] [doi: 10.1111/resp.12966] [Medline: 27992099]
45. Marquis E, Jung B, Schormans AF, Lukmanji S, Wilton R, Baptiste S. Developing inclusive educators: enhancing the accessibility of teaching and learning in higher education. Int J Acad Dev 2016 May 19;21(4):337-349 [FREE Full text] [doi: 10.1080/1360144x.2016.1181071]
46. Hou J, Yang R, Yang Y, Tang Y, Deng H, Chen Z, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e10201 [FREE Full text] [doi: 10.2196/10201] [Medline: 30785406]
47. Chaplin E, Hewitt S, Apps L, Bankart J, Pulikottil-Jacob R, Boyce S, et al. Interactive web-based pulmonary rehabilitation programme: a randomised controlled feasibility trial. BMJ Open 2017 Mar 31;7(3):e013682 [FREE Full text] [doi: 10.1136/bmjopen-2016-013682] [Medline: 28363923]
48. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with chronic obstructive pulmonary disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
49. Dinesen B, Huniche L, Toft E. Attitudes of COPD patients towards tele-rehabilitation: a cross-sector case study. Int J Environ Res Public Health 2013 Nov 15;10(11):6184-6198 [FREE Full text] [doi: 10.3390/ijerph10116184] [Medline: 24247995]
50. Broadbent E, Garrett J, Jepsen N, li Ogilvie V, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
51. Kairy D, Lehoux P, Vincent C, Visintin M. A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. Disabil Rehabil 2009;31(6):427-447. [doi: 10.1080/09638280802062553] [Medline: 18720118]
52. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
53. Crotty M, Killington M, van den Berg M, Morris C, Taylor A, Carati C. Telerehabilitation for older people using off-the-shelf applications: acceptability and feasibility. J Telemed Telecare 2014 Oct;20(7):370-376. [doi: 10.1177/1357633X14552382] [Medline: 25399997]
54. Shulver W, Killington M, Morris C, Crotty M. 'Well, if the kids can do it, I can do it': older rehabilitation patients' experiences of telerehabilitation. Health Expect 2017 Feb;20(1):120-129 [FREE Full text] [doi: 10.1111/hex.12443] [Medline: 26890025]
55. Tyagi S, Lim DS, Ho WH, Koh YQ, Cai V, Koh GC, et al. Acceptance of tele-rehabilitation by stroke patients: perceived barriers and facilitators. Arch Phys Med Rehabil 2018 Dec;99(12):2472-7.e2. [doi: 10.1016/j.apmr.2018.04.033] [Medline: 29902469]
56. World Health Organization. 2017. Chronic Obstructive Pulmonary Disease (COPD) URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2020-05-11]

XSL•FO
RenderX

57. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. Lancet 2012 Dec 15;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0] [Medline: 23245604]
58. Salvi S. The silent epidemic of COPD in Africa. Lancet Glob Health 2015 Jan;3(1):e6-e7 [FREE Full text] [doi: 10.1016/S2214-109X(14)70359-6] [Medline: 25539971]
59. Lomborg B. Global Problems, Smart Solutions: Costs and Benefits. Cambridge, UK: Cambridge University Press; 2013.
60. Harris D, Hayter M, Allender S. Improving the uptake of pulmonary rehabilitation in patients with COPD: qualitative study of experiences and attitudes. Br J Gen Pract 2008 Oct;58(555):703-710 [FREE Full text] [doi: 10.3399/bjgp08X342363] [Medline: 18826782]
61. Jolly K, Majothi S, Sitch A, Heneghan N, Riley RD, Moore DJ, et al. Self-management of health care behaviors for COPD: a systematic review and meta-analysis. Int J Chron Obstruct Pulmon Dis 2016;11:305-326 [FREE Full text] [doi: 10.2147/COPD.S90812] [Medline: 26937183]
62. Zwerink M, Brusse-Keizer M, van der Valk PD, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2014 Mar 19(3):CD002990 [FREE Full text] [doi: 10.1002/14651858.CD002990.pub3] [Medline: 24665053]
63. Trappenburg J, Jonkman N, Jaarsma T, van Os-Medendorp H, Kort H, de Wit N, et al. Self-management: one size does not fit all. Patient Educ Couns 2013 Jul;92(1):134-137. [doi: 10.1016/j.pec.2013.02.009] [Medline: 23499381]
64. Bryanton C, Bossé J, Brien M, McLean J, McCormick A, Sveistrup H. Feasibility, motivation, and selective motor control: virtual reality compared to conventional home exercise in children with cerebral palsy. Cyberpsychol Behav 2006 Apr;9(2):123-128. [doi: 10.1089/cpb.2006.9.123] [Medline: 16640463]
65. Brütsch K, Koenig A, Zimmerli L, Mérillat-Koeneke S, Riener R, Jäncke L, et al. Virtual reality for enhancement of robot-assisted gait training in children with central gait disorders. J Rehabil Med 2011 May;43(6):493-499 [FREE Full text] [doi: 10.2340/16501977-0802] [Medline: 21491072]
66. Brütsch K, Schuler T, Koenig A, Zimmerli L, -Koeneke SM, Lünenburger L, et al. Influence of virtual reality soccer game on walking performance in robotic assisted gait training for children. J Neuroeng Rehabil 2010 Apr 22;7:15 [FREE Full text] [doi: 10.1186/1743-0003-7-15] [Medline: 20412572]
67. Camporesi C, Kallman M, Han J. VR Solutions for Improving Physical Therapy. In: Proceedings of the IEEE Virtual Reality Conference. 2013 Presented at: VR'13; March 18-20, 2013; Lake Buena Vista, FL, USA. [doi: 10.1109/vr.2013.6549371]
68. Keller MS, Park HJ, Cunningham ME, Fouladian JE, Chen M, Spiegel BM. Public perceptions regarding use of virtual reality in health care: a social media content analysis using Facebook. J Med Internet Res 2017 Dec 19;19(12):e419 [FREE Full text] [doi: 10.2196/jmir.7467] [Medline: 29258975]
69. Cho KH, Lee KJ, Song CH. Virtual-reality balance training with a video-game system improves dynamic balance in chronic stroke patients. Tohoku J Exp Med 2012 Sep;228(1):69-74 [FREE Full text] [doi: 10.1620/tjem.228.69] [Medline: 22976384]
70. Rahman S. Efficacy of virtual reality-based therapy on balance in children with down syndrome. World Appl Sci J 2010;10(3):254-261 [FREE Full text]
71. Rostami HR, Arastoo AA, Nejad SJ, Mahany MK, Malamiri RA, Goharpey S. Effects of modified constraint-induced movement therapy in virtual environment on upper-limb function in children with spastic hemiparetic cerebral palsy: a randomised controlled trial. NeuroRehabilitation 2012;31(4):357-365. [doi: 10.3233/NRE-2012-00804] [Medline: 23232158]
72. Mosadeghi S, Reid MW, Martinez B, Rosen BT, Spiegel BM. Feasibility of an immersive virtual reality intervention for hospitalized patients: an observational cohort study. JMIR Ment Health 2016 Jun 27;3(2):e28 [FREE Full text] [doi: 10.2196/mental.5801] [Medline: 27349654]
73. Chen C, Hong W, Cheng HK, Liaw M, Chung C, Chen C. Muscle strength enhancement following home-based virtual cycling training in ambulatory children with cerebral palsy. Res Dev Disabil 2012;33(4):1087-1094. [doi: 10.1016/j.ridd.2012.01.017] [Medline: 22502833]
74. Ma H, Hwang W, Fang J, Kuo J, Wang C, Leong I, et al. Effects of virtual reality training on functional reaching movements in people with Parkinson's disease: a randomized controlled pilot trial. Clin Rehabil 2011 Oct;25(10):892-902. [doi: 10.1177/0269215511406757] [Medline: 21632652]
75. Corbetta D, Imeri F, Gatti R. Rehabilitation that incorporates virtual reality is more effective than standard rehabilitation for improving walking speed, balance and mobility after stroke: a systematic review. J Physiother 2015 Jul;61(3):117-124 [FREE Full text] [doi: 10.1016/j.jphys.2015.05.017] [Medline: 26093805]
76. Lopes JP, de Almeida N, Lazzari R, Oliveira C. Virtual reality in the rehabilitation process for individuals with cerebral palsy and down syndrome: a systematic review. J Bodyw Mov Ther 2018 Jun:- (forthcoming) [FREE Full text] [doi: 10.1016/j.jbmt.2018.06.006]
77. Laver K, Lange B, George S, Deutsch J, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
78. Broeren J, Rydmark M, Sunnerhagen KS. Virtual reality and haptics as a training device for movement rehabilitation after stroke: a single-case study. Arch Phys Med Rehabil 2004 Aug;85(8):1247-1250. [doi: 10.1016/j.apmr.2003.09.020] [Medline: 15295748]

XSL•FO
RenderX

79. Lewis GN, Rosie JA. Virtual reality games for movement rehabilitation in neurological conditions: how do we meet the needs and expectations of the users? Disabil Rehabil 2012;34(22):1880-1886. [doi: 10.3109/09638288.2012.670036] [Medline: 22480353]
80. Burridge JH, Hughes A. Potential for new technologies in clinical practice. Curr Opin Neurol 2010 Dec;23(6):671-677. [doi: 10.1097/WCO.0b013e3283402af5] [Medline: 20962639]
81. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 3;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
82. Threapleton K, Drummond A, Standen P. Virtual rehabilitation: what are the practical barriers for home-based research? Digit Health 2016;2:2055207616641302 [FREE Full text] [doi: 10.1177/2055207616641302] [Medline: 29942551]
83. Braun V, Clarke V. What can 'thematic analysis' offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
84. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
85. Gill IJ, Wall G, Simpson J. Clients' perspectives of rehabilitation in one acquired brain injury residential rehabilitation unit: a thematic analysis. Brain Inj 2012;26(7-8):909-920. [doi: 10.3109/02699052.2012.661118] [Medline: 22571407]
86. Holliday RC, Ballinger C, Playford ED. Goal setting in neurological rehabilitation: patients' perspectives. Disabil Rehabil 2007 Mar 15;29(5):389-394. [doi: 10.1080/09638280600841117] [Medline: 17364791]
87. Saunders JC, Byrne MM. A thematic analysis of families living with schizophrenia. Arch Psychiatr Nurs 2002 Oct;16(5):217-223. [doi: 10.1053/apnu.2002.36234] [Medline: 12434327]
88. Zoeckler N, Kenn K, Kuehl K, Stenzel N, Rief W. Illness perceptions predict exercise capacity and psychological well-being after pulmonary rehabilitation in COPD patients. J Psychosom Res 2014 Feb;76(2):146-151. [doi: 10.1016/j.jpsychores.2013.11.021] [Medline: 24439691]
89. Didehbani N, Allen T, Kandalaft M, Krawczyk D, Chapman S. Virtual reality social cognition training for children with high functioning autism. Comput Hum Behav 2016 Sep;62:703-711 [FREE Full text] [doi: 10.1016/j.chb.2016.04.033]
90. British Lung Foundation. 2018. Chronic Obstructive Pulmonary Disease (COPD) Statistics URL: https://statistics.blf.org.uk/copd [accessed 2020-05-11]

Abbreviations

COPD: chronic obstructive pulmonary disease
HR: heart rate
HRQoL: health-related quality of life
MRC: Medical Research Council
PR: pulmonary rehabilitation
VR: virtual reality

Edited by G Eysenbach; submitted 06.04.19; peer-reviewed by R Davies, M Duplaga; comments to author 12.12.19; revised version received 31.03.20; accepted 10.04.20; published 07.07.20.

Please cite as:
Jung T, Moorhouse N, Shi X, Amin MF
A Virtual Reality–Supported Intervention for Pulmonary Rehabilitation of Patients With Chronic Obstructive Pulmonary Disease: Mixed Methods Study
J Med Internet Res 2020;22(7):e14178
URL: https://www.jmir.org/2020/7/e14178
doi:10.2196/14178
PMID:32673224

©Timothy Jung, Natasha Moorhouse, Xin Shi, Muhammad Farhan Amin. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Dietary Mobile App for Patients Undergoing Hemodialysis: Prospective Pilot Study to Improve Dietary Intakes

Cosette Fakih El Khoury[1], BSc, MSc, PhD; Rik Crutzen[2], PhD; Jos M G A Schols[1,3], MD, PhD; Ruud J G Halfens[1], PhD; Mirey Karavetian[4], BSc, MSc, PhD

[1]Department of Health Services Research, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[2]Department of Health Promotion, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[3]Department of Family Medicine, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[4]Department of Health Sciences, Zayed University, Dubai, United Arab Emirates

Corresponding Author:
Mirey Karavetian, BSc, MSc, PhD
Department of Health Sciences
Zayed University
PO Box 19282
Dubai
United Arab Emirates
Phone: 971 562446865
Email: Mirey.Karavetian@zu.ac.ae

Abstract

Background: Mobile technology has an impact on the health care sector, also within dietetics. Mobile health (mHealth) apps may be used for dietary assessment and self-monitoring, allowing for real-time reporting of food intakes. Changing eating behaviors is quite challenging, and patients undergoing hemodialysis, particularly, struggle to meet the target intakes set by dietary guidelines. Usage of mobile apps that are developed in a person-centered approach and in line with recommendations may support both patients and health care practitioners.

Objective: This study is a pilot that aims at estimating the potential efficacy of a dietary intervention using a theory-based, person-centered smartphone app. Results will be used to improve both the app and a planned large-scale trial intended to assess app efficacy thoroughly.

Methods: A prospective pilot study was performed at the hemodialysis unit of Al Qassimi Hospital (The Emirate of Sharjah). All patients that fulfilled the study inclusion criteria were considered eligible to be enrolled in the pilot study. Upon successful installation of the app, users met with a dietitian once a week. Outcomes were measured at baseline (T0) and 2 weeks post app usage (T1). This pilot is reported as per guidelines for nonrandomized pilot and feasibility studies and in line with the CONSORT 2010 checklist for reporting pilot or feasibility trials.

Results: A total of 23 patients completed the pilot intervention. Mean energy intakes increased from 24.4 kcal/kg/day (SD 8.0) to 29.1 kcal/kg/day (SD 7.8) with a medium effect size (d=0.6, 95% CI 0.0-1.2). Mean protein intakes increased from 0.9 g/kg/day (SD 0.3) to 1.3 g/kg/day (SD 0.5) with a large effect size (d=1.0, 95% CI 0.4-1.6); mean intake of high biological value (%HBV) proteins also increased from 58.6% (SD 10.1) to 70.1% (SD 10.7) with a large effect size (d=1.1, 95% CI 0.5-1.7). Dietary intakes of minerals did not change, apart from sodium which decreased from a mean intake of 2218.8 mg/day (SD 631.6) to 1895.3 mg/day (SD 581.0) with a medium effect size (d=0.5, 95% CI 0.1-1.1). Mean serum phosphorus, potassium, and albumin levels did not change relevantly. Mean serum iron increased from 7.9 mg/dL (SD 2.8) to 11.5 mg/dL (SD 7.9) postintervention with a medium effect size (d=0.6, 95% CI 0.0-1.2).

Conclusions: This pilot study showed that the KELA.AE app has the potential to improve dietary intakes. Processes related to procedure, resources, tools, and app improvement for a future trial were assessed. A more extended intervention using a randomized controlled trial is required to estimate parameters concerning app efficacy accurately.

(J Med Internet Res 2020;22(7):e17817) doi:10.2196/17817

XSL•FO
RenderX

KEYWORDS

mHealth; dietary app; hemodialysis diet

Introduction

Mobile technology has the potential to improve health care coverage, especially in low- to middle-income countries where people may be more likely to have access to a smartphone than to basic needs such as safe water and electricity [1]. Self-tracking and wearable technologies have become popular, particularly in the areas of diet and fitness [2]. Previous research on mobile health (mHealth) interventions provides evidence regarding their effectiveness [3-5]. However, their role as educational tools or as supportive tools to standard care remains inconclusive [3,6].

Mobile technology may assist in the introduction of new methods of dietary assessment and self-monitoring, allowing for real-time reporting of food intakes [7]. Self-monitoring is an important factor in successful dietary behavior changes [8]. In a previous study on a weight loss intervention, the more participants recorded food intakes on a mobile app, the more likely they were to lose weight [9]. However, patients may not always be interested and willing to track intakes [10], and commitment to self-monitoring decreases over time, even with the use of mobile technology [11].

Changing eating behaviors is difficult [12], and adherence to guidelines is challenging, especially for patients with chronic conditions such as chronic kidney disease (CKD) [13]. In particular, patients undergoing hemodialysis struggle to meet the target intakes set by dietary guidelines [14]. The diet during hemodialysis is somewhat restrictive, requiring the management of potassium, phosphorus, sodium, and fluids while maintaining adequate protein and energy intakes to prevent malnutrition [15]. In a recent study assessing adherence to diet guidelines among patients undergoing hemodialysis, 77% and 50% of the patients, respectively, consumed less energy and proteins than recommended. These low intakes could be attributed to the restrictive nature of the dialysis diet. In the same study, participants were also found to consume excessive saturated fats and inadequate intakes of fibers and micronutrients [14]. Maintaining adequate intake of minerals such as potassium, phosphorus, and calcium is essential in the prevention of dialysis-related complications such as heart failure, metabolic bone disorders, and mortality [16,17]. Protein-energy malnutrition is also a contributor to complications, and malnutrition among patients undergoing dialysis is a predictor of mortality [18]. Accordingly, it seems that the diet quality of these patients is often poor and that they do not follow dialysis-specific guidelines [14,19].

The main challenge in changing dietary behavior is developing interventions that are comprehensive and sustainable, promoting long-term changes in eating habits and lifestyle [12]. There is a need for nephrologists and dietitians to adopt approaches that strengthen educational and clinical interventions [14]. Possibly, mobile technology could be used to enhance dietetic practice by providing support to patients and dietitians. Thus, adding mobile apps to in-person counseling may provide more accessible and flexible dietetic services at lower costs [20]. However, research on the role of mHealth in dietary behavior in patients undergoing hemodialysis is still scarce [21].

This is a pilot study that aims at estimating the potential of a dietary intervention using a smartphone app in patients undergoing hemodialysis. The results from this pilot study will be used to improve the app itself as well as the study design of the planned randomized controlled trial.

Methods

Participants

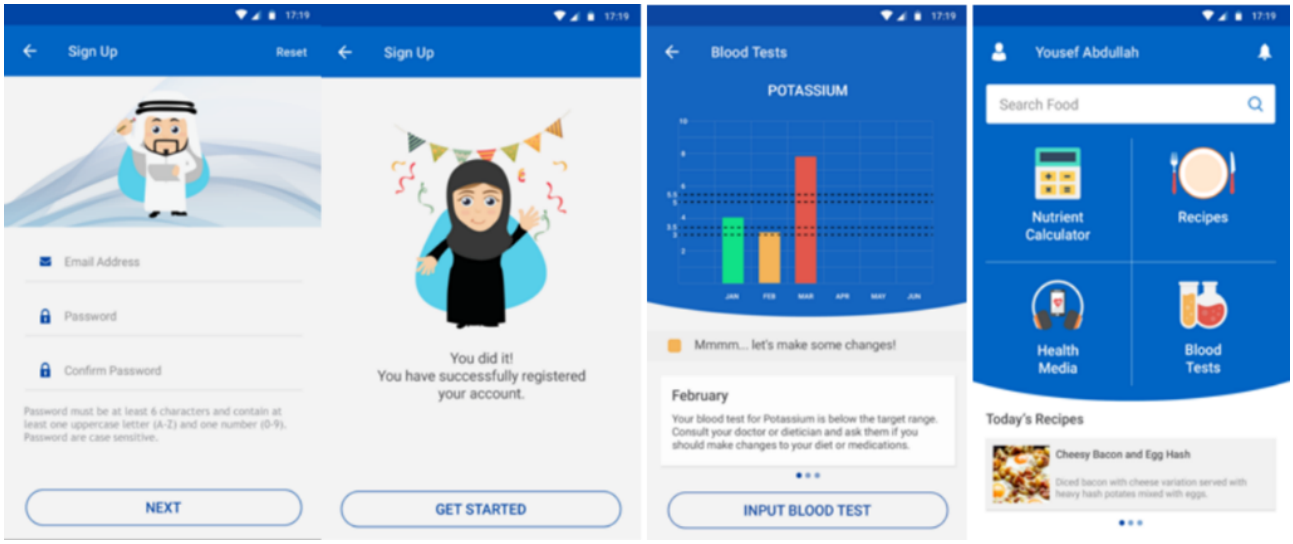
All patients at the hemodialysis unit of Al Qassimi Hospital (The Emirate of Sharjah) who fulfilled the study inclusion criteria were considered eligible to be enrolled in the pilot study. Inclusion criteria were patients undergoing hemodialysis for at least three months; free of life-threatening conditions; able to read, write, listen, and communicate; owning an Android smartphone; and not having been hospitalized in the past 6 months. All patients approached who accepted to enroll in the study signed written consent forms. A total of 26 participants were enrolled in the pilot, of which 23 downloaded and used the app for 2 weeks. Data were collected between February 2019 and April 2019. Post-hoc calculations of sample size in pilot studies, assuming detection of unanticipated problems with a probability of at least 15% (π=0.15) and a 95% confidence level, resulted in a required sample of 19 patients [22].

KELA.AE App

KELA.AE (Kidney Education for Lifestyle Application) is an Arabic, culturally specific, educational, and self-monitoring app that was developed in a person-centered, theory-based approach. Educational materials were developed based on the transtheoretical model [23] and constructs of the reasoned action approach [24]. Included educational materials are podcasts, videos, notifications, and recipes. The app also provides self-monitoring features that allow patients to track food intakes and blood parameters. The app was developed following the IDEAS (*i*ntegrate, *de*sign, *a*ssess, and share) framework [25]; a detailed description of the app development is described elsewhere [10]. Figure 1 depicts screenshots from the app.

XSL•FO
RenderX

Figure 1. Screenshots of the KELA.AE app used during the pilot study.



Procedure

This pilot is reported as per guidelines for nonrandomized pilot and feasibility studies [26] and in line with the CONSORT 2010 checklist for reporting a pilot or feasibility trial (items pertinent to randomization were considered not applicable) [27]. Ethical approval was received from the Institutional Review Board of Zayed University, Dubai (Ethical Approval number ZU17_066_F).

Participants who consented to enroll in the pilot study downloaded the app and were provided with usernames that allowed them to sign-in to the app. Upon successful installation of the app, users met with a dietitian once a week (two times in total); the dietitians responded to concerns pertinent to app usage, as well as questions about the renal diet by reinforcing the critical messages relayed by the educational materials. The dietitians also collected the data from patients before and after app usage. All outcomes were measured before (T0) and after 2 weeks (T1) of app usage.

Dietary Intakes

Face-to-face 24-hour recalls were collected [28] at T0 and T1 for all participants; nutrient compositions were derived for energy, carbohydrates, proteins, high biological value (%HBV) proteins, total fat, potassium, phosphorus, and sodium. Intakes were then compared with dietary guidelines for patients undergoing hemodialysis [15,29]. Protein targets were considered as 1.2 g/kg or more with 50% or higher HBV protein; energy as 30-35 kcal/kg; phosphorus as 1000 mg/day for participants with serum phosphorus below 5.5 mg/day and 12 mg/g of protein intake for participants with serum phosphorus below 5.5 mg/dL. Sodium and potassium targets were considered less than 2400 mg/day [15,29]. Standard body weight from the National Health and Nutrition Examination Study was used for calculations; however, adjusted edema-free body weight was used for calculating nutrient needs for individuals with less than 95% or greater than 115% of standard body weight [29]. Textbox 1 illustrates the dietary guidelines that were used as targets and that are also included in the adherence index.

Textbox 1. Dietary guidelines for patients undergoing hemodialysis used for adherence index [15,29]. HBV: high biological value, PTH: parathyroid hormone, aBWef: adjusted edema-free body weight, SBW: standard body weight, BWef: edema-free body weight.

1. Protein≥1.2 g/kg of body weight, ≥50% HBV protein
2. Energy for those <60 years of age: kilogram (kg) of body weight × 35 kcal; >60 years of age: kilogram (kg) of bodyweight × 30 kcal/kg to 35 kcal/kg
3. Sodium less than 2.4 g/day
4. Potassium less than 2.4 g/day
5. Phosphorus 800 mg/day to 1000 mg/day or 10-12 mg phosphorus/g of protein when serum phosphorus >5.5 mg/dL or intact PTH is elevated
6. aBW_{ef} is recommended for calculating nutrient needs for individuals with <95% or >115% of SBW using the formula aBW_{ef} (kg) = BW_{ef}+ [(SBW – BW_{ef}) × 0.25]. SBW from the National Health and Nutrition Examination Study is used otherwise

Biochemical Parameters

Serum phosphorus, potassium, and iron were retrieved from the patients' medical records. These biochemical parameters were measured as part of the routine tests performed in the hemodialysis unit. All tests were conducted postdialysis (in sessions). Comparative standards for serum potassium were considered between 3.5 and 5.5 mEq/L; and between 3.5 and 5.5 mg/dL for phosphorus [15,29].

Statistical Analysis

The Shapiro–Wilk normality test was performed to ensure that data are normally distributed. Cohen *d* effect sizes and confidence intervals (95%) were derived from means and pooled

XSL•FO
RenderX

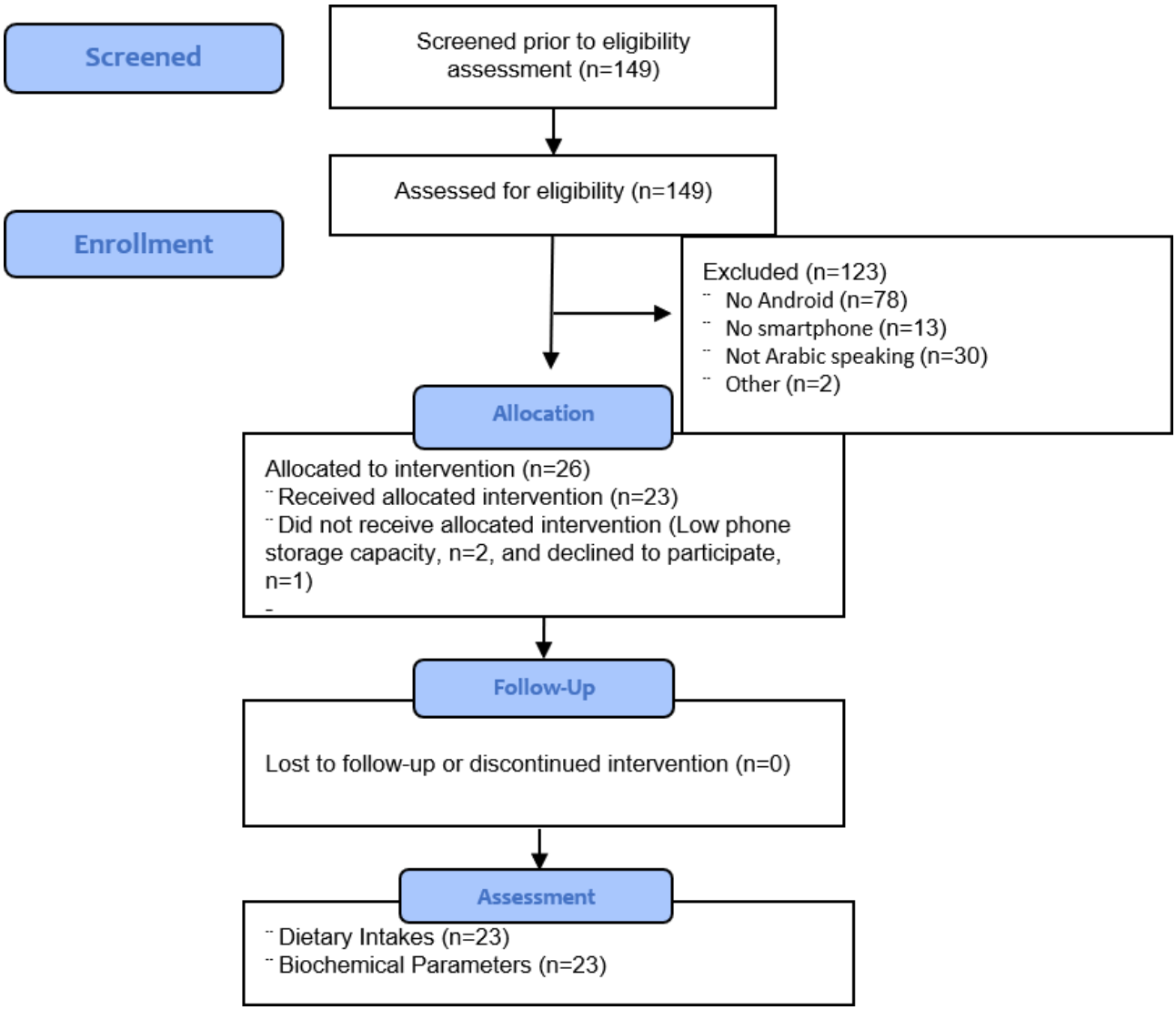
standard deviations. The effect size was considered small at 0.2, medium at 0.5, and large at 0.8 and above [30]. Effect sizes were calculated to understand the magnitude of the reported effects along with the probability by means of *P* values [31]. Paired *t* tests were performed to compare the mean scores before and after the intervention. Two-tailed *P* values are reported. Frequencies and percentages were used to describe categorical variables, whereas means and standard deviations were used for continuous variables. Statistical software IBM SPSS Statistics 21 Data Editor was used to perform all statistical analyses.

Results

Participant Eligibility and Baseline Characteristics

Of the 149 patients at the hemodialysis unit, 26 were eligible, of which 23 downloaded the app and completed the study. Participants that were not eligible were mainly those who did not own a smartphone or who owned a smartphone with an iOS operating system; 2 participants did not download the app due to limitations in phone storage, and 1 was not interested in downloading the app. Figure 2 depicts the CONSORT flow diagram. The sample's mean age (years) was 48.5 (SD 13.7), and mean BMI (kg/m^2) was 31.9 (SD 7.9); participants had been on dialysis for a mean of 29.7 (SD 37.3) months. More than half of the participants were males (n=14), and more than half suffered from hypertension or diabetes or both. Demographic data are detailed in Table 1.

Figure 2. CONSORT study flow diagram.



XSL•FO
RenderX

Table 1. Demographic and baseline characteristics of the sample (N=23).

Characteristics	Value
Age, mean (SD)	48.5 (13.7)
BMI, mean (SD)	31.9 (7.9)
Months on dialysis, mean (SD)	29.7 (37.3)
Gender, n (%)	
Male	14 (61)
Smokers, n (%)	6 (26)
Comorbidities, n (%)	
Hypertension	16 (70)
Diabetes	11 (48)
Dyslipidemia	2 (9)
Cancer	1 (4)
Liver disease	1 (4)

Anthropometry and Dietary Intakes

There was no change in weight and BMI postintervention; however, mean dietary intakes changed mainly for energy and macronutrients. Mean energy intakes increased from 24.4 kcal/kg/day (SD 8.0) to 29.1 kcal/kg/day (SD 7.8) with a medium effect size (d=0.6, 95% CI 0.0-1.2). Mean protein intakes increased from 0.9 g/kg/day (SD 0.3) to 1.3 g/kg/day (SD 0.5) with a large effect size (d=1.0, 95% CI 0.4-1.6); mean intake of HBV protein (%) also increased from 58.6% (SD 10.1) to 70.1% (SD 10.7) with a large effect size (d=1.1, 95% CI 0.5-1.7). Total fat intakes increased from baseline with medium effect size (d=0.5, 95% CI 0.0-1.1). Dietary intakes of minerals did not change, apart from sodium which decreased from a mean intake of 2218.8 mg/day (SD 631.6) to 1895.3 mg/day (SD 581.0) with a medium effect size (d=0.5, 95% CI 0.1-1.1).

Biochemical Parameters

Serum phosphorus, potassium, and albumin did not change relevantly. Mean serum iron increased from 7.9 mg/dL (SD 2.8) to 11.5 mg/dL (SD 7.9) postintervention with a medium effect size (d=0.6, 95% CI 0.0-1.2). Table 2 details the results of the anthropometry, dietary intakes, and blood parameters.

XSL•FO
RenderX

Table 2. Dietary intakes and laboratory data at baseline and postintervention.

Dietary intakes and laboratory data	Baseline (T0), mean (SD)	Postintervention (T1), mean (SD)	Cohen *d* (95% CI)	*P* value
Weight (kg)	85.5 (23.1)	84.1 (24.1)	0.1 (–0.5 to 0.6)	.37
BMI (kg/m^2)	31.9 (7.9)	31.3 (8.5)	0.1 (–0.5 to 0.6)	.39
Energy intake (kcal/day)	1918.3 (570.4)	2206.2 (378.2)	0.6 (0.0 to 1.2)	.003*
Energy intake (kcal/kg/day)	24.4 (8.0)	29.1 (7.8)	0.6 (0.0 to 1.2)	.002*
Dietary protein (g/day)	71.1 (26.4)	103.8 (37.8)	1.0 (0.4 to 1.6)	<.001*
Dietary protein (g/kg/day)	0.9 (0.3)	1.3 (0.5)	1.1 (0.4 to 1.7)	<.001*
HBV protein[a] (%)	58.6 (10.1)	70.1 (10.7)	1.1 (0.5 to 1.7)	<.001*
Total dietary CHO[b] (g/day)	224.7 (88.0)	215.2 (40.7)	0.1 (–0.4 to 0.7)	.59
Total dietary fat (g/day)	87.3 (30.1)	103.0 (26.8)	0.5 (0.0 to 1.1)	.02*
Dietary potassium (mg/day)	1831.2 (728.4)	2046.1 (555.1)	0.3 (0.2 to 0.9)	.19
Dietary phosphorus (mg/day)	1152.5 (489.7)	1343.1 (398.0)	0.4 (0.2 to 1.0)	.09
Dietary sodium (mg/day)	2218.8 (631.6)	1895.3 (581.0)	0.5 (0.1 to 1.1)	.03*
Serum potassium (mg/dL)	4.7 (0.7)	4.7 (0.7)	0.0 (–0.6 to 0.6)	.92
Serum phosphorus (mg/dL)	5.2 (1.5)	5.5 (2.2)	0.15 (–0.4 to 0.7)	.60
Serum iron (mg/dL)	7.9 (2.8)	11.5 (7.9)	0.6 (0.0 to 1.2)	.03*
Serum aluminum (g/dL)	3.0 (0.4)	3.2 (0.8)	0.3 (–0.3 to 0.9)	.37

*P<.05.

[a]HBV: high biological value.

[b]CHO: carbohydrates.

Adherence to Dietary Guidelines in Hemodialysis

Adherence to dietary guidelines in hemodialysis improved for energy intakes, protein intakes, and %HBV proteins with medium to large sizes, respectively (d=0.4, 95% CI 0.2-1.0 for energy; d=0.9, CI 0.3, 1.5 for proteins; d=1.1, 95% CI 0.5-1.7 for %HBV proteins). Adherence to fat, potassium, and phosphorus intakes did not change, whereas adherence to sodium further dropped to achieve intakes below 2400 mg/day with a medium effect size (d=0.5, CI 0.1-1.1). The number of patients adhering to dietary guidelines increased for energy, proteins, and sodium, whereas a larger number of patients became nonadherent to fat and phosphorus intakes. There was no relevant change in the number of patients concerning adherence to potassium intakes. Table 3 describes the difference between the recommended intakes and dietary intakes before and after the intervention. Figure 3 details the results of percent compliance to dietary recommendations.

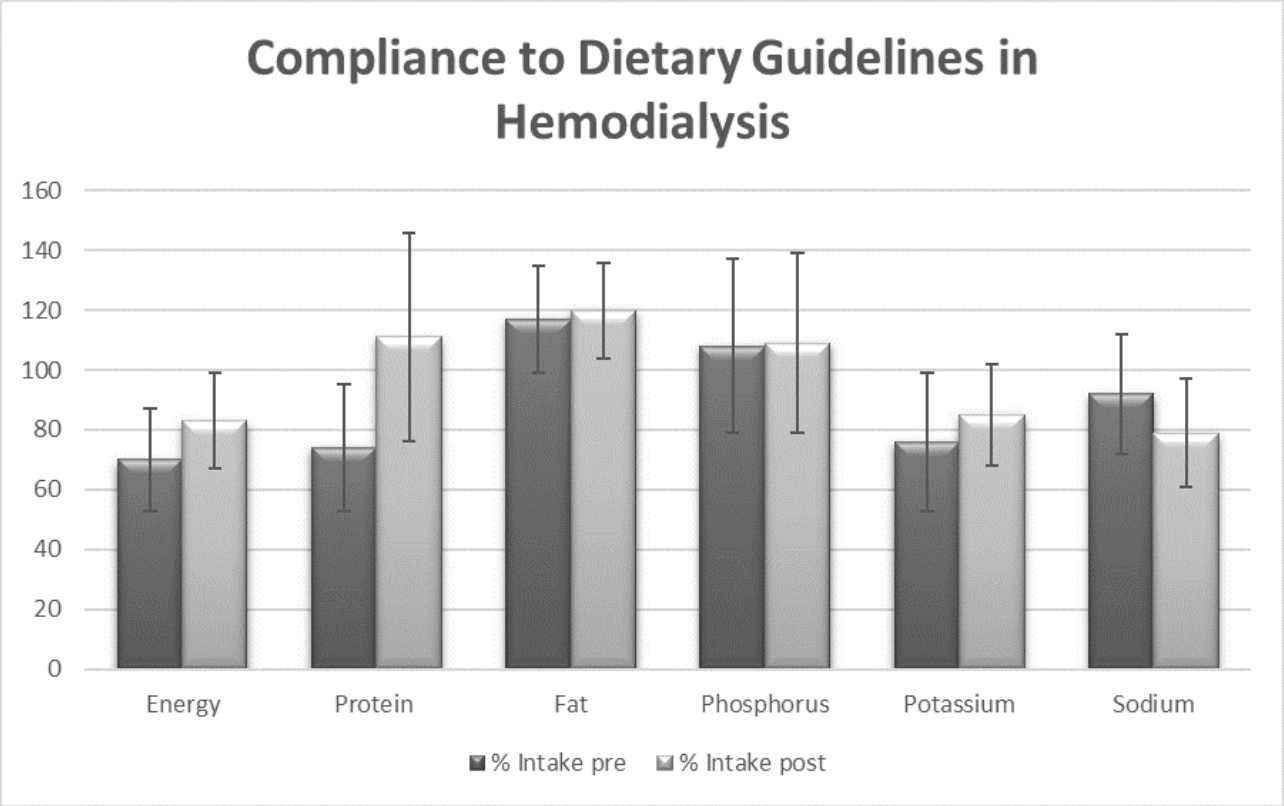
Table 3. Adherence to dietary intakes reported as the difference between intake and recommendation at baseline and postintervention.

Dietary intakes	Baseline, mean (SD)	Postintervention, mean (SD)	Cohen *d* (95% CI)	*P* value
Energy (kcal/day)	–903.1 (705.4)	–581.3 (779.8)	0.4 (0.2 to 1.0)	.004*
Dietary protein (g/day)	–27.3 (26.6)	+6.5 (43.9)	0.9 (0.3 to 1.5)	<.001*
HBV proteins[a] (%)	58.6 (10.1)	70.1 (10.7)	1.1 (0.5 to 1.7)	<.001*
Total dietary fat (% energy)	41.0 (7.6)	41.9 (7.3)	0.1 (–0.7 to 0.4)	.66
Dietary potassium (mg/day)	–568.1 (728.4)	–353.9 (555.1)	0.3 (–0.9 to 0.2)	.19
Dietary phosphorus (mg/day)	+84.7 (445.8)	250.3 (369.9)	0.4 (–0.9 to 0.2)	.14
Dietary sodium (mg/day)	–181.1 (631.6)	–504.6 (581.0)	0.5 (0.1 to 1.1)	.03*

[a]HBV: high biological value.

XSL•FO
RenderX

Figure 3. Percent (%) adherence to dietary intakes pre- and post-intervention. Values are presented as mean (SD).



User Acceptability and App Usability

A user acceptability tool was used during this pilot; however, all participants responded to all questions with a *strongly agree* option based on a Likert scale. Accordingly, the results were not considered useful to understand better the acceptability and, therefore, have not been reported. Assessment of acceptability will be reassessed in the future trial. The validated end-user version of the Mobile App Rating Scale (uMARS) [32] will be used in the trial. Participants may not have well understood that the questionnaire aims to improve the app; thus, qualitative data will also be collected to ask specific questions about the KELA.AE usability and acceptability by referring to each feature in the app individually. Qualitative data will also address the culturally specific features of the app such as language and recipes.

Field notes revealed that all 23 participants attended the weekly sessions with the dietitians. Usage data available from analytics were only able to provide the last date of access for each participant, and all 23 had accessed the app in the last week of the trial. No data on usage frequency could be retrieved due to a lack of in-app analytics.

Discussion

Principal Results

Potential Efficacy

In this study, short-term app usage had a potential impact on energy and protein intakes among patients undergoing hemodialysis. In comparison to the target dietary guidelines for hemodialysis, patients started with a baseline intake of energy and proteins below recommendations. As an outcome of the intervention, mean energy intakes reached 29.1 kcal/kg/day (SD 7.8) as compared with the recommended 30-35 kcal/kg/day. In addition, protein intakes were also considerably low as compared with guidelines (mean consumption at baseline of 0.9 g/kg/day [SD 0.3]) and improved to become in line with the recommended 1.2 g/kg or more of body weight and 50% or higher HBV protein [15,29].

No changes were observed in intakes of minerals; nonetheless, baseline intakes of potassium and sodium were already within the recommendations of less than 2400 mg/day. Mean sodium intakes further dropped to 1895.3 mg/day (SD 581.0). Fat and phosphorus intakes, however, were already elevated at baselines and remained above recommendations after the intervention.

No changes were observed in the serum laboratory parameters for potassium and phosphorus; levels remained within recommendation at baseline and after the intervention. The duration of the intervention is not long enough to detect relevant changes in serum values; accordingly, the upcoming trial is expected to capture the effect of app usage on laboratory parameters better. Mean serum iron increased with a medium effect size from 7.9 mg/dL (SD 2.8) to 11.5 mg/dL postintervention. The latter may be explained by the increase in the dietary protein of HBV, which mainly consists of animal protein sources.

Study and App Feasibility

Research methodology, resources, and tools used in this pilot were assessed to refine and modify the planned trial and improve the app. The feasibility of the processes related to recruitment, retention, and refusal seemed adequate; there were no patients

XSL•FO
RenderX

that withdrew from the study; and all eligible patients agreed to participate. However, the number of eligible patients were about 17.4% (26/149) out of all patients undergoing dialysis at the unit. It would mean that to achieve an appropriate sample size for the planned randomized trial, many dialysis units will need to be recruited. Thus, the number of dietitians needed to be involved in the future study will need adequate resources to meet patients weekly, as it was the case in this study. Accordingly, an assessment of resource allocation should be performed to consider app improvement and development on iOS mobile operating system to increase eligible patients within the same dialysis unit.

Digital interventions may be used as tools to support a reciprocal relationship between patients and health care practitioners and enhance patient-centered care [33]. Accordingly, dietitians would be expected to manage and recommend the usage of KELA.AE app in real practice settings. Thus, the future trial will include dietitians following up with patients to simulate app integration in real practice.

In-app analytics are also essential for the future trial to assess the effect of app usage on outcomes. In-app analytics should track usage frequency as well as the usage of each feature separately. The results would help to better understand if educational or self-monitoring features of the app are preferred by the user.

The future trial should further analyze the fatty acid profile (saturated, monounsaturated, and polyunsaturated) on top of total fats to assess diet quality from an atherogenic point of view. In addition, further assessment of the nutritional status of patients undergoing dialysis should be performed to identify patients with malnourishment that may need additional support from the research dietitians.

The KELA.AE app will be re-evaluated to include educational materials addressing diet quality pertinent to dietary fatty acid profiles to promote dietary intakes in line with the KDOQI Clinical Practice Guidelines for the Management of Dyslipidemias in Patients With Kidney Disease [34]. This is an update to be considered in the next version of the app should dietary fat intakes not be in line with guidelines.

This pilot study has demonstrated that it is feasible to integrate a dietary app into dietetic practice, allowing it to be a tool in addition to regular meetings with dietitians. The study has also identified how to adjust the design, procedures, data collection tools, and outcome measurements for a future trial. Based on the information collected, we feel ready to proceed to a larger randomized controlled trial.

Comparison With Prior Work

Energy and protein recommendations in hemodialysis are higher than the recommendations for healthy individuals. Whereas in the general population, intakes of 0.8 g/kg are adequate to maintain nitrogen balance [35], such intakes are inadequate in hemodialysis [36]. Hemodialysis sessions cause losses in proteins and amino acids that need to be replenished by the diet [36]. Protein-energy wasting is dangerous and has been correlated with mortality, adverse clinical outcomes [37], and poor quality of life [18]. Both Inadequate dietary proteins and energy intakes are criteria used in the diagnosis of protein-energy wasting in patients undergoing hemodialysis [38]. Similar to the findings in this pilot, many studies have identified inadequate energy and protein intakes among patients undergoing dialysis [14,39,40]. This pilot intervention seems to show potential in the improvement of energy and protein intakes among this population.

However, multiple dietary components need to be adjusted to improve the clinical outcomes of these patients, and the latter includes management of potassium and phosphorus, along with a diet quality that is cardioprotective [15]. In the sample studied in this pilot, both serum potassium levels and potassium intakes were within targets before and after the intervention. This was also observed by others [14,19], whereby similar to our findings, patients undergoing hemodialysis tend to be adherent to potassium intakes and laboratory targets. However, serum phosphorus and phosphorus intakes were borderline high both before and after the intervention. This is also a common finding in this population, in which the discrepancies between the phosphorus and protein recommendations [15] make it challenging to achieve the required targets. Accordingly, it seems that further assessment of the factors that may influence phosphorus management, such as food sources of phosphorus and compliance to phosphate binders, should be explored. In addition, it is expected that a longer intervention that includes education and follow-up with dietitians may lead to better outcomes.

Total fat intakes were also found to be elevated by our pilot study, and their intakes remained elevated after the intervention as well. Others have previously explored the diet quality of patients undergoing hemodialysis as compared with the recommendations of the American Heart Association and KDOQI Clinical Practice Guidelines for Management of Dyslipidemias in Patients With Kidney Disease [34], and found the current diet intakes to be proatherogenic in nature [14]. Given that protein intakes increased and that their sources are mainly from animal proteins (an increase in %HBV proteins were also observed), it can be predicted that saturated fat intakes also increased as a result of these changes.

There are only few available apps that target patients with CKD specifically; however, most of them are available on app stores, and information on their efficacy and usability are rather scarce [41]. A recent content analysis of mobile apps for CKD revealed that available apps fail to provide the continuity of patient-centered care that is needed to support patients with CKD [42].

Limitations

Given the pilot nature of the study, some limitations may lead to a bias in the interpretation of the results and their generalizability. The duration of this study was short; therefore, its effects are only focused on dietary intakes and laboratory parameters. Behavioral interventions need to be of longer duration and should be comprehensive to change dietary intakes [43]. The results of this pilot will instead be used to improve the app and modify the study procedures in the future trial.

XSL•FO
RenderX

In addition, the results identified a possible improvement in nutritional status and a potentially proatherogenic diet quality among the study's sample. Thus, the future trial should assess the prevalence of patients with malnourishment along with the fatty acid profile of the diet in line with the KDOQI Clinical Practice Guidelines for Management of Dyslipidemias in Patients With Kidney Disease [34].

The intervention included face-to-face sessions with dietitians that may have influenced the outcomes. The reinforcement of the dietitians and their availability may have impacted the results similarly or more than the app itself. Accordingly, the future trial should include a control group whereby dietitians see patients with the same frequency but without app usage.

Given that app analytics were not available, we were unable to track how many times the users accessed the app. The only available data are the last access for each user, and all users had accessed the app during the last week.

Conclusions

This pilot study showed that KELA.AE app has the potential to improve dietary intakes. Processes related to the procedures, resources, tools, and app improvement for a future trial were assessed. A more extended intervention using a randomized controlled trial is required to estimate parameters concerning app efficacy accurately.

Acknowledgments

We acknowledge the staff and students from Zayed University who volunteered by contributing to the app development at many steps. A research grant by Zayed University, Dubai, funded the study (Grant No. R18060).

Conflicts of Interest

None declared.

References

1. Executive Board, 142 (World Health Organization). mHealth: use of appropriate digital technologies for public health: report by the Director-General, World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2020-06-25]
2. Zaidan S, Roehrer E. Popular Mobile Phone Apps for Diet and Weight Loss: A Content Analysis. JMIR Mhealth Uhealth 2016 Jul 11;4(3):e80. [doi: 10.2196/mhealth.5406]
3. Fakih El Khoury C, Karavetian M, Halfens RJG, Crutzen R, Khoja L, Schols JMGA. The Effects of Dietary Mobile Apps on Nutritional Outcomes in Adults with Chronic Diseases: A Systematic Review and Meta-Analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651. [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
4. Helle C, Hillesund ER, Wills AK, Øverby NC. Evaluation of an eHealth intervention aiming to promote healthy food habits from infancy -the Norwegian randomized controlled trial Early Food for Future Health. Int J Behav Nutr Phys Act 2019 Jan 03;16(1):1 [FREE Full text] [doi: 10.1186/s12966-018-0763-4] [Medline: 30606197]
5. Järvelä-Reijonen E, Karhunen L, Sairanen E, Muotka J, Lindroos S, Laitinen J, et al. The effects of acceptance and commitment therapy on eating behavior and diet delivered through face-to-face contact and a mobile app: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Feb 27;15(1):22 [FREE Full text] [doi: 10.1186/s12966-018-0654-8] [Medline: 29482636]
6. Gonzalez-Sanchez J, Recio-Rodriguez JI, Fernandez-delRio A, Sanchez-Perez A, Magdalena-Belio JF, Gomez-Marcos MA, On behalf the EVIDENT Investigators group. Using a smartphone app in changing cardiovascular risk factors: A randomized controlled trial (EVIDENT II study). Int J Med Inform 2019 May;125:13-21. [doi: 10.1016/j.ijmedinf.2019.02.004] [Medline: 30914176]
7. Lemacks JL, Adams K, Lovetere A. Dietary Intake Reporting Accuracy of the Bridge2U Mobile Application Food Log Compared to Control Meal and Dietary Recall Methods. Nutrients 2019 Jan 19;11(1):199 [FREE Full text] [doi: 10.3390/nu11010199] [Medline: 30669430]
8. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
9. Kerr DA, Dhaliwal SS, Pollard CM, Norman R, Wright JL, Harray AJ, et al. BMI is Associated with the Willingness to Record Diet with a Mobile Food Record among Adults Participating in Dietary Interventions. Nutrients 2017 Mar 07;9(3):244 [FREE Full text] [doi: 10.3390/nu9030244] [Medline: 28272343]
10. Fakih El Khoury C, Karavetian M, Halfens RJG, Crutzen R, El Chaar D, Schols JMGA. Dietary Application for the Management of Patients with Hemodialysis: A Formative Development Study. Healthc Inform Res 2019 Oct;25(4):262-273 [FREE Full text] [doi: 10.4258/hir.2019.25.4.262] [Medline: 31777669]
11. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining Adherence to Mobile Dietary Self-Monitoring and Assessing Tracking Over Time: Tracking at Least Two Eating Occasions per Day Is Best Marker of

XSL•FO
RenderX

Adherence within Two Different Mobile Health Randomized Weight Loss Interventions. J Acad Nutr Diet 2019 Sep;119(9):1516-1524. [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
12. Kelly MP, Barker M. Why is changing health-related behaviour so difficult? Public Health 2016 Jul;136:109-116 [FREE Full text] [doi: 10.1016/j.puhe.2016.03.030] [Medline: 27184821]
13. Hayashi A, Yamaguchi S, Waki K, Fujiu K, Hanafusa N, Nishi T, et al. Testing the Feasibility and Usability of a Novel Smartphone-Based Self-Management Support System for Dialysis Patients: A Pilot Study. JMIR Res Protoc 2017 Apr 20;6(4):e63 [FREE Full text] [doi: 10.2196/resprot.7105] [Medline: 28428168]
14. Luis D, Zlatkis K, Comenge B, García Z, Navarro JF, Lorenzo V, et al. Dietary Quality and Adherence to Dietary Recommendations in Patients Undergoing Hemodialysis. J Ren Nutr 2016 May;26(3):190-195. [doi: 10.1053/j.jrn.2015.11.004] [Medline: 26827131]
15. Kopple JD. National kidney foundation K/DOQI clinical practice guidelines for nutrition in chronic renal failure. Am J Kidney Dis 2001 Jan;37(1 Suppl 2):S66-S70. [doi: 10.1053/ajkd.2001.20748] [Medline: 11158865]
16. Tentori F, Blayney MJ, Albert JM, Gillespie BW, Kerr PG, Bommer J, et al. Mortality risk for dialysis patients with different levels of serum calcium, phosphorus, and PTH: the Dialysis Outcomes and Practice Patterns Study (DOPPS). Am J Kidney Dis 2008 Sep;52(3):519-530. [doi: 10.1053/j.ajkd.2008.03.020] [Medline: 18514987]
17. Kovesdy CP, Regidor DL, Mehrotra R, Jing J, McAllister CJ, Greenland S, et al. Serum and dialysate potassium concentrations and survival in hemodialysis patients. Clin J Am Soc Nephrol 2007 Sep;2(5):999-1007 [FREE Full text] [doi: 10.2215/CJN.04451206] [Medline: 17702709]
18. Borges MCC, Vogt BP, Martin LC, Caramori JCT. Malnutrition Inflammation Score cut-off predicting mortality in maintenance hemodialysis patients. Clin Nutr ESPEN 2017 Feb;17:63-67. [doi: 10.1016/j.clnesp.2016.10.006] [Medline: 28361749]
19. Kim H, Lim H, Choue R. A Better Diet Quality is Attributable to Adequate Energy Intake in Hemodialysis Patients. Clin Nutr Res 2015 Jan;4(1):46-55 [FREE Full text] [doi: 10.7762/cnr.2015.4.1.46] [Medline: 25713792]
20. Rollo ME, Burrows T, Vincze LJ, Harvey J, Collins CE, Hutchesson MJ. Cost evaluation of providing evidence-based dietetic services for weight management in adults: In-person versus eHealth delivery. Nutr Diet 2018 Feb;75(1):35-43. [doi: 10.1111/1747-0080.12335] [Medline: 29411491]
21. Ni Z, Jin H, Jiang G, Wang N, Peng A, Guo Z, et al. A Telemedicine-Based Registration System for the Management of Renal Anemia in Patients on Maintenance Hemodialysis: Multicenter Study. J Med Internet Res 2019 May 08;21(5):e13168 [FREE Full text] [doi: 10.2196/13168] [Medline: 31344676]
22. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015 Nov;68(11):1375-1379. [doi: 10.1016/j.jclinepi.2015.04.014] [Medline: 26146089]
23. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [Medline: 10170434]
24. Fishbein M. A reasoned action approach to health promotion. Med Decis Making 2008;28(6):834-844 [FREE Full text] [doi: 10.1177/0272989X08326092] [Medline: 19015289]
25. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A Framework and Toolkit of Strategies for the Development of More Effective Digital Interventions to Change Health Behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
26. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5:114 [FREE Full text] [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
27. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
28. Van Horn L, Beto JA. Successful Approaches in Nutrition and Dietetics. Cleveland, OH: Academy of Nutrition and Dietetics; 2019.
29. Academy of Nutrition and Dietetics. Nutrition Care Manual. 2019. URL: https://www.nutritioncaremanual.org/ [accessed 2020-06-25]
30. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
31. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
32. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
33. Ammenwerth E. From eHealth to ePatient: The Role of Patient Portals in Fostering Patient Empowerment. ejbi 2018;14(2):1-4. [doi: 10.24105/ejbi.2018.14.2.4]
34. Kidney Disease Outcomes Quality Initiative (K/DOQI) Group. K/DOQI clinical practice guidelines for management of dyslipidemias in patients with kidney disease. Am J Kidney Dis 2003 Apr;41(4 Suppl 3):I-IV, S1. [Medline: 12671933]

XSL•FO
RenderX

35. Trumbo P, Schlicker S, Yates AA, Poos M, FoodNutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. J Am Diet Assoc 2002 Nov;102(11):1621-1630. [doi: 10.1016/s0002-8223(02)90346-9] [Medline: 12449285]
36. Sabatino A, Regolisti G, Karupaiah T, Sahathevan S, Sadu Singh BK, Khor BH, et al. Protein-energy wasting and nutritional supplementation in patients with end-stage renal disease on hemodialysis. Clin Nutr 2017 Jun;36(3):663-671. [doi: 10.1016/j.clnu.2016.06.007] [Medline: 27371993]
37. Kovesdy CP. Malnutrition in Dialysis Patients--The Need for Intervention Despite Uncertain Benefits. Semin Dial 2016;29(1):28-34. [doi: 10.1111/sdi.12410] [Medline: 26190025]
38. Fouque D, Kalantar-Zadeh K, Kopple J, Cano N, Chauveau P, Cuppari L, et al. A proposed nomenclature and diagnostic criteria for protein-energy wasting in acute and chronic kidney disease. Kidney Int 2008 Feb;73(4):391-398 [FREE Full text] [doi: 10.1038/sj.ki.5002585] [Medline: 18094682]
39. Carvalho KT, Silva MIB, Bregman R. Nutritional profile of patients with chronic renal failure. J Ren Nutr 2004 Apr;14(2):97-100. [doi: 10.1053/j.jrn.2004.01.009] [Medline: 15060874]
40. Kopple JD, Berg R, Houser H, Steinman TI, Teschan P. Nutritional status of patients with different levels of chronic renal insufficiency. Modification of Diet in Renal Disease (MDRD) Study Group. Kidney Int Suppl 1989 Nov;27:S184-S194. [Medline: 2636655]
41. Lewis RA, Lunney M, Chong C, Tonelli M. Identifying Mobile Applications Aimed at Self-Management in People With Chronic Kidney Disease. Can J Kidney Health Dis 2019;6:2054358119834283 [FREE Full text] [doi: 10.1177/2054358119834283] [Medline: 30899533]
42. Lee Y, Cui Y, Tu M, Chen Y, Chang P. Mobile Health to Maintain Continuity of Patient-Centered Care for Chronic Kidney Disease: Content Analysis of Apps. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e10173 [FREE Full text] [doi: 10.2196/10173] [Medline: 29678805]
43. Nooriani N, Mohammadi V, Feizi A, Shahnazi H, Askari G, Ramezanzade E. The Effect of Nutritional Education Based on Health Belief Model on Nutritional Knowledge, Health Belief Model Constructs, and Dietary Intake in Hemodialysis Patients. Iran J Nurs Midwifery Res 2019;24(5):372-378 [FREE Full text] [doi: 10.4103/ijnmr.IJNMR_124_18] [Medline: 31516524]

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by C Gibson, E Ding; comments to author 14.03.20; revised version received 05.05.20; accepted 03.06.20; published 20.07.20.

Please cite as:
Fakih El Khoury C, Crutzen R, Schols JMGA, Halfens RJG, Karavetian M
A Dietary Mobile App for Patients Undergoing Hemodialysis: Prospective Pilot Study to Improve Dietary Intakes
J Med Internet Res 2020;22(7):e17817
URL: http://www.jmir.org/2020/7/e17817/
doi:10.2196/17817
PMID:32706698

©Cosette Fakih El Khoury, Rik Crutzen, Jos M G A Schols, Ruud J G Halfens, Mirey Karavetian. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Experiences With Wearable Activity Data During Self-Care by Chronic Heart Patients: Qualitative Study

Tariq Osman Andersen[1*], DPhil; Henriette Langstrup[2*], DPhil; Stine Lomborg[3*], DPhil

[1]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

[2]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[3]Department of Communication, University of Copenhagen, Copenhagen, Denmark

[*]all authors contributed equally

Corresponding Author:
Tariq Osman Andersen, DPhil
Department of Computer Science
University of Copenhagen
Universitetsparken 5
Copenhagen, 2100
Denmark
Phone: 45 26149169
Email: tariq@di.ku.dk

Abstract

Background: Most commercial activity trackers are developed as consumer devices and not as clinical devices. The aim is to monitor and motivate sport activities, healthy living, and similar wellness purposes, and the devices are not designed to support care management in a clinical context. There are great expectations for using wearable sensor devices in health care settings, and the separate realms of wellness tracking and disease self-monitoring are increasingly becoming blurred. However, patients' experiences with activity tracking technologies designed for use outside the clinical context have received little academic attention.

Objective: This study aimed to contribute to understanding how patients with a chronic disease experience activity data from consumer self-tracking devices related to self-care and their chronic illness. Our research question was: "How do patients with heart disease experience activity data in relation to self-care and chronic illness?"

Methods: We conducted a qualitative interview study with patients with chronic heart disease (n=27) who had an implanted cardioverter-defibrillator. Patients were invited to wear a FitBit Alta HR wearable activity tracker for 3-12 months and provide their perspectives on their experiences with step, sleep, and heart rate data. The average age was 57.2 years (25 men and 2 women), and patients used the tracker for 4-49 weeks (mean 26.1 weeks). Semistructured interviews (n=66) were conducted with patients 2–3 times and were analyzed iteratively in workshops using thematic analysis and abductive reasoning logic.

Results: Of the 27 patients, 18 related the heart rate, sleep, and step count data directly to their heart disease. Wearable activity trackers actualized patients' experiences across 3 dimensions with a spectrum of contrasting experiences: (1) knowing, which spanned gaining insight and evoking doubts; (2) feeling, which spanned being reassured and becoming anxious; and (3) evaluating, which spanned promoting improvements and exposing failure.

Conclusions: Patients' experiences could reside more on one end of the spectrum, could reside across all 3 dimensions, or could combine contrasting positions and even move across the spectrum over time. Activity data from wearable devices may be a resource for self-care; however, the data may simultaneously constrain and create uncertainty, fear, and anxiety. By showing how patients experience self-tracking data across dimensions of knowing, feeling, and evaluating, we point toward the richness and complexity of these data experiences in the context of chronic illness and self-care.

(J Med Internet Res 2020;22(7):e15873) doi:10.2196/15873

KEYWORDS

consumer health information; wearable electronic devices; self-care; chronic illness; patient experiences

XSL•FO
RenderX

Introduction

Consumer Wearable Activity Trackers in Chronic Care

Consumer health information technologies such as wearable activity trackers are increasingly being considered to improve chronic care management [1-4]. Contrary to traditional health information technologies, these devices are developed as consumer devices and not as clinical devices. Most commercial activity trackers aim to monitor and motivate sport activities, healthy living, and similar wellness purposes. Wristbands such as Fitbit and smart watches that track bodily signs (eg, heart rate) do not provide diagnostic services or disorder-specific information, and they are regulated less rigorously than are monitoring devices aimed at specific patient groups and clinical measures. This makes them readily available to consumers. Moreover, their designs aim at easy and noninvasive integration into everyday life, by way of automated tracking. These features are attractive for application in chronic disease management, where a healthy lifestyle can be a central part of treatment, rehabilitation, and prevention [1,5-8].

The separate realms of "wellness tracking" and "disease self-monitoring" and "activity data" and "medical data" are thus blurred, which is somewhat mirrored in an increasing prominence of concepts such as "patient-generated health data" and "personal health technology" where the focus is on the individual producer of data, rather than on the specific context or purpose of use. Applying leisure activity tracking to chronic care management provides new opportunities but is often based on assumptions about what characterizes these devices: easy, applicable, user-friendly, empowering, and motivating technology that can collect data with relevance for self-care and treatment. While there are great expectations and promising results emerging [2,3,9], little attention has been paid to the embodied and embedded experiences of self-tracking among patients with a chronic disease using consumer devices, which are not integrated in the health care system.

In this paper, we explored how patients who use consumer wearable devices make sense of them and of the data they produce and display vis-à-vis their embodied disorders. We recruited 27 patients with chronic heart disease who had an implantable cardioverter-defibrillator (ICD) to wear a Fitbit wearable tracker to understand the experiential qualities of how they relate self-tracking and activity data to their disease to explore the question: How do the embodied experiences and self-care practices of dealing with a specific health condition respond to the introduction of activity trackers?

Background and Significance: Experiences with Self-Monitoring in Health Care and Leisure Contexts

The existing literature of experiences with self-tracking comprises two related fields: (1) rich literature on how patients, as part of their prescribed treatment, engage in and make sense of clinically validated data related to self-care for chronic illnesses and (2) emerging literature that primarily explores users' experiences with leisurely oriented self-tracking technologies that are outside of the health care system. Patients' experiences with consumer activity-tracking technologies designed for use outside the clinical context have received scant academic attention. There is an important research gap in understanding the relation between the rich human-information interaction and the contexts of activity tracking such as self-care [1,10]. Consequently, studies that explore the experiences of people coping with illnesses — while recognizing the specific ramifications that self-tracking might have for those who have severe health problems — are necessary.

Self-Care and Chronic Illness in eHealth

Many patients routinely engage with data from digital devices that are part of the prescribed treatment. These clinically integrated data-producing devices affect self-care activities as they enable managing symptoms, taking medicine, dealing with the emotional impact, and tackling lifestyle changes [11-13]. Fostering self-care has been a central ambition of much telecare and electronic health (eHealth) development during the last 20 years. Early telecare technologies were typically designed from a clinical standpoint with measures to support remote decision making such as blood glucose tracking in diabetes [14,15], oxygen saturation, pulse rate and respiration rate tracking in chronic obstructive pulmonary disease [16], and heart arrhythmia detection through remote monitoring of cardiac implantable electronic devices such as pacemakers and ICDs [17]. In recent years, a more participatory agenda frames eHealth innovation, aiming to enhance independence and enable patients to become more active participants in managing their own disorder (eg, through use of wearable activity trackers) [18,19].

While much research has focused on measurable outcomes of digital self-tracking and self-management on clinical parameters, emergent studies have also explored the experiential qualities of patients' "data work" [20-22]. Central to such data work, along with the experiential qualities of patients engaging with self-tracking data in chronic care management, are the affective aspects. It is known that self-monitoring of blood glucose data by patients with diabetes and their caretakers is tightly bound to an emotional struggle ranging between control and freedom, peace of mind and anxiety, and empowerment and the burden of managing technology [23]. Similarly, it is found that self-tracking data in fertility self-monitoring promotes the achievement of certain positive goals but may accentuate negative emotions such as feeling burdened or abandoned [24]. Others have studied patient experiences in heart arrhythmia telemonitoring and found that not having access to data or feedback from clinicians has an emotional and life-changing impact, which in turn creates doubt, guilt, and concern [25].

Self-Tracking and Activity Data

While clinically validated self-monitoring technologies have become the standard in chronic care, the rise of low-cost sensors has accelerated the consumer market, making wearable activity trackers and mobile data-logging applications a widespread commodity. Large corporations like FitBit and Apple are entering the medical domain with consumer wearable devices, most prominent through automated activity measures like step count [1,26] to support disease monitoring and rehabilitation of cardiac, pulmonary, and cancer patients [8,27-29], among

XSL•FO
RenderX

others, and most recently, through large-scale interventions using the Apple Watch for screening of atrial fibrillation.

Literature on the so-called Quantified Self over the past decade has explored users' (ie, self-trackers) experiences with these technologies and the data they produce when applied voluntarily for leisure or wellness activities [30-34]. Lomborg and Frandsen [35] showed that, while self-tracking is often depicted as an entirely individual endeavor to retrieve calculable, reliable knowledge, it is experienced by users as a deeply communicative activity, with an often playful and pleasurable quality to it.

In their exploration of self-tracking cycling, Lupton et al [36] presented the concept of "data sense" to describe how people's experience with data from sensors is not just a matter of cognitive "knowing" — as often assumed in data literacy approaches — but equally involves sensory and affective dimensions such as alerting cyclists to new bodily sensations while possibly invoking feelings of frustration or even embarrassment. Thus, what emerges are accounts of self-tracking experiences as complex encounters between the metric and sensuous, between knowing and feeling. This, in turn, suggests that self-tracking, while often a purposeful and systematic practice, is not necessarily guided by goals of improving the self, forming new and healthier habits, or getting to know oneself better, as suggested by many of the wellness technologies currently entering the market [37].

The critical question is then: what happens when technologies, practices, and data from the consumer market for self-tracking are introduced in the context of chronic self-care? What kinds of experiences do they offer in a setting where self-tracking is "pushed" [38] to patient-consumers who live with a chronic disorder?

Several studies have examined patients' experiences with wearable activity trackers, and the focus tends to be on patient acceptability and feasibility among the elderly [39-41] and in the chronic care context [1,5-8,28]. Some support positive outcomes like ease-of-use and willingness among patients to wear activity trackers and integrating them into clinical care [5,28]. Rosenberg et al [28] conducted a 3-week study examining the acceptability of the Fitbit Zip and attitudes towards integrating fitness tracking into clinical care among men with prostate cancer. All participants were willing to wear the device and endorsed its value in ensuring they engaged in a "minimal amount of activity." However, several barriers to use were found, including health-related limitations (like pain and injuries making it difficult to walk) and practical or technical problems with syncing devices and experiencing data inaccuracies (eg, not capturing the activities).

Other studies present negative aspects and challenges with integrating patient-generated data in clinical settings [6,7]. Zhu et al [6] found technical challenges (such as security and privacy issues and the practical work of clinicians transferring self-tracking data to the electronic medical record), social challenges (such as health professionals adapting to new forms of care where patients are collaborative partners), and organizational challenges (like organizational policies and workflows that do not include attending to patient-self tracking data). Ancker et al [7] conducted an interview study to explore self-tracking among patients with multiple chronic conditions. They found that patients associated several negative experiences with self-tracking and self-tracking data can negatively influence the patient-clinician relationship owing to a lack of trust in the data. For these patients, self-tracking thus became burdensome, which contrasts the pleasurable and playful experiences promoted in wellness self-tracking.

Objective

The objective of this study was to understand how patients with chronic heart disease, as opposed to healthy individuals, experience activity data from consumer self-tracking devices when engaging in self-care. With this study, we contribute to the emergent literature on how patients' experiences with consumer wearable activity trackers are related to their illness and their self-care activities and the implications that arise for design and deployment of these devices. There is a need for going beyond acceptability and feasibility studies and conducting more fine-grained analysis of the experiential qualities of interacting with personal health data outside the context of clinical practice among the increasing number of people with chronic illnesses.

Methods

Overview

We conducted a qualitative study to understand how patients with an ICD experience self-tracking of activity data in relation to their embodied condition and daily practices of dealing with a chronic heart condition. As we know from the self-tracking literature, such experiences may change over time [33]. To grasp this, we followed patients over a 49-week period from January 2018 to December 2018, during which we observed their activity tracking data and interviewed them repeatedly (2-3 times each) about their experiences and insights gleaned from the data.

Setting

This study was part of a larger research and development project, SCAUT (Self-, Collaborative- and AUTo-detection of signs and symptoms of deterioration), 2014–2018, which aimed to improve early detection of deterioration and communication among patients with a cardiac device and health professionals. The overall project was carried out at a cardiac device clinic at the Rigshospital, University of Copenhagen, Denmark, which is one of the largest cardiac device remote monitoring centers in Europe, following more than 3500 patients.

Recruitment of Participants

The study comprised a sample of 27 patients with chronic heart disease who had a secondary prevention ICD and were already part of an (R&D project). Secondary prevention ICDs were offered to individuals who survived sudden cardiac arrest or had a history of dangerous and recurrent abnormal heart rhythms, which is relevant for this study due to the chronicity of the disease and related self-care activities. While patients were similar in having an ICD, their underlying cardiac diagnosis, possible comorbidities, and psychosocial situation differed substantially. Participants were recruited through a mix of purposive sampling and self-signup to ensure that patients

XSL•FO
RenderX

were interested and not too ill to participate. Of the 65 ICD patients we invited to participate, 27 ICD patients provided written informed consent to this study, which explored the experiences of activity data as related to being an ICD patient and the data's potential for predictive analytics of dangerous arrhythmias. Of the 27 patients, 25 patients were male (93%), and 2 patients were female (7%); the average age was 57 years. The sample largely reflected the demographic profile of ICD patients in Denmark in 2017, when 18% of patients were female and the majority of procedures were carried out on patients aged 55-74 years [42].

The participants were provided with and instructed to wear a Fitbit Alta HR (Fitbit, San Francisco, CA), which is a wristband activity tracker that can record and visualize heart rate, sleep, and steps onscreen and in a Fitbit mobile app. They were informed that wearing the activity tracker was unrelated to their treatment at the clinic and that our purpose was to explore how they experience the relationship between activity data and their heart disease.

Data Collection: Semi-Structured Interviews

Data were collected with 66 semistructured interviews in 3 overall iterations using 3 interview guides. The first iteration aimed to create a baseline of patients' expectations concerning activity-illness relationships and get a sense of their embodied experience of everyday living with an ICD. The second iteration aimed to understand the initial 2-week experiences with activity tracking using concrete examples from their own tracking data and asking them what they had learned or wondered about when using the Fitbit and looking at the data. The third iteration aimed to understand the longer-term experiences and data-sensing practices and any ambivalences arising from using the Fitbit (4–49 weeks).

Patients were interviewed individually (sometimes with relatives) in their homes or in locations convenient to them (eg, workplace or hospital office space). Each interview lasted between 25 minutes and 1 hour and 40 minutes. During interviews, the interviewers took field notes and pictures of participants showing concrete examples of activity data in the Fitbit mobile app to support the analysis of the data. All interviews were audio-recorded and transcribed either in full or in selected passages.

Data Analysis

The interview transcripts were hand-coded iteratively, following an abductive reasoning logic [43,44], starting with a joint analysis workshop after the second interview with all 27 patients to identify emergent themes regarding data sensing and insights from the data to be followed up in the final interviews. At this stage, field notes and pictures from the interviews offered an interpretive aid, offering contextual guidance for those of us who had not been present at the interview. Upon the final round of interviews, another joint workshop solidified and elaborated the initial insights and ideas against the body of related work, to develop a joint analytical framework and coding protocol. This framework combined dimensions of data sensing — knowing, feeling, and evaluating the self through data with contextual embedding and experiences of illness in daily chronic living [45] — to clarify how patients with an ICD make sense of Fitbit data relating to their heart disease. Finally, we recoded the complete empirical material manually according to these dimensions, first individually and then together, to ensure reliability in producing a thematically organized analysis of patients' data sensing [46].

Study Approval and Ethical Considerations

This paper was based on a substudy of the SCAUT research and development project, which was approved by the Danish Data Protection Agency and reviewed by the National Board of Health and Danish National Committee on Health Research Ethics (H-19029475). We took several measures to respond to possible ethical concerns. First, we ensured voluntary participation through an open invitation with self-signup, and we emphasized in interviews that participants could opt out at any time. Second, we communicated with all participants between interviews to ensure they were comfortable with wearing the wristband. We made sure that the participants understood that the Fitbit activity tracker was a consumer device and not a clinical device and that the data did not have diagnostic validity. Finally, we adjusted the conversation in interviews accordingly if patients expressed specific health concerns brought on by their interaction with the Fitbit; specifically, we urged them to contact their health professional for guidance.

Results

Device Engagement

Our results showed how patients with an implanted ICD device engaged with and made sense of activity data from the Fitbit in the context of chronic illness and self-care. Most (18 of the 27 participants) related their real-time heart rate, sleep, and step count data directly to their heart disease (Table 1). The remaining participants, however, connected the data to leisure activities, wellness, and exercise. Some portrayed themselves as not being a patient, explaining that they mostly did not have symptoms. Two patients chose to opt out after only wearing the tracker a few times owing to finding the wristband annoying to wear or simply losing interest (P6, P7). Patients used the activity tracker for an average of 26.1 weeks and took breaks from using it an average of 5.9 weeks.

The patients who did relate the data directly to their illness did so in 3 overall dimensions (Table 2): as something that generated new knowledge, as something that raised affective responses, and as something that could be used to evaluate themselves and their overall health. Within these 3 dimensions of experience, patients accounted for a range of positive, negative, and ambivalent experiences with activity data. For an extended analysis of the affective dimension and the consequences for patients' interpretation of Fitbit data, see [47].

XSL•FO
RenderX

Table 1. Overview of participating patients with chronic heart disease and an implantable cardioverter-defibrillator (n=27).

Patient number	Age (years)	Sex	Year of ICD[a] implant	Symptoms experienced	Number of weeks using the Fitbit (not using)	Experienced Fitbit data relating to heart disease
P1	67	Male	1998	No symptom experiences; experienced palpitations before	18 (0)	No
P2	61	Male	2015	Severe chest pain and shortness of breath	47.5 (0.5)	Yes
P3	41	Male	2009	No symptom experiences; experienced shortness of breath before	41 (6)	No
P4	55	Male	2014	Dizziness and sometimes fainting	13 (5)	Yes
P5	66	Male	2010	Dizziness and sometimes fainting	8 (23)	Yes
P6	67	Male	2015	No symptom experiences	9.5 (1.5)	N/A[b]
P7	28	Male	2008	No symptom experiences	6 (3)	N/A
P8	69	Male	2015	No symptom experiences (primary prophylaxis)	36.5 (11)	Yes
P9	47	Male	2008	No symptoms experiences related to his heart disease; lung disease; difficulties exercising	33,5 (14.5)	Yes
P10	61	Male	2010	Sometimes feeling very tired	30 (8)	Yes
P11	59	Male	2006	Shortness of breath and sometimes sleep problems; finds it difficult to feel his heart rate	8.5 (0.5)	Yes
P12	66	Male	2015	Dizziness and sometimes fainting; anxious about getting a shock and experiences depression	49 (0)	Yes
P13	67	Male	2017	No symptom experiences; rarely experience fainting; leg tenderness and muscle fatigue	44.5 (0.5)	Yes
P14	52	Female	2008	No symptom experiences	49 (0)	No
P15	61	Female	2004	No symptom experiences; sometimes being anxious about the ICD/irregular heartbeats	14 (1)	Yes
P16	47	Male	2014	Dizziness and sometimes fainting	35.5 (11.5)	No
P17	45	Male	2013	Dizziness and shortness of breath during high activity levels	35 (7)	Yes
P18	67	Male	2009	No symptom experiences	18 (29)	Yes
P19	66	Male	2005	Palpitations daily; experiences periods of depression and has restless leg syndrome	20.5 (1.5)	No
P20	69	Male	2014	No symptom experiences, except sometimes shortness of breath	39.5 (8.5)	Yes
P21	38	Male	2008	No symptom experiences; worries daily about having a cardiac arrest again	13.5 (0)	Yes
P22	59	Male	2001	Shortness of breath and sometimes dizziness when exercising and running; rapid heartbeats and chest pain	19.5 (0.5)	Yes
P23	49	Male	2017	No symptom experiences	47 (2)	Yes
P24	74	Male	2017	No symptom experiences	42.5 (6)	Yes
P25	51	Male	2014	No symptom experiences	9 (15)	No
P26	56	Male	2010	No symptom experiences; sometimes feels palpitations and shortness of breath	4 (0)	No
P27	58	Male	2014	Shortness of breath; sometimes sleep problems; kneepain due to osteoarthritis	12 (5)	Yes

[a]ICD: implantable cardioverter-defibrillator.

[b]N/A: not available because the patient did not describe a relation between the activity data and his or her heart disease.

XSL•FO
RenderX

Table 2. Dimensions of how patients experienced activity tracking related to their disease.

Experiential dimension	Experience
Knowing	
Positive: gaining insight	Learning that heart disease increases one's average resting heart rate (P2, P4)
	Learning that medication influences the heart rate (P5, P22, P23, P27)
	Learning that activity improves one's average heart rate (P4, P21)
	Using activity data to monitor heart pumping ability (P10)
Negative: evoking doubts	No new learnings: Sensing is more useful than activity data (P1, P5, P16, P22)
	Doubting heart rate data (P2, P22)
	When doubt becomes mistrust (P12)
Feeling	
Positive: being reassured	Feeling safe through Fitbit reassurance (P11, P12, P17)
	Reassurance prompts activity (P24)
Negative: becoming anxious	Both insights and doubts can introduce new anxieties (P12, P13, P15, P23)
Evaluating	
Positive: promoting improvement	Being nudged and getting praise (P19, P20, P23, P24)
Negative: exposing failure	Recognizing a nudge but not knowing what to do about it (P13)
	Not getting the proper reward: the invisibility of "good" activities (P18)
	Self-disappointment with poor numbers (P8, P17, P24)
	Ignoring or resisting nudges (P18, P19)

Knowing: Gaining Insight and Evoking Doubts

As part of their motivation for participating in the study, many participants expressed an interest in knowing more about their health and body and about the possible relationships between their daily activities and their heart. In our interviews, it became clear that patients actively sought knowledge about their heart and health through the readings on the Fitbit and that some gained insight from the data. For others, the knowledge they gained from sensing their own bodies was more useful than the Fitbit data. Finally, some participants experienced that the data did not align with their activities and sensory experiences; thus, they doubted the accuracy and trustworthiness of the data.

Learning That Heart Disease Increases One's Average Resting Heart Rate

One patient gained insight into his average resting heart rate and discovered it was higher than he expected:

> *My resting heart rate ought to be 60, but it is 80, and as soon as I start moving, it goes up to 100 or 120.* [P2]

It did not concern him since there is not much the clinicians can do about it, as he said. Another patient had the same kind of insight:

> *Then, the pulse is in a zone where I am physically active. That is what it signals. But I am not physically active. I am just walking. That has been an eye-opening experience.* [P4]

Learning That Medication Influences the Heart Rate

Several patients knew, speculated, or learned about how medication can affect the heart rate. By looking at the Fitbit data, P5 noticed that his average heart rate decreased when on vacation, and he considered that his medication might have an effect. P27 also noticed that his medication influenced his heart rate:

> *Obviously, the heart rate follows how active you are; but, for healthy people, it's not abnormal that it reaches 116. But I get 13 pills in the morning and 3 in the afternoon.*

P23 also noticed that medication influences his heart rate:

> *My beta blockers have been reduced in dose, which may also have something to do with the changes in the resting heart rate.*

P22 had always been exercising; however, after an event a few years ago, he was prescribed a double dosage of beta-blockers, and it was decided to reduce the dosage again because his heart rate had dropped too low. Now, he finds it exciting to use Fitbit to learn about the relationship between medication, exercise, and heart rate and to confirm his heart rate data:

> *I still monitor what happens. This morning I saw an increase in my average heart rate of 12% and my maximum heart rate of just over 15%. And that's what's fun and what I use it for because it has annoyed me extremely that I couldn't run like I used to.*

XSL•FO
RenderX

Learning That Activity Improves One's Average Heart Rate

P21 noticed that there is a relationship between exercising and average heart rate:

> *After serious ice hockey training, the average heart rate is higher on the day after, and then it drops in the weekend.*

Similarly, P4 discovered that when he is more active, his average heart rate drops:

> *During the 20 days I have been wearing the Fitbit, I have seen a drop in my heart rate by 10 heartbeats per minute. I think that's crazy. I have been more active, yes, and I have been walking more. I have set exercise goals.*

He experienced this as a positive thing:

> *I think it's fine that my resting heart rate goes down because, if it goes down, the heart does not have to work as hard. It makes a difference whether my heart rate is 70 or 60 beats per minute when resting.*

Using Activity Data to Monitor Heart Pumping Ability

Several patients had experience with other activity trackers. One patient (P10) described how he used the data to monitor his heart condition. He was diagnosed with heart failure, and his heart had a reduced pumping ability, "Down to around half of what is normal," he said. He used the Fitbit to address his concern that his heart rate will drop even lower:

> *I use it to keep an eye on my condition. When I go spinning, I do the same intervals, and I can therefore see if I have burned the same calories in that hour. So, I can say, okay, I'm fairly stable, or I can see if it has gone down. Because then it might be my heart getting worse in its ability to pump.*

No New Learnings: Sensing is More Useful Than Activity Data

Several patients had learned over the years to sense developments in their heart condition. One patient explained that, in the past, when he got rapid and dangerous heartbeats and the ICD began treatment with antitachycardia pacing to terminate the arrhythmia, he sat down and waited until it passed. He used to call the Heart Centre when he felt the symptoms to get confirmation about the episodes. For him, the Fitbit did not generate any new insights apart from what he already knew from sensing and listening to his body when exercising in the gym:

> *I know I cannot do physically very demanding exercises. I have come to terms with that. So, I have not received any new extra information via Fitbit.* [P1]

P22 found it useful to use the activity tracker to get confirmation on fast heartbeats, for example, when running in the woods. However, he trusts his senses more:

> *If there is a connection between what I feel in my body and what the tracker shows, then I react. But when the tracker shows something that I don't notice in my body, I consider it an IT error.*

Another patient explained that he has tried to look at the Fitbit data when he gets sudden dizziness; however, it provides no explanation:

> *Sometimes I suddenly experience a "dive," and then I just have to hold on to something. It's very different how often it happens. But when I look at [the Fitbit activity data], it does not show anything. So, I can't find the reason for getting so dizzy.* [P5]

Similarly, P16 explained that there is no connection between symptoms and heart rate:

> *There is no connection. I've tried to get symptoms when doing gymnastics where I had the pulse all the way up, and I've tried to get symptoms while I was sleeping. I have not found any connection at all.*

Doubting Heart Rate Data

Heart rate data also created doubt among some participants. One patient doubted that the heart rate data presented on his Fitbit were accurate when walking:

> *But I have seen sometimes when out for a walk, my heart rate is really high. Average 166—I think that is a little high. And, sometimes, I have seen it going up to 197 beats, and I don't know why.* [P2]

Another patient found the heart rate data "weird" at times. He experienced chest pain a few times when he was out running; however, no answers could be found in the activity data:

> *I felt a little uncomfortable — one might suddenly think of a blood clot. But I can't see in any of the trackers that the pulse has been particularly high or particularly low.* [P22]

At other times, when he was just walking, there were sudden peaks in the heart rate data:

> *It seems strange. It's just right there — a peak up. It almost seems like a mistake that it goes from 80 to almost 170. It seems completely messed up what happens here.* [P22]

Several other patients experienced that their heart rate showed unusual fluctuations on Fitbit.

When Doubt Becomes Mistrust

One patient had disease-related anxiety and had difficulties sleeping. Initially, the Fitbit data concerning his sleep revealed that he slept more than he thought he did when he woke in the morning after a difficult night. However, when he later saw that the Fitbit had registered "sleep" while he was calmly watching a movie, he and his wife started distrusting the measurements of sleep altogether:

> *This Saturday, we looked at sleep, for instance, and we could not make the numbers fit because he had been awake a lot. The data was wrong, and we do not know how it works. Then you go on to think, can you trust this device at all?* [wife of P12]

XSL•FO
RenderX

Across participants, the sleep tracking measures of Fitbit were noted as unreliable and not reflective of actual sleep. For some participants, discovering that sleep data were inaccurate led to mistrust and a general, critical understanding of the measurements provided by Fitbit.

Feeling: Being Reassured and Becoming Anxious

Being diagnosed with a heart condition often introduces profound anxiety into patients' (and their relatives') lives. It is well-known that patients with an ICD are at an increased risk of being diagnosed with depression and anxiety [48]. In our interviews, some patients had different levels of anxiety and used the Fitbit to reassure themselves that their heart was doing okay. For some, this helped them engage more in physical activity — something they might have held back on owing to fear of provoking an attack (ie, kinesiophobia) [49]. However, the Fitbit — partly owing to the doubts and uncertainties introduced as described — could also spark new and additional anxieties and negative feelings regarding participants' health.

Feeling Safe Through Fitbit Reassurance

For some patients, having a heart condition raised their embodied attention, making them very alert to bodily signs:

> *As soon as there is even a little thing in these zones [in his chest region], and I would even say just one, like a sprain, I get nervous.* [P12]

This patient, who is very affected by anxiety and depression, used the Fitbit to reassure himself that he is not having a cardiac arrest:

> *Then, I was out chopping firewood. I felt like it began to beat both in a weird way, and it felt like it beat really fast, but it did not. There was nothing. There was nothing to be seen [on the Fitbit] anyway.* [P12]

For this patient, seeing his heart rate within the normal spectrum reassured and calmed him down:

> *Now I get certainty. Is something wrong or not.*

Another patient, who experienced getting a shock from his ICD after walking the stairs, explained that*:*

> *Being able to see my heart rate is normal creates a sense of security because I'm not able to feel when my heart rate rises. A normal rhythm means that there is nothing to be afraid of — no danger is underway.* [P11]

Similarly, P17 used the heart rate data as support in vulnerable situations:

> *I've tried it so many times, to have those VTs [rapid heartbeat] and I know what it leads to, and that's what I fear.*

One time he was lying down, and he used the heart rate on the Fitbit tracker to get confirmation on the duration of rapid heartbeats:

> *If it lasted more than 5 minutes, I would have called 112 [emergency services]. Because, sometimes, I think when I get that feeling of fast heartbeats, it may well be imagination.*

Reassurance Prompts Activity

Holding back during physical activity was something several of the informants touched upon, as some had had a heart attack while exercising or had experienced an ICD shock when climbing a flight of stairs. Checking their heart rate while doing more physically demanding tasks and sports motivated them to increase their activity:

> *I'm not afraid to have a high pulse when we do the Bikefit exercise at gymnastics … because I can see it goes down again [his heart rate].* [P24]

Both Insights and Doubts Can Introduce New Anxieties

In the first section, we touched upon some of the doubts concerning the validity of the Fitbit data. For the anxious patient in need of reassurance, this uncertainty can be stressful, as noted by the wife of P12. P15 described it as follows:

> *There are plenty of worries when you have a heart disease. You don't need unnecessary things that make you worry more.*

The Fitbit sleep data, for example, created unwanted attention to what it meant for her health:

> *What does it mean for my health? Am I sleeping enough or too little, and what can I do about it? Such concerns arise, which I could do well without.*

She also experienced getting a high pulse that was "completely unprovoked" and then seeing it on the Fitbit:

> *But I can't do anything about it, and I can't use it for anything—unless they can see it in the clinic.*

One patient noticed that Fitbit wants him to sleep 8 hours per night — a goal he rarely reaches:

> *I have always had this sleep pattern, and it did not bother me until I got this Fitbit, which says I should sleep 8 hours a night … I get worried and start to question whether I ought to sleep more.* [P13]

Finally, the introduction of new concerns and anxieties by Fitbit goes beyond patients' individual experiences. P23 explained that he tried to avoid drawing attention to his Fitbit wristband when being around his teenage kids:

> *I think it reminds them a little bit of something bad … it's interpreted negatively—like someone has to keep an eye on me.*

Evaluating: Promoting Improvement and Exposing Failure

As we presented in the previous sections, patients used Fitbit's numerical representations to make sense of their bodily sensations in the context of self-care. The fact that the Fitbit device allows them to see — as numbers, icons, and graphics — something that they usually relate to as sensations provides a new form of motivation. Setting targets for activity, getting notifications, and seeing achievements represented visually can be encouraging. Concurrently, however, the Fitbit also exposes unmet goals, thereby inducing self-disappointment or even shame. The Fitbit does not register or "see" all the activity that the patients found to be relevant as fair representations.

XSL•FO
RenderX

Being Nudged and Getting Praise

Several of the patients talked about how the Fitbit nudged them to stay physically active — it is "*a kick in the butt*," as one patient noted (P19). Users can set their activity goals as they please; however, most went with the default setting of 10,000 steps a day. All patients looked daily to see if they had reached that goal, and some found it motivating to see the numbers and get positive feedback (visually provided in the form of stars) from the user interface. One patient described himself laughingly as:

> *Addicted to it because it says you have to walk 10,000 steps a day, and that fits with some of our walks. It becomes a sport; it gets me going.* [P20]

Others noted how the gamification element invoked in the Fitbit led to small changes in their daily activities. For P23, Fitbit prompted him to take the stairs instead of the elevator to get more steps. It also prompted him to get up when it beeped every hour to take 250 steps by walking up and down the hall during work breaks. For these patients, Fitbit is "a little push in the right direction" (P23) and "an inspiration to continue" (P24).

Recognizing a Nudge But Not Knowing What to Do About It

Whereas simple nudging features such as awarding stars for accomplishing specific activity benchmarks seemed to motivate participants, there were also examples of participants becoming unsure of what constitutes appropriate activity. For example, Fitbit (by default) beeps once every hour during the day to pace the wearer to walk 250 steps every hour. For P13, this nudge suggested that average but regular activity throughout the day might be preferable to his usual practice of lumping activity together for more intense periods of exercise. It made him wonder if he should organize his workout differently, even if this wondering did not lead him to make any actual change.

Not Getting the Proper Reward: The Invisibility of "Good" Activities

After wearing the Fitbit for some time and having acquainted themselves with the collected activity data, some participants reported being frustrated that the Fitbit did not really measure all their activity. They did not feel "seen" by the device and rewarded properly for their efforts. This is particularly the case for those participants who did cycling, CrossFit, and other activities beyond walking and running as part of their everyday life activities. One participant lamented:

> *It is actually misleading because most of my activity is on a bike, and it does not register this. But it is also exercise.* [P18]

These experiences added to the doubt in the data and the accuracy of measurements described previously.

Self-Disappointment With Poor Numbers

If positive feedback is seen as motivating further activity tracking, conversely, the negative feedback from Fitbit made some participants feel disappointed with or ashamed of themselves because it highlighted that they had not been active enough:

> *Well, I guess it has to do with that bad conscience you get the next day, if you cheat.* [P8]

This was mentioned by several patients, and, for some, it made the Fitbit less attractive. P24 talked of self-disappointment when receiving negative feedback from Fitbit and noted that there might be someone else looking at their data, surveilling whether they reached their goals:

> *Yes, because it gossips all the time, noting that I have not walked far enough … I think you could always find an excuse for not walking; like, it's raining.*

For P17, the low step count created negative emotions on "bad days" and became linked to his heart condition:

> *It gives me a little guilty conscience that I do not get much exercise because, I have no doubt, the more weight I gain, the more fat is generated around my heart, and the harder it is for the heart to pump.*

Ignoring or Resisting Nudges

Some participants tempered their engagement with Fitbit by actively resisting to do what the device suggests:

> *I know that it tells me, every once in a while, that it is time to go for a walk. But I decide when it is time to go for a walk. And then it says, let's go, and I'm like no way because I don't have the time right now.* [P18]

For some, the Fitbit's nudging was simply a source of annoyance:

> *It can be a little irritating watching all the green "pling pling." I don't want that; I don't care about it. I just want the info.* [P19]

Yet, for others, such as P18, resisting the nudge to walk every hour was followed by a deeper reflection about whether activity tracking is good at all for his health:

> *I actually think it is a little unhealthy to measure oneself all the time. It comes to take up a lot, in my life, and I don't think it is that important.*

Discussion

Principal Findings

Our study contributes to understanding how patients with chronic heart disease, as opposed to healthy individuals, experience activity data from consumer self-tracking devices in self-care. We found that patients with an implanted ICD relate the activity data to their illness experience and their self-care activities in 3 overall dimensions: as something that generated new or destabilized existing knowledge, as something that raised affective responses, and as something that could be used to evaluate themselves and their overall health.

The distribution of patients' experiences on a continuum from positive to negative suggests that activity data had "dual effects," which means that the data created as much as it solved the problem of chronic illness [50]. The problems that people with chronic illness have to deal with become mediated in new ways and what may count as "normal," "good," "problematic," or

XSL•FO
RenderX

"bad" may change accordingly. Positive and negative effects potentially co-exist and support the experiential ambivalence that studies of self-tracking and activity data have also found among leisurely users and quantified self-enthusiasts [22,45,51,52]. The concept of "ambivalence" unites this stream of research in which patients' attitudes towards digital health devices "neither are consistently negative (implied by the notion of 'rejection') nor consistently positive (implied by the notion of 'acceptance')" [45]. Conflicting or ambivalent experiences appear constitutive of self-tracking: "doubt, guilt, fear, shame, dismay, disappointment, and hesitation as well as joy, relief, excitement, enthusiasm, and pride" [51].

Generating knowledge from interpreting activity data is often portrayed as the essence of self-tracking. For healthy individuals, it may comprise discoveries about physical performance in everyday life and adopting healthier behavior [33,35]. For the patients in this study, performance-oriented and fitness-oriented development of self-knowledge also surfaced. Patients obtained new insights about how exercising improves their average heart rate and that their heart disease may be the reason for a higher resting heart rate.

Other more disease-specific reflections surfaced as using Fitbit data to monitor the status and development of heart failure (heart pumping ability) and speculating about how heart medication affects the pulse. Unusually high heart rate data created doubt when walking or when connected to chest pain while running. Therefore, Fitbit data became part of generating a type of lay and personal expertise for, at best, supporting day-to-day self-care activities and living with a chronic disease and, at worst, creating uncertainty. This kind of "experiential knowledge" or "patient knowledge" [53-55] is often considered distinct from medical and scientific knowledge in that it is a by-product of bodily sensing and coping with daily practicalities of the disease as well as it is re-appropriated medical knowledge used to contribute, but also dispute, the biomedical perspective [55].

For patients, Fitbit data can provide support for self-care with informational cues alongside bodily sensations and experiences in the development of "know-now" [53] (ie, understanding what is going on or deciding what action to take). As opposed to healthy individuals' knowledge-making with Fitbit, the unsupported lay interpretation of medically unvalidated heart rate data poses a risk for patients taking inappropriate action, for example, using Fitbit heart rate numbers to diagnose a cardiac arrest when running and deciding whether to keep on running or when getting chest pain and becoming dizzy in the office and then using Fitbit to decide what to do. The practical implications of patient knowledge generation from Fitbit suggest that patients should not be left alone with interpreting activity data as part of self-care. Deploying self-tracking and activity data in chronic care should be carefully accompanied by a purposeful clinical intervention such as rehabilitation and training programs where clinical staff can support patients in interpreting activity data, and data visualization should be designed to support meaningful action in the context of self-care.

The affective dimension of self-tracking when living with a chronic heart disease also emerged as loaded with ambivalence. Fitbit numbers may provide numerical reassurance, which can relieve acute anxiety related to unclear bodily sensations and provide confidence to exercise. Concurrently, heightened attention to Fitbit data can also introduce new uncertainties and anxieties. Moreover, it is important to underline that the reassurance of the Fitbit data is not based on clinical evidence and, while reduction of acute anxiety is important to patients' wellbeing, there is a risk that the numbers provide pseudoproof not sufficiently reliable to indicate anything clinically relevant about the patient's condition. Given the prevalence of mental health issues, such as anxiety and depression, related to heart disease and the lack of mental health services for these patients, it is vital to consider the potential negative interactions between health tracking and mental health. Patients with chronic mental health comorbidities should not be left to try to cope with serious mental health issues alone with consumer devices.

Taken together, we see a tension between Fitbit's promotion of success and exposure of failure to comply with set standard activity levels in the actual experience of using Fitbit. The ambivalence of knowing, feeling, and evaluating one's chronic health condition against activity data from consumer devices, as opposed to clinically validated instruments, poses a concern for how engagement with data is placed in chronic care contexts as well as the purposes inscribed in the design of these new devices. In their analysis of ambivalence in mobile health for HIV care, Marent and colleagues [45] argued for the need to consider how the tension implied with ambivalence is embodied by particular bodily conditions and embedded in particular relationships and environments. Our study concerns people who, in clinical terms, have a chronic condition; however, the embodiment of this condition varies among participants. For some, the disease has a continuous and very challenging presence related to managing and coping with severe symptoms (see Table 1), while others tell us they do not feel sick at all. We suggest that the ambivalence of Fitbit data is more problematic when used by people who have a chronic condition and even more so for people who feel very challenged by their disease.

This relates to a second point about embeddedness. The ambivalence of Fitbit data should be understood in relation to its embeddedness in everyday contexts unrelated to clinical contexts of treatment. To what extent do people in these contexts have support from others such as relatives, peers, or health professionals who choose to engage in handling the ambivalences they encounter? What resources can they mobilize to act when experiencing doubt, anxieties, or other concerns when self-monitoring with Fitbit? With our paper being specifically concerned with patients who engaged with Fitbit data outside the established relationships of health care institutions, these questions become critical. Navigating benefits and harms of this form of active engagement with personal health data is, to a large degree, dependent on individual circumstances, resources, and networks, leaving inequalities potentially less mitigated by public health systems. We find that these are very central insights to take into account in research that focuses on how self-care practices can be furthered by harnessing the power of data and personal health technology. Too often, this literature focuses narrowly on individual

XSL•FO
RenderX

information processing and empowerment while disregarding the relational and situational embeddedness of chronic disease management. It not only neglects insights from health information-seeking literature, which has convincingly shown that patients' information behavior is more often based on serendipity, avoidance, blissful ignorance, indolence, bewilderment, and indolence than on rational choices and reflections [56], but also neglects that self-care practices are always collective in the sense that they are embedded in complex sociomaterial relationships [57,58]. What our study further adds to the literature on self-care and technology, then, is that technology and data mobilized outside the established arrangements of health care with consumer tracking devices may introduce new ambivalences that some patients may have difficulties managing without professional assistance.

For human-computer interaction research and the design of activity-data visualization for patients, our study aligns with the understanding that data supports different levels of reflection and serves multiple purposes [59]. The implications of our study, we suggest, is that consumer activity-tracking devices deployed in health care contexts should be designed to also support collaborative reflection (ie, "co-reflection") with health professionals, rather than focusing mainly on individual reflection [60]. Research on ways to support the shared work of tracking [61], co-reflection, and "co-care" [62] might be necessary to consider for personal informatics research in chronic self-care [25,47,63].

As an increasing number of people are generating and interacting with digital and individual health data outside the context of clinical practice, issues of inequality in health must be considered. Thus, there are important implications to consider in this typically optimistic, yet blurred, realm of "personal health data" (actualized for wellness purposes) and "patient-generated data" (actualized for clinical purposes).

Conclusions

We presented the findings from an explorative intervention study of how patients with a heart arrhythmia who have an implanted ICD experience activity data from Fitbit concerning their self-care and chronic illness. The aim was to further emergent literature and offer crucial empirical insight into the introduction of wellness tracking devices to various forms of chronic care management and the associated user experiences.

Through repeated semistructured interviews with 27 patients equipped with a Fitbit wristband, we offer support and further elaboration on existing work on patients' ambivalent experiences. We found that wearable activity trackers actualize patients' experiences across 3 dimensions on a spectrum: (1) gaining new knowledge versus evoking doubt, (2) feeling reassured versus becoming anxious, and (3) evaluating one's health by celebrating improvements and exposing failure. The experiences of individual patients can reside more on one end of the spectrum, can reside across all 3 dimensions, or can combine contrasting positions and even move across the spectrum over time. While activity data from wearable devices may be a resource for self-care through reassurance and motivation, they may also constrain patients and create increased uncertainty, fear, and anxiety.

The ramifications of knowing, feeling, and evaluating one's chronic health condition against activity data from consumer devices, as opposed to clinically validated instruments, are largely unexplored. Our study suggests that we need critical attention in scholarship and health care practice concerning how engagement with such data is practiced in chronic care contexts, not least to assess how the purposes inscribed in the design of these new devices may be molded and twisted in self-care when meeting the logics, needs, and abilities of patients in different health care circumstances. Designers and health authorities should consider this complexity and ambiguity when determining the usefulness of self-tracking data in chronic illness.

Acknowledgments

The authors wish to thank all the patient participants and their relatives who shared their experiences with us for this study. We also thank Tanja Munk Warmdahl, Amalie Lund Bendtsen, Oliver Carl Clemmensen, Andreas Millarch, and Anders Riis Vestergaard for their collaboration in interviewing, transcribing, and analysis workshops. This study is cofunded by the Innovation Fund Denmark #72-2014-1 and the University of Copenhagen, Vital Beats, and Medtronic. This study was also, in part, supported by a grant from the Danish Velux Foundations #33295.

Conflicts of Interest

None declared.

References

1. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
2. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
3. Garge GK, Balakrishna C, Datta SK. Consumer Health Care: Current Trends in Consumer Health Monitoring. IEEE Consumer Electron. Mag 2018 Jan;7(1):38-46. [doi: 10.1109/mce.2017.2743238]

XSL•FO
RenderX

4. Weiner K, Will C. Thinking with care infrastructures: people, devices and the home in home blood pressure monitoring. Sociol Health Illn 2018 Feb;40(2):270-282. [doi: 10.1111/1467-9566.12590] [Medline: 29464773]
5. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
6. Zhu H, Colgan J, Reddy M, Choe E. Sharing Patient-Generated Data in Clinical Practices: An Interview Study. 2016 Presented at: AMIA Annual Symposium Proceedings; 2016; Chicago, IL p. 1303-1312.
7. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You Get Reminded You're a Sick Person": Personal Data Tracking and Patients With Multiple Chronic Conditions. J Med Internet Res 2015 Aug 19;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
8. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
9. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
10. Liu YH, Scifleet P, Given LM. Contexts of Information Seeking in Self-tracking and the Design of Lifelogging Systems. 2014 Presented at: MindTheGap@ iConference; 2014; Berlin, Germany p. 44-50.
11. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-Care Technologies in HCI. ACM Trans. Comput.-Hum. Interact 2015 Dec 14;22(6):1-45. [doi: 10.1145/2803173]
12. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Education and Counseling 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]
13. Schermer M. Telecare and self-management: opportunity to change the paradigm? J Med Ethics 2009 Nov 30;35(11):688-691. [doi: 10.1136/jme.2009.030973] [Medline: 19880706]
14. Greenwood DA, Young HM, Quinn CC. Telehealth Remote Monitoring Systematic Review: Structured Self-monitoring of Blood Glucose and Impact on A1C. J Diabetes Sci Technol 2014 Mar 21;8(2):378-389 [FREE Full text] [doi: 10.1177/1932296813519311] [Medline: 24876591]
15. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Health Sociology Review 2016 Aug 05;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]
16. Chau JP, Lee DT, Yu DS, Chow AY, Yu W, Chair S, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. Int J Med Inform 2012 Oct;81(10):674-682. [doi: 10.1016/j.ijmedinf.2012.06.003] [Medline: 22789911]
17. Burri H, Senouf D. Remote monitoring and follow-up of pacemakers and implantable cardioverter defibrillators. Europace 2009 Jun 26;11(6):701-709 [FREE Full text] [doi: 10.1093/europace/eup110] [Medline: 19470595]
18. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
19. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]
20. Fiske A, Prainsack B, Buyx A. Data Work: Meaning-Making in the Era of Data-Rich Medicine. J Med Internet Res 2019 Jul 09;21(7):e11672 [FREE Full text] [doi: 10.2196/11672] [Medline: 31290397]
21. Piras EM. Beyond self-tracking: Exploring and unpacking four emerging labels of patient data work. Health Informatics J 2019 Sep;25(3):598-607. [doi: 10.1177/1460458219833121] [Medline: 30848690]
22. Ruckenstein M, Schüll ND. The Datafication of Health. Annu. Rev. Anthropol 2017 Oct 23;46(1):261-278. [doi: 10.1146/annurev-anthro-102116-041244]
23. Kaziunas E, Ackerman M, Lindtner S, Lee J. Caring through Data: Attending to the Social and Emotional Experiences of Health Datafication. 2017 Presented at: CSCW '17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; Feb 2017; Portland, Oregon p. 2260-2272. [doi: 10.1145/2998181.2998303]
24. Figueiredo MC, Caldeira C, Reynolds TL, Victory S, Zheng K, Chen Y. Self-Tracking for Fertility Care: Collaborative Support for a Highly Personalized Problem. 2017 Dec 06 Presented at: Proceedings of the ACM on Human-Computer Interaction (CSCW); 2017; Portland, Oregon p. 1-21. [doi: 10.1145/3134671]
25. Andersen TO, Andersen PRD, Kornum AC, Larsen TM. Understanding Patient Experience: A Deployment Study in Cardiac Remote Monitoring. 2017 May Presented at: the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2017; Barcelona, Spain p. 221-230. [doi: 10.1145/3154862.3154868]
26. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of Different Activity Monitors to Count Steps in an Inpatient Rehabilitation Setting. Phys Ther 2017 May 01;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]
27. Benzo R. Activity monitoring in chronic obstructive pulmonary disease. J Cardiopulm Rehabil Prev 2009;29(6):341-347 [FREE Full text] [doi: 10.1097/HCR.0b013e3181be7a3c] [Medline: 19940638]

XSL•FO
RenderX

28. Rosenberg D, Kadokura EA, Bouldin ED, Miyawaki CE, Higano CS, Hartzler AL. Acceptability of Fitbit for physical activity tracking within clinical care among men with prostate cancer. 2016 Presented at: AMIA Annual Symposium Proceedings; 2016; Chicago, IL p. 1050-1059.
29. Thorup CB, Grønkjær M, Spindler H, Andreasen JJ, Hansen J, Dinesen BI, et al. Pedometer use and self-determined motivation for walking in a cardiac telerehabilitation program: a qualitative study. BMC Sports Sci Med Rehabil 2016;8:24 [FREE Full text] [doi: 10.1186/s13102-016-0048-7] [Medline: 27547404]
30. Ruckenstein M, Pantzar M. Datafied Life: Techno-Anthropology as a Site for Exploration and Experimentation. Techné: Research in Philosophy and Technology 2015;19(2):191-210. [doi: 10.5840/techne20159935]
31. Didžiokaitė G, Saukko P, Greiffenhagen C. The mundane experience of everyday calorie trackers: Beyond the metaphor of Quantified Self. New Media & Society 2017 Mar 24;20(4):1470-1487. [doi: 10.1177/1461444817698478]
32. Sharon T, Zandbergen D. From data fetishism to quantifying selves: Self-tracking practices and the other values of data. New Media & Society 2016 Mar 09;19(11):1695-1709. [doi: 10.1177/1461444816636090]
33. Kristensen DB, Ruckenstein M. Co-evolving with self-tracking technologies. New Media & Society 2018 Feb 21;20(10):3624-3640. [doi: 10.1177/1461444818755650]
34. Lomborg S, Thylstrup NB, Schwartz J. The temporal flows of self-tracking: Checking in, moving on, staying hooked. New Media & Society 2018 May 31;20(12):4590-4607. [doi: 10.1177/1461444818778542]
35. Lomborg S, Frandsen K. Self-tracking as communication. Information, Communication & Society 2015 Aug 03;19(7):1015-1027. [doi: 10.1080/1369118X.2015.1067710]
36. Lupton D, Pink S, LaBond C, Sumartojo S. Digital Traces in Context: Personal Data Contexts, Data Sense, and Self-Tracking Cycling. International Journal of Communication 2018;12:647-665.
37. Schüll ND. Data for life: Wearable technology and the design of self-care. BioSocieties 2016 Oct 13;11(3):317-333. [doi: 10.1057/biosoc.2015.47]
38. Lupton D. Self-Tracking Modes: Reflexive Self-Monitoring and Data Practices. SSRN Journal 2014 Aug:2483549 [FREE Full text] [doi: 10.2139/ssrn.2483549]
39. Ehn M, Eriksson LC, Åkerberg N, Johansson A. Activity Monitors as Support for Older Persons' Physical Activity in Daily Life: Qualitative Study of the Users' Experiences. JMIR Mhealth Uhealth 2018 Feb 01;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]
40. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User Acceptance of Wrist-Worn Activity Trackers Among Community-Dwelling Older Adults: Mixed Method Study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
41. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
42. Hjerteforeningen (Danish Heart Association). ICD (Implantable Cardioverter Defibrillator). HjerteTal. 2018. URL: https://hjerteforeningen.dk/alt-om-dit-hjerte/hjertetal/hjertetaldk [accessed 2020-03-20]
43. Timmermans S, Tavory I. Theory Construction in Qualitative Research. Sociological Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
44. Blaikie N. Abduction. In: Lewis-Beck MS, Bryman A, Liao TF, editors. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: SAGE Publications; 2004.
45. Marent B, Henwood F, Darking M, EmERGE Consortium. Ambivalence in digital health: Co-designing an mHealth platform for HIV care. Soc Sci Med 2018 Oct;215:133-141. [doi: 10.1016/j.socscimed.2018.09.003] [Medline: 30232053]
46. Boyatzis RE. Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, CA: SAGE Publications; 1998:1-169.
47. Lomborg S, Langstrup H, Andersen TO. Interpretation as luxury: Heart patients living with data doubt, hope, and anxiety. Big Data & Society 2020 May 08;7(1):205395172092443-205395172092413. [doi: 10.1177/2053951720924436]
48. Pedersen SS, von Känel R, Tully PJ, Denollet J. Psychosocial perspectives in cardiovascular disease. Eur J Prev Cardiol 2017 Jun 16;24(3_suppl):108-115. [doi: 10.1177/2047487317703827] [Medline: 28618908]
49. Hoffmann JM, Hellwig S, Brandenburg VM, Spaderna H. Measuring Fear of Physical Activity in Patients with Heart Failure. Int.J. Behav. Med 2017 Dec 11;25(3):294-303. [doi: 10.1007/s12529-017-9704-x]
50. Lehoux P. The duality of health technology in chronic illness: how designers envision our future. Chronic Illn 2008 Jun;4(2):85-97. [doi: 10.1177/1742395308092475] [Medline: 18583444]
51. Salmela T, Valtonen A, Lupton D. The Affective Circle of Harassment and Enchantment: Reflections on the ŌURA Ring as an Intimate Research Device. Qualitative Inquiry 2018 Sep 22;25(3):260-270. [doi: 10.1177/1077800418801376]
52. Fotopoulou A, O'Riordan K. Training to self-care: fitness tracking, biopedagogy and the healthy consumer. Health Sociology Review 2016 Jun 02;26(1):54-68. [doi: 10.1080/14461242.2016.1184582]
53. Pols J. Knowing Patients: Turning Patient Knowledge into Science.. Science, Technology, & Human Values 2013 Sep 13;39(1):73-97. [doi: 10.1177/0162243913504306]
54. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011 Aug;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]

XSL•FO
RenderX

55. Storni C. Patients' lay expertise in chronic self-care: a case study in type 1 diabetes. Health Expect 2013 Aug 28;18(5):1439-1450. [doi: 10.1111/hex.12124]
56. Johnson DJ. Health-related information seeking: Is it worth it? Information Processing & Management 2014 Sep;50(5):708-717 [FREE Full text] [doi: 10.1016/j.ipm.2014.06.001]
57. Pols J. Care at a Distance. On the Closeness of Technology. Amsterdam: Amsterdam University Press; 2012:1-163.
58. Mol A. The logic of care: Health and the problem of patient choice. London and New York: Routledge; 2008:1-95.
59. Fleck R, Fitzpatrick G. Reflecting on reflection: framing a design landscape. 2010 Presented at: OZCHI '10: Proceedings of the 22th conference of the computer-human interaction special interest group of Australia on Computer-human interaction; Nov 22, 2010; Brisbane Australia p. 216-223. [doi: 10.1145/1952222.1952269]
60. Choe EK, Lee B. Toward supporting personalized tracking experiences in healthcare. Interactions 2019 Dec 26;27(1):84-87. [doi: 10.1145/3371287]
61. Vizer LM, Eschler J, Koo BM, Ralston J, Pratt W, Munson S. "It's Not Just Technology, It's People": Constructing a Conceptual Model of Shared Health Informatics for Tracking in Chronic Illness Management. J Med Internet Res 2019 Apr 29;21(4):e10830 [FREE Full text] [doi: 10.2196/10830] [Medline: 31033452]
62. von Thiele Schwarz U. Co-care: Producing better health outcome through interactions between patients, care providers and information and communication technology. Health Serv Manage Res 2016 Apr 04;29(1-2):10-15. [doi: 10.1177/0951484816637746]
63. Nunes F, Fitzpatrick G. Self-Care Technologies and Collaboration. International Journal of Human-Computer Interaction 2015 Jul 30;31(12):869-881. [doi: 10.1080/10447318.2015.1067498]

Abbreviations

eHealth: electronic health
ICD: implantable cardioverter-defibrillator

Edited by G Eysenbach, S Nelson; submitted 15.08.19; peer-reviewed by YH Liu, B Dinesen, D Amiri, GE Iyawa; comments to author 24.01.20; revised version received 20.03.20; accepted 14.05.20; published 20.07.20.

Please cite as:
Andersen TO, Langstrup H, Lomborg S
Experiences With Wearable Activity Data During Self-Care by Chronic Heart Patients: Qualitative Study
J Med Internet Res 2020;22(7):e15873
URL: https://www.jmir.org/2020/7/e15873
doi:10.2196/15873
PMID:32706663

©Tariq Osman Andersen, Henriette Langstrup, Stine Lomborg. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions and Acceptability of Digital Interventions Among Tuberculosis Patients in Cambodia: Qualitative Study of Video-Based Directly Observed Therapy

Lila Rabinovich[1], MPhil; James Steven Molton[2], MBBS; Wei Tsang Ooi[3], PhD; Nicholas Iain Paton[2], MD; Shelly Batra[4], MD; Joanne Yoong[1], PhD

[1]Center for Economic and Social Research, University of Southern California, Washington, DC, United States

[2]Department of Medicine, National University of Singapore, Singapore, Singapore

[3]Department of Computer Science, National University of Singapore, Singapore, Singapore

[4]Operation ASHA, Mumbai, India

Corresponding Author:
Lila Rabinovich, MPhil
Center for Economic and Social Research
University of Southern California
1090 Vermont Avenue, NW
Washington, DC, 20005
United States
Phone: 1 2138210537
Email: lilarabi@usc.edu

Abstract

Background: Despite the development of effective drugs for treatment, tuberculosis remains one of the leading causes of death from an infectious disease worldwide. One of the greatest challenges to tuberculosis control is patient adherence to treatment. Recent research has shown that video-based directly observed therapy is a feasible and effective approach to supporting treatment adherence in high-income settings. However, few studies have explored the potential for such a solution in a low- or middle-income country setting. Globally, these countries' rapidly rising rate of mobile penetration suggests that the potential for translation of these results may be high.

Objective: We sought to examine patient perceptions related to the use of mobile health, and specifically video-based directly observed therapy, in a previously unstudied patient demographic: patients with tuberculosis in a low-income country setting (Cambodia).

Methods: We conducted a cross-sectional qualitative study in urban and periurban areas in Cambodia, consisting of 6 focus groups with tuberculosis patients who were receiving treatment (standard directly observed therapy) through a nongovernmental organization.

Results: Familiarity with mobile technology and apps was widespread in this population, and overall willingness to consider a mobile app for video-based directly observed therapy was high. However, we identified potential challenges. First, patients very much valued their frequent in-person interactions with their health care provider, which may be reduced with the video-based directly observed therapy intervention. Second, there may be technical issues to address, including how to make the app suitable for illiterate participants.

Conclusions: While video-based directly observed therapy is a promising technology, even in country settings where mobile penetration is reportedly almost universal, it should be introduced with caution. However, the results were generally promising and yielded important insights that not only will be translated into the further adaptation of key features of video-based directly observed therapy for tuberculosis patients in Cambodia, but also can inform the future design and successful implementation of video-based directly observed therapy interventions in low- and middle-income settings more generally.

(J Med Internet Res 2020;22(7):e16856) doi:10.2196/16856

XSL•FO
RenderX

KEYWORDS

directly observed therapy; video recording; telemedicine; mobile health; mHealth; tuberculosis; low-income settings; developing countries; patient acceptance of health care; patient acceptability; Cambodia

Introduction

Background

Despite the development of effective drugs for treatment, tuberculosis (TB) remains one of the leading causes of death from an infectious disease worldwide [1]. One of the greatest challenges to TB control is patient adherence to treatment: the need to take drugs daily for 6 months often proves too burdensome, leading to late initiation (treatment delay), disruption (missed treatment), or early termination (treatment default). Nonadherence prolongs the period of infectivity, potentiates relapse, and contributes to the emergence of multidrug-resistant TB cases every year [2].

Conventional directly observed therapy (DOT; the standard TB care whereby patients are strictly monitored daily over the course of their treatment, either by a health care worker or by trained community or family members [3]) faces significant implementation challenges. For patients and their households, adherence entails continuing nonpecuniary costs even when TB drugs themselves are free, including physical side effects, foregone productive time, travel costs, social stigma, and potential affront to autonomy [3]. Reliance on in-person interactions for oversight and support, particularly in resource-poor settings where patients are geographically isolated and hard to reach, is another persistent difficulty.

To address these challenges, numerous adherence-enhancement strategies based on DOT have been tried. Interventions that aim to enhance adherence (eg, reminder systems, defaulter action, education, and peer assistance) without addressing the underlying incentive problem have either shown mixed results or suggested that more high-quality context-specific evidence is needed [4,5]. Studies of financial incentives, however, have shown that patients receiving material rewards are significantly more likely to complete TB treatment [6,7], although these findings were mostly obtained in developed country settings.

With the spread of mobile technology, TB stakeholders are also increasingly eager to explore mobile health (mHealth) solutions to overcoming these constraints [8,9]. The simplest and most widely available approach is the use of text messages or apps to provide support or reminders. However, in the case of TB, use of text message reminders in low- and middle-income country (LMIC) populations has had limited effectiveness [10,11]. More sophisticated interventions, such as smart pills and electronic pill boxes, have been explored and some have shown promise, but practical apps suited to public TB control programs in LMICs are still limited [12-18].

Video-based DOT (VDOT) apps are another mHealth solution that can be integrated into a model of care. By leveraging the ability to take and share videos, VDOT platforms offer a wide range of intervention approaches beyond reminders alone, enhancing monitoring and support by allowing patients to record and send videos of themselves taking medication and enabling new modes of communication with health workers. If successful, such apps would reduce ongoing labor and transport costs by eliminating the need for health workers to travel to patients' homes and allowing patients to take the drugs daily when convenient, further reducing patient opportunity costs. VDOT has already been shown to be effective in increasing adherence in some high-income country settings [19-25]. However, few studies have explored the potential for such a solution in an LMIC setting; where studies exist (eg, in India, Vietnam, Kenya, and Uganda) they show promise terms of feasibility and acceptability [15,26-29]. Globally, these countries' rapidly rising rate of mobile penetration suggests that the potential for translation of these results may be high, but at the same time, the speed of such development may result in new and unanticipated challenges.

Objective

In this context, we sought to examine patient perceptions of issues relevant to the use of mHealth, specifically perceived barriers to treatment adherence, current experience with mobile phone features, and acceptability of VDOT and mobile cash transfers, in a previously unstudied patient demographic: TB patients in a low-income country setting undergoing rapid expansion of mobile penetration. Ultimately, the study aimed specifically to inform the future implementation of an mHealth platform to support TB treatment adherence in this population in Cambodia and, more generally, to highlight considerations when planning and developing such an intervention.

Methods

Setting

Our study location, Cambodia, presents a unique opportunity, with a high prevalence of TB, declining TB treatment completion, and increased mobile phone penetration. Approximately two-thirds of all Cambodians carry the TB bacterium, one of the highest rates in the world. DOT can be a challenge for the 80% of the population that lives in rural communities with poor roads and difficult access, served by a mix of health care workers and DOT volunteers. Although Cambodia has maintained outstanding national average treatment success rates of over 90% for well over a decade, disturbing overall trends and subnational disparities have emerged. Since 2009, treatment success and completion rates have declined steadily. Rising rates of drug-resistant TB also indicate that poor adherence may be a concern [30].

Regarding mobile phone penetration, in 2019 there were more mobile phones than people in Cambodia, with more than half the population owning smartphones [7]. This is an increase of 30% over the previous year, a trend that is likely to expand in the lower-income segments of the population, supported by continued declines in prices, as well as the recent introduction of Khmer-language support and system translation for Android version 4.4 and higher (previously available on Nokia and

XSL•FO
RenderX

iPhone only). These conditions suggest a large and growing opportunity for mHealth to improve monitoring and enhancement.

Study Design

We conducted a cross-sectional qualitative study in urban and periurban areas of Cambodia, with the support of Operation ASHA, a nongovernmental organization providing TB treatment and other health services to low-income communities in Cambodia and other Asian countries. Operation ASHA's community health workers visit patients in their homes every day during the first (intensive) phase of treatment and once weekly during the second (continuous) phase, with patients self-administering the drug the rest of the week.

We conducted 6 focus groups with TB patients being served by Operation ASHA: 3 groups in Phnom Penh (an urban setting) and 3 in Takeo (a semirural province). We determined the number of focus groups using a rule of thumb typically cited in the qualitative literature, namely, that 2 to 5 groups are required per category of participants. Because we had 2 locations (urban and periurban) and for logistical reasons limiting a greater number of focus groups, we settled on 3 groups per location. We recruited 12 participants per group to ensure a minimum of 8 assuming some attrition; ultimately, each group had between 10 and 12 participants. The 2 areas have similar rates of smear-positive results, at around 9% [31].

Participants

Study participants were men and women aged 18 years and older who had a diagnosis of TB and were currently receiving treatment from Operation ASHA, and who were physically able to travel to Operation ASHA local headquarters. Participants were recruited by Operation ASHA staff using the recruitment script developed by the research team and approved by the institutional review board of the University of Southern California, Los Angeles, CA, USA, which also covers Operation ASHA's Cambodia work by an agreement of deferral for this study. All participants in the study were assured of confidentiality and anonymity and all identifiers have been removed to protect participant confidentiality. Participants provided oral informed consent, including for audiotaping and transcription, both during recruitment and at the start of the discussion. As a token of appreciation, participants received a US $10 cash incentive for participation. Participants who did not or were unwilling to provide informed consent were excluded from the study.

Instruments

A semistructured discussion protocol was designed by the lead author with qualitative research expertise (LR) to collect information on patient perceptions of adherence and mHealth interventions in general, as well as to inform the development of a mobile VDOT intervention to support TB treatment adherence. Issues for discussion included perceived barriers to treatment adherence, views on mHealth, and the acceptability of a potential VDOT approach to treatment. The protocol further focused on general issues related to basic features of a VDOT system: (1) sending and receiving text messages, (2) taking and sending videos, and (3) providing incentives via cash or mobile payments as incentives for treatment adherence. We pilot tested the instrument in the first focus group, after which we made minor amendments (primarily clarifying certain questions) for the subsequent groups.

Participants were also asked to assess an existing VDOT app developed by members of the study team (NIJP, JSM, WTO) and asked to evaluate the usability of such an app [32,33]. To avoid priming participants, we presented them with a short description of this VDOT mHealth app only after the general discussion about mobile phone ownership and current use. Users of this platform are provided with a personalized pill schedule that is monitored via the VDOT app. Using the VDOT mobile app, users are able to take a video of themselves taking their medication daily and upload the recorded video for a review by their health worker, together with a report of any adverse effects from the day before. Other features include a color-coded calendar indicating video submission and approval status by day, a function for explaining reasons for missing entries, and a messaging function for communication with the health worker. We also showed participants printouts of what the various screens of the app would look like. We then solicited impressions of the app, including their interest in such an approach, and concerns about its introduction.

We designed a short survey to collect background information prior to the focus group discussion. The questionnaire included questions on demographic and socioeconomic status to provide a complete picture of the background characteristics of our sample. All instruments and the screenshots of the VDOT app were prepared in English, translated into Khmer by bilingual translators, reviewed by an experienced local physician, and further refined where required.

Data Collection

We conducted all focus groups in July 2018. The focus groups were held in private meeting rooms within Operation ASHA's health facilities in both study locations and facilitated by a moderator and note taker, both of whom were bilingual native Khmer speakers locally hired and trained in study protocols, facilitation methods, and data capture by 2 study leads (JY and LR). All groups were facilitated in Khmer with only the moderator, note taker, and 1 researcher from the University of Southern California present during the discussion. Discussions lasted approximately 90 minutes and were recorded with permission, transcribed verbatim, and translated into English for analysis.

Coding and Analysis

We conducted a thematic analysis, in which we coded transcripts using a deductive approach, with codes generated from a review of the transcripts and the study's research questions and aims. This approach aimed to uncover the key themes and the most significant messages inherent in the raw data [34]. Where needed, the resulting codes were organized into broader, overarching categories. All transcripts were coded by the lead author (LR) using the final version of the codebook. We used the qualitative data analysis software ATLAS.ti version 8.4.4 (1135) (ATLAS.ti Scientific Software Development GmbH) to support the organization, review, and coding of the raw data.

XSL•FO
RenderX

Results

Sample Characteristics

We recruited 64 adults receiving TB treatment from Operation ASHA into our study. The sample comprised 34 men and 30 women with a median age of 55 years. The majority (n=44) were married, and 10 were widowed. This is broadly in line with available statistics on TB patients: in the most recently available Cambodian national TB prevalence survey, those aged 45 years or older accounted for 75% of smear-positive TB and 63% of smear-negative, culture-positive TB cases identified, with more males than females having smear-positive and smear-negative, culture-positive TB [35]. Educational attainment was generally low, with 19% (12/64) having no education, 53% (34/64) having an elementary school education, and the rest (18/64, 28%) having either lower or upper secondary education. No one in our sample had any tertiary education. While data are limited on the educational background of TB patients, these numbers show that our sample had overall lower educational attainment than the national population: in 2017, 19% of Cambodians aged 25 years and older reported having no or less than primary education, and 6% reported tertiary education [36]. In our own sample, almost half of participants required assistance from the research staff to complete the paper-based prefocus group questionnaire.

In terms of labor force participation, less than half our sample (29/64) were employed full-time, and 31% were not working at all (20/64). The rest were working part-time (8/64), retired (n=3), or students (n=1). Our sample differed from the general population in this respect: in 2018, labor force participation for all adults aged 15 years and older was estimated at 85% [37]. This difference is likely a function of at least both the median age of our sample (55 years) and the fact that they were all TB patients in treatment.

Regarding TB treatment, 61% (39/64) of our participants had begun their TB treatment less than 4 months previously, with the rest (n=25) in treatment for 5 months or longer.

Reported Adherence to TB Treatment

In addition to basic sociodemographic information, the questionnaire asked participants about their experience with TB treatment, and their mobile phone ownership and use. The vast majority of our participants reported never having missed any of their medication (55/64, 86%); this is likely because Operation ASHA makes home visits. However, these home visits are not uniformly performed daily [2]: most reported seeing their health worker as part of their TB treatment only once a week (35/64, 55%) or less (n=7, 11%); the rest (n=22, 35%) reported seeing their health worker twice or more a week.

Our focus group discussions revolved around 3 main themes that would inform the rollout of a VDOT intervention in Cambodia: (1) challenges to TB treatment adherence, (2) familiarity with and likely acceptability of features of such a platform, and (3) receptiveness to an mHealth app to support treatment, including use of appropriate cash incentives. We discuss our findings in detail below and provide translated direct quotations to illustrate participants' views of these different themes.

Challenges to Treatment Adherence

As expected, participants agreed that access to treatment itself had been easy for them because Operation ASHA provides treatment free of charge and through home visits. Across the groups, participants expressed high levels of satisfaction with the existing Operation ASHA model of home visits:

> *I speak truly, there is no difficulty [adhering to treatment]. The medicine, they bring it home; [when it] runs out, they bring [more] home.* [Female, focus group 5]

> *It was not difficult like before, [when] we took the medicine at the doctor's place. Now they bring the medicine to our home.* [Male, focus group 2]

> *[For] some patients [it is] difficult to travel to the health center, so now we [have] home-based service, it's [easier] for patients.* [Female, focus group 3]

Participants tended to highlight the practical benefits of the home visits more than they did the financial benefit of medications for which they did not have to pay out of pocket. Issues of stigma and privacy regarding an individual's TB and treatment status were not reported as significant concerns, even in the context of receiving a health care worker at home. In addition, there was widespread agreement that home visits not only were more convenient, but also enabled the health care provider to provide the encouragement, reassurance, and support that patients needed to adhere to the treatment, especially in the difficult first 2 months:

> *The doctor in charge of TB is very helpful. He is very attentive to patients. When I was on treatment he would come in every day to make sure I wouldn't miss [any dose of] my treatment. I would have missed my treatment if he didn't push me. After I took the medications I felt so nauseated.* [Male, focus group 4]

> *Some [patients] could give up the medicines, so the doctor has to come every day. They travel by motorcycle on difficult road conditions.* [Female, focus group 4]

While most participants agreed that the no-cost, home-based nature of their TB treatment was a critical factor for accessibility and adherence, they nevertheless recognized that, for themselves and for people they knew, there are significant challenges to treatment adherence, even when the medication is provided at home and at no out-of-pocket cost to the patient.

The main challenges participants identified were physical adverse effects, which can include itching and rashes, nausea, vomiting, lack of appetite, abdominal pain, and fatigue. Across the groups, participants told us of many individuals they knew in their communities who had interrupted treatment because of intolerance to adverse effects:

> *[Some people] do not want to be healed. They throw away the medicine which cause them some adverse effects.* [Male, focus group 1]

XSL•FO
RenderX

> *The reason why patients refuse to take medicines is because of adverse effects. They make them nauseated or feel itchy.* [Male, focus group 4]
>
> *[A neighbor] did not take [the medicine]. He said it's difficult and [has a] bad smell.* [Male, focus group 5]

Many also reported that they themselves had experienced difficulties "because there are very terrible adverse effects" (female, focus group 1). Adverse effects, they explained, can be confusing and scary, as in this participant's experience:

> *For me after I took the medicines I developed numb joints and the lower part of my body became numb. I was panicked...* [Male, focus group 4]

Our participants noted that a key difference between nonadherent patients and themselves was that the participants were able to overcome their discomfort with the adverse effects and remain in treatment, as well as their determination to complete their treatment:

> *When the doctor came here he gave me some medications, which were difficult to swallow for the first 2 months. I had itching, couldn't stand their strong odor, and felt nauseated. I had so much discomfort for the first 2 months. It was so horrible that I felt I was going to die. My weight then dropped to 56 kg, down from 60 kg. The smell was so repugnant and even if my children cooked porridge and fried or grilled cured dry beef for me I turned them away. But when I finished the larger tablets [and] I began to take the smaller ones I could eat very well and was able to eat anything including pickled vegetables, and my weight now is 63 kg.* [Male, focus group 4]
>
> *We want to be cured, we have to try to be patient with eating, have to think appropriately for taking medicine to be cured. Do not give [away], do not throw [it] out, [we] take it all, and do not leave it.* [Female, focus group 5]

Another issue participants identified was the pill itself. When asked what may lead a TB patient to interrupt treatment, a few participants agreed that the pills are large and difficult to swallow, which may lead some patients to interrupt treatment very early on. As one participant noted: "Some people complain that the medicine is too big to take" (male, focus group 6). Another one argued that the pill size was especially hard for a particular group: "For elders, [they] cannot swallow" (male, focus group 5). A participant told the group about a neighbor's experience:

> *I saw one neighbor, she is a woman. She said when she takes medicine in the first session, the tablets are big. She said when she took medicine, it's difficult to take it, and [she was] afraid of the medicine. So, she took [it] only for a few days, [then] she gave up, stop taking [the medicine].* [Female, focus group 3]

A few participants recognized that the length of treatment poses challenges for adherence, even as the majority agreed that following the health care professional's advice and continuing treatment exactly as prescribed are the key to recovering from the disease. Some admitted that this was true even for themselves, as in the following example:

> *[W]hen we take [the] medicine, it is difficult and sometime it's too long time to take. Even I sometime do not want to be cured.* [Female, focus group 3]

Another participant commented more generally:

> *[S]ome people take [the medicine] for half the period, and they feel better, then they give up at the half period. They think they are better, they are cured. Actually, this disease, we need time. So, they give up at the half period, they think they are better, they don't think it will be reactive.* [Male, focus group 3]

Some participants also expressed the view that there are people with TB who may be "less likely to believe in medicines; they believe in traditional medicine" (male, focus group 1). For example, speaking of a neighbor who had received treatment but not followed it appropriately, a participant said: "Maybe he does not believe. He does not believe that the medicine could treat well" (male, focus group 6).

Experience With Mobile Phone Features

Most participants (n=41) said they owned a mobile phone, but a significant minority (n=23) reported they did not own one or use a phone owned by a family member. This is well below estimates in 2016 that over 96% of Cambodians aged 18 to 65 years reported owning a mobile phone [7]. A possible explanation for this discrepancy is the higher concentration of older participants in our sample (n=24, 38% individuals ≥60 years old) relative to the national average (estimated at around 7%) [38].

The majority of our participants who owned mobile phones used them for making and receiving phone calls. Other basic functions were far less common. Only 8 people reported using their phones to send or receive text messages, 7 reported using their phones to send or receive money, and only 1 used them to send or receive videos. Nevertheless, while the use of mobile phones to send and receive money was not widespread, many people reported being familiar with Wing (the largest mobile money platform in Cambodia) and other mobile money services when these were brought up in the discussion.

We probed our participants about reasons for not using their phones for money transfers, an element that would allow integration of incentives directly into a mobile app. Those who did not make any mobile cash transfers typically said that it was either because they did not know how to use these apps, or because they had no one to transfer money to or from (or a combination of both these reasons). Among those who had used their mobile phones to send and receive money, the view was that these services were convenient and easy to use:

> *In terms of using phones to transfer money, I think if we know [how to use it], it's easier than going to receive [the money] at the bank or organization. It's easier because the phone is in our hands.* [Male, focus group 6]

XSL•FO
RenderX

Receptiveness to VDOT and Cash Incentives

Overall, participants were receptive to the general concept of a mobile app for VDOT. There was widespread recognition that an app could be convenient, as it would save time for both patients (in particular those who still had to attend health facilities for their treatment) and providers:

> *If we spend time to meet doctor at the health center, it's difficult. It takes a long time. Sometime, some patients do not have time. So, if you introduce us to use this, it's easier and saves time.* [Female, focus group 3]

> *We can alleviate the doctor's workload because he would not have to come every day.* [Female, focus group 4]

At the same time, participants expressed some concerns about this approach. First, some participants worried about the implications of seeing less of their health care provider, since VDOT would replace most of the home visits of the provider. A few individuals across the groups wondered where they would get their medications from if the home visits were discontinued: "If we don't go to meet the doctor, where can we get the medicine, if he does not bring it for us?" (female, focus group 5).

Second, a related question was whom they would ask questions about their health if their face-to-face interactions with their providers were decreased. Two participants, for example, expressed a clear preference for seeing their health care providers in person rather than having them follow their treatment remotely:

> *Seeing the doctor in person is better, as she can explain and advise us more.* [Male, focus group 4]

> *Go to meet doctor is better...because we can talk face to face for longer discussion.* [Male, focus group 5]

When presented with the concrete example of the existing VDOT app, there was widespread concern about the technical aspects of the app. Even though the moderator described the app in some detail using printouts with pictures of the various screens and functions, there remained some apprehension among participants that they would not know how to use the app. Nevertheless, there was also agreement that if participants were taught how to use it on their phones they would be willing to try it.

Third, participants who did not own phones at the time of our focus group wondered how the app would apply to them.

In a few instances, participants worried about illiterate patients and how they would be able to navigate an app that, as they saw in the printouts, includes some text.

We also asked participants about their views of cash transfers proposed as incentives to adherence, which could potentially be a component of VDOT platforms. Most people welcomed the incentive idea, but many did not want to comment on how much the incentive should be or said they did not know what would be suitable. Those who agreed to comment on how much would be appropriate argued that somewhere between US $0.50 and US $1 per intake would be acceptable (although a very small number of participants argued for higher incentives, of US $2 or more).

Nevertheless, while our participants generally welcomed a cash incentive, many said that getting the medicines for free was sufficient incentive and money was unnecessary: "Having medicines is good already, sir. There is no need to have cash prize" (female, focus group 4).

Discussion

Principal Findings

Our findings showed that familiarity with mobile technology and apps was widespread in this population, and overall willingness to consider a mobile app for VDOT was high. The themes that emerged are in line with VDOT acceptability studies in low-income countries including India, Vietnam, Kenya, and Uganda [15,26-29] and broadly consistent with well-accepted frameworks for assessing the potential acceptability of technology to individuals, such as the technology acceptance model or the unified theory of use and acceptance of technology [39,40].

Our results, however, also highlighted some potential challenges that might arise with transitioning from a face-to-face interaction to VDOT and the resulting implications for interventions. First, many patients related their overall levels of satisfaction with and adherence to frequent, reliable, and high-quality in-person interactions and encouragement. This may then manifest in erosion of both patient satisfaction and adherence in the context of an app-only intervention. In addition, the most frequently cited barriers to adherence were adverse effects, which required physician intervention, and dissatisfaction with the treatment regimen in general, rather than forgetfulness or inconvenience (which are typical factors that work in favor of introducing regular VDOT over periodic in-person visits). Any VDOT intervention may thus need to prioritize timely and frequent feedback and encouragement, as well as ensuring regular interactions with health professionals at intervals, in particular during the first (intensive) phase of treatment, to ensure mitigation of any potential physical and mental harms from less frequent in-person contact.

Second, despite good internet network coverage in our selected areas, not all patients in our study owned a phone. However, almost all patients had access to one. In addition, for most individuals, especially in less urban areas, the use of the phone beyond calls was limited, and familiarity with key aspects of a VDOT platform such as texting and videos was low, with only 1 individual using video regularly. With access to a phone being household-based rather than individual, there may be implications for patient privacy and adherence to a regular schedule of use.

Third, given the rapidly expanding penetration of mobile networks across Cambodia, we expected a higher degree of ownership than we found in our focus group sample, with only 41 individuals personally owning a mobile phone. This may limit the accessibility of an app and lead to concerns such as confidentiality and reliability of access, which may not be assured on a borrowed phone. On the other hand, these findings

XSL•FO
RenderX

also suggest that providing phones to participants as part of the intervention may prove a strong incentive to participate. In the focus groups, when the moderator explained that those without phones could be provided with one, enthusiasm and receptiveness to an app significantly increased.

Fourth, the benefits of the app are not likely to accrue to all groups equally. The demographic profile of users may be significant. Younger, more educated, and employed individuals may be most able to use the app and most likely to benefit from a more efficient interaction, whereas older adults may not only find the app more challenging, but also be most likely to experience adverse effects or be averse to losing the in-person interaction. Another issue is how to make the app suitable for illiterate patients. Literacy rates among Cambodian adults have been estimated at around 80%, which makes illiteracy a very significant problem [41]. While the proportion of illiterate TB patients is unclear, it is likely that it will at least mirror, if not exceed, that of the general population. The initial response to the presentation of a specific app platform was relatively subdued among illiterate users. The rollout of such an app may therefore not be universal, and instead would need to take the likely user profile into consideration, with training possibly extended to family members living with or close to the patient, who may act as treatment supporters especially with illiterate or older patients. Future development could focus on designing a language-free interface with visual cues in the form of symbols or icons instead of words.

In spite of these concerns, it is clear from our focus groups that participants were broadly receptive of our proposed innovation, as long as basic safeguards are put in place (in particular phone provision, training, uninterrupted access to the medications, continued access to a health care provider, and suitability for illiterate patients). Moreover, an app was perceived to be convenient not only for themselves; surprisingly, patients felt positive about lessening the burden on providers as well, which is likely to ease a transition to a new platform.

Limitations

First, one limitation of our study was the reliance on a convenience sample. Individuals who are willing to attend focus groups about adherence may be more likely to be generally adherent or positively disposed toward treatment or Operation ASHA in general.

Second, the use of a hypothetical, graphical presentation of a possible mobile app limited our ability to predict real acceptability and use. This may be especially problematic, as our sample demonstrated remarkably low exposure to apps overall, in spite of network coverage and access to smartphones, potentially due to a lack of familiarity or the lack of relevant service provision via apps at the time of the study. While a more in-depth investigation of factors driving this lack of exposure is beyond the scope of our study, it is worth more detailed exploration. We should reiterate, regardless, that participants were enthusiastic about the possibility of using an app as it was presented to them.

Third, some of our findings are context specific. The level of care provided by ASHA with extensive outreach may not be typical of DOT models in many parts of the world. Furthermore, none of our participants had used mobile apps for TB treatment, let alone health, before. These circumstances may limit broader generalizability across other treatment settings and models of care.

Conclusion

Our findings suggest that, while VDOT is a promising technology as long as appropriate safeguards are put in place, even in country settings where mobile penetration is reportedly almost universal, it should be introduced with caution. However, the results are generally promising and yield important insights, which not only will be translated into the further adaptation of key features of the VDOT app for TB patients in Cambodia, but also can inform the future design and successful implementation of VDOT interventions in LMIC settings more generally.

Conflicts of Interest

None declared.

References

1. World Health Organization. Global Tuberculosis Report 2019. Geneva, Switzerland: WHO; 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf [accessed 2020-06-01]
2. World Health Organization. Treatment of Tuberculosis Guidelines. 4th edition. Geneva, Switzerland: WHO; 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/44165/9789241547833_eng.pdf?sequence=1 [accessed 2020-06-01]
3. Dean A, Dias HM, Falzon D, Floyd K, Raviglione M, Weil D, et al. Drug-Resistant TB Surveillance & Response. Supplement: Global Tuberculosis Report 2014. Geneva, Switzerland: World Health Organization; 2014. URL: https://www.who.int/tb/publications/global_report/gtbr14_supplement_web_v3.pdf [accessed 2020-06-01]
4. Volmink J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2007 Oct 17(4):CD003343. [doi: 10.1002/14651858.CD003343.pub3] [Medline: 17943789]
5. Liu Q, Abba K, Alejandria MM, Sinclair D, Balanag VM, Lansang MAD. Reminder systems to improve patient adherence to tuberculosis clinic appointments for diagnosis and treatment. Cochrane Database Syst Rev 2014;11:CD006594 [FREE Full text] [doi: 10.1002/14651858.CD006594.pub3] [Medline: 25403701]

XSL•FO
RenderX

6. Nglazi MD, Bekker L, Wood R, Hussey GD, Wiysonge CS. Mobile phone text messaging for promoting adherence to anti-tuberculosis treatment: a systematic review. BMC Infect Dis 2013 Dec 02;13:566 [FREE Full text] [doi: 10.1186/1471-2334-13-566] [Medline: 24295439]
7. Kanagaraj P. A boom in smartphone sales in the Kingdom. Phnom Penh, Cambodia: CapitalCambodia; 2019 Mar 15. URL: https://capitalcambodia.com/a-boom-in-smartphone-sales-in-the-kingdom/ [accessed 2020-06-01]
8. Haberer JE, Robbins GK, Ybarra M, Monk A, Ragland K, Weiser SD, et al. Real-time electronic adherence monitoring is feasible, comparable to unannounced pill counts, and acceptable. AIDS Behav 2012 Mar;16(2):375-382 [FREE Full text] [doi: 10.1007/s10461-011-9933-y] [Medline: 21448728]
9. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
10. Au-Yeung KY, Moon GD, Robertson TL, Dicarlo LA, Epstein MS, Weis SE, et al. Early clinical experience with networked system for promoting patient self-management. Am J Manag Care 2011;17(7):e277-e287 [FREE Full text] [Medline: 21819175]
11. Belknap R, Weis S, Brookens A, Au-Yeung KY, Moon G, DiCarlo L, et al. Feasibility of an ingestible sensor-based system for monitoring adherence to tuberculosis therapy. PLoS One 2013;8(1):e53373 [FREE Full text] [doi: 10.1371/journal.pone.0053373] [Medline: 23308203]
12. DeMaio J, Schwartz L, Cooley P, Tice A. The application of telemedicine technology to a directly observed therapy program for tuberculosis: a pilot project. Clin Infect Dis 2001 Dec 15;33(12):2082-2084. [doi: 10.1086/324506] [Medline: 11698993]
13. Garfein R, Collins K, Munoz F, Moser K, Cerecer-Callu P, Sullivan M, et al. High tuberculosis treatment adherence obtained using mobile phones for video directly observed therapy: results of a binational pilot study. J Mobile Technol Med 2012 Dec;1(4S):30-30. [doi: 10.7309/jmtm.52]
14. Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, Flick A, et al. Experience with mobile technology among patients with tuberculosis in San Diego, California and Tijuana, Mexico. J Mobile Technol Med 2013 Dec;2(4S):13-13. [doi: 10.7309/jmtm.2.4S.10]
15. Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, et al. Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010 Jul;39(1):78-80. [doi: 10.1016/j.amepre.2010.02.018] [Medline: 20537846]
16. Sagbakken M, Bjune GA, Frich JC. Humiliation or care? A qualitative study of patients' and health professionals' experiences with tuberculosis treatment in Norway. Scand J Caring Sci 2012 Jul;26(2):313-323. [doi: 10.1111/j.1471-6712.2011.00935.x] [Medline: 22043979]
17. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
18. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
19. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]
20. Safren SA, Hendriksen ES, Desousa N, Boswell SL, Mayer KH. Use of an on-line pager system to increase adherence to antiretroviral medications. AIDS Care 2003 Dec;15(6):787-793. [doi: 10.1080/09540120310001618630] [Medline: 14617500]
21. Simoni JM, Huh D, Frick PA, Pearson CR, Andrasik MP, Dunbar PJ, et al. Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: a randomized controlled trial. J Acquir Immune Defic Syndr 2009 Dec 1;52(4):465-473 [FREE Full text] [Medline: 19911481]
22. Bock NN, Sales RM, Rogers T, DeVoe B. A spoonful of sugar...: improving adherence to tuberculosis treatment using financial incentives. Int J Tuberc Lung Dis 2001 Jan;5(1):96-98. [Medline: 11263524]
23. Petry NM, Rash CJ, Byrne S, Ashraf S, White WB. Financial reinforcers for improving medication adherence: findings from a meta-analysis. Am J Med 2012 Sep;125(9):888-896 [FREE Full text] [doi: 10.1016/j.amjmed.2012.01.003] [Medline: 22800876]
24. Tulsky JP, Pilote L, Hahn JA, Zolopa AJ, Burke M, Chesney M, et al. Adherence to isoniazid prophylaxis in the homeless: a randomized controlled trial. Arch Intern Med 2000 Mar 13;160(5):697-702. [doi: 10.1001/archinte.160.5.697] [Medline: 10724056]
25. Lutge EE, Wiysonge CS, Knight SE, Volmink J. Material incentives and enablers in the management of tuberculosis. Cochrane Database Syst Rev 2012 Jan 18;1:CD007952. [doi: 10.1002/14651858.CD007952.pub2] [Medline: 22258976]
26. Nguyen TA, Pham MT, Nguyen TL, Nguyen VN, Pham DC, Nguyen BH, et al. Video directly observed therapy to support adherence with treatment for tuberculosis in Vietnam: a prospective cohort study. Int J Infect Dis 2017 Dec;65:85-89 [FREE Full text] [doi: 10.1016/j.ijid.2017.09.029] [Medline: 29030137]

XSL•FO
RenderX

27. Sekandi JN, Buregyeya E, Zalwango S, Dobbin KK, Atuyambe L, Nakkonde D, et al. Video directly observed therapy for supporting and monitoring adherence to tuberculosis treatment in Uganda: a pilot cohort study. ERJ Open Res 2020 Jan;6(1):1-11 [FREE Full text] [doi: 10.1183/23120541.00175-2019] [Medline: 32280670]
28. Garfein RS, Liu L, Cuevas-Mota J, Collins K, Muñoz F, Catanzaro DG, et al. Tuberculosis treatment monitoring by video directly observed therapy in 5 health districts, California, USA. Emerg Infect Dis 2018 Dec;24(10):1806-1815. [doi: 10.3201/eid2410.180459] [Medline: 30226154]
29. Kumar AA, De Costa A, Das A, Srinivasa GA, D'Souza G, Rodrigues R. Mobile health for tuberculosis management in South India: is video-based directly observed treatment an acceptable alternative? JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11687 [FREE Full text] [doi: 10.2196/11687] [Medline: 30942696]
30. Sam S, Shapiro AE, Sok T, Khann S, So R, Khem S, et al. Initiation, scale-up and outcomes of the Cambodian National MDR-TB programme 2006-2016: hospital and community-based treatment through an NGO-NTP partnership. BMJ Open Respir Res 2018;5(1):e000256 [FREE Full text] [doi: 10.1136/bmjresp-2017-000256] [Medline: 29955361]
31. Morishita F, Furphy VB, Kobayashi M, Nishikiori N, Eang MT, Yadav R. Tuberculosis case-finding in Cambodia: analysis of case notification data, 2000 to 2013. Western Pac Surveill Response J 2015;6(1):15-24 [FREE Full text] [doi: 10.2471/WPSAR.2014.5.4.005] [Medline: 25960919]
32. Pang Y, Molton JS, Ooi WT, Paton NI, He H. Preliminary effects of a mobile interactive supervised therapy intervention on people living with HIV: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e15702 [FREE Full text] [doi: 10.2196/15702] [Medline: 32217500]
33. Molton JS, Pang Y, Wang Z, Qiu B, Wu P, Rahman-Shepherd A, et al. Prospective single-arm interventional pilot study to assess a smartphone-based system for measuring and supporting adherence to medication. BMJ Open 2016 Dec 20;6(12):e014194 [FREE Full text] [doi: 10.1136/bmjopen-2016-014194] [Medline: 27998903]
34. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
35. National Tuberculosis Control Program. Report: Second National Tuberculosis Prevalence Survey Cambodia, 2011. Phnom Penh, Cambodia: Ministry of Health, National Center for Tuberculosis and Leprosy Control; 2011 Dec. URL: http://open_jicareport.jica.go.jp/pdf/12120325.pdf [accessed 2020-06-01]
36. Cambodia Socio-Economic Survey 2017. Phnom Penh, Cambodia: National Institute of Statistics, Ministry of Planning; 2017. URL: http://www.nis.gov.kh/nis/CSES/Final%20Report%20CSES%202017.pdf [accessed 2020-06-01]
37. International Labour Organization. ILOSTAT database. Washington, DC: World Bank Group; 2018. Labor force participation rate, total (% of total population ages 15+) (modeled ILO estimate) - Cambodia URL: https://data.worldbank.org/indicator/SL.TLF.CACT.ZS?end=2018&locations=KH&start=1990&view=chart [accessed 2020-06-01]
38. Demographics of Population Ageing in Cambodia, 2012. Phnom Penh: National Institute of Statistics, Ministry of Planning, Cambodia; 2012 Oct. URL: https://cambodia.unfpa.org/sites/default/files/pub-pdf/Ageing_FinalReport_NIS2012.pdf [accessed 2020-06-01]
39. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
40. Venkatesh V, Morris MG, Davi GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
41. United Nations Educational, Scientific and Cultural Organization, UNESCO Institute for Statistics. Data for the Sustainable Development Goals. Montreal, QC: UNESCO Institute for Statistics; 2020. Cambodia URL: http://uis.unesco.org/en/country/kh [accessed 2020-06-01]

Abbreviations

DOT: directly observed therapy
LMIC: low- and middle-income country
mHealth: mobile health
TB: tuberculosis
VDOT: video-based directly observed therapy

XSL•FO
RenderX

Edited by G Eysenbach; submitted 31.10.19; peer-reviewed by A Hayward, Y Pang, R Subbaraman; comments to author 29.02.20; revised version received 24.04.20; accepted 13.05.20; published 27.07.20.

Please cite as:
Rabinovich L, Molton JS, Ooi WT, Paton NI, Batra S, Yoong J
Perceptions and Acceptability of Digital Interventions Among Tuberculosis Patients in Cambodia: Qualitative Study of Video-Based Directly Observed Therapy
J Med Internet Res 2020;22(7):e16856
URL: https://www.jmir.org/2020/7/e16856
doi:10.2196/16856
PMID:32716309

©Lila Rabinovich, James Steven Molton, Wei Tsang Ooi, Nicholas Iain Paton, Shelly Batra, Joanne Yoong. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis

Diana Ramírez-Cifuentes[1], MSc; Ana Freire[1], PhD; Ricardo Baeza-Yates[1], PhD; Joaquim Puntí[2,3], PhD; Pilar Medina-Bravo[4], PhD; Diego Alejandro Velazquez[5], MSc; Josep Maria Gonfaus[6], PhD; Jordi Gonzàlez[5], PhD

[1]Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

[2]Hospital de Día de Adolescentes, Servicio de Salud Mental, Consorci Corporació Sanitària Parc Taulí, Sabadell, Spain

[3]Departamento de Psicología Clínica y de la Salud, Universitat Autònoma de Barcelona, Barcelona, Spain

[4]Department of Communication, Universitat Pompeu Fabra, Barcelona, Spain

[5]Computer Vision Center, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

[6]Visual Tagging Services, Bellaterra (Barcelona), Spain

Corresponding Author:
Diana Ramírez-Cifuentes, MSc
Department of Information and Communication Technologies
Universitat Pompeu Fabra
Carrer de Tànger, 122-140
Barcelona, 08018
Spain
Phone: 34 643294184
Email: diana.ramirez@upf.edu

Abstract

Background: Suicide risk assessment usually involves an interaction between doctors and patients. However, a significant number of people with mental disorders receive no treatment for their condition due to the limited access to mental health care facilities; the reduced availability of clinicians; the lack of awareness; and stigma, neglect, and discrimination surrounding mental disorders. In contrast, internet access and social media usage have increased significantly, providing experts and patients with a means of communication that may contribute to the development of methods to detect mental health issues among social media users.

Objective: This paper aimed to describe an approach for the suicide risk assessment of Spanish-speaking users on social media. We aimed to explore behavioral, relational, and multimodal data extracted from multiple social platforms and develop machine learning models to detect users at risk.

Methods: We characterized users based on their writings, posting patterns, relations with other users, and images posted. We also evaluated statistical and deep learning approaches to handle multimodal data for the detection of users with signs of suicidal ideation (suicidal ideation risk group). Our methods were evaluated over a dataset of 252 users annotated by clinicians. To evaluate the performance of our models, we distinguished 2 control groups: users who make use of suicide-related vocabulary (focused control group) and generic random users (generic control group).

Results: We identified significant statistical differences between the textual and behavioral attributes of each of the control groups compared with the suicidal ideation risk group. At a 95% CI, when comparing the suicidal ideation risk group and the focused control group, the number of friends (P=.04) and median tweet length (P=.04) were significantly different. The median number of friends for a focused control user (median 578.5) was higher than that for a user at risk (median 372.0). Similarly, the median tweet length was higher for focused control users, with 16 words against 13 words of suicidal ideation risk users. Our findings also show that the combination of textual, visual, relational, and behavioral data outperforms the accuracy of using each modality separately. We defined text-based baseline models based on bag of words and word embeddings, which were outperformed by our models, obtaining an increase in accuracy of up to 8% when distinguishing users at risk from both types of control users.

Conclusions: The types of attributes analyzed are significant for detecting users at risk, and their combination outperforms the results provided by generic, exclusively text-based baseline models. After evaluating the contribution of image-based predictive models, we believe that our results can be improved by enhancing the models based on textual and relational features. These methods can be extended and applied to different use cases related to other mental disorders.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e17758) doi:10.2196/17758

KEYWORDS

social media; mental health; suicidal ideation; risk assessment; machine learning

Introduction

Background

Mental disorders are a serious health issue worldwide. According to the mortality data presented by the World Health Organization, the number of deaths because of suicide is equivalent to a person dying every 40 seconds [1]. Considering that the signs and symptoms of these disorders have been proven to be traceable on social media, scientists have started to work on the development of automated methods to detect signs and symptoms of these conditions [2-5] by addressing the importance of early detection [6,7].

State-of-the-art approaches for the detection of mental disorders in social media involve the use of machine learning techniques mainly based on textual features extracted from the posts [8-11]. These models require the previous acquisition of annotated data, which are extracted from a selected social platform. Considering this as a classification task, which manipulates sensitive data [12], a proper annotation process is required to obtain a reliable dataset. This has become one of the main challenges because diagnosing mental disorders involves the intervention of specialized and experienced clinicians who follow strict screening proceedings [11].

The methods used for analyzing user-generated data related to suicide focus on tracking social networks at a post level, *that is*, a tweet [13,14], or at a user level, *that is*, a sample of their tweets or posts [15]. The latter is more related to risk assessment, as more data from a single user can be explored. In terms of the type of information extracted and explored, state-of-the-art approaches perform an exhaustive analysis of textual information contained in posts. This has been proven to be relevant for screening and risk assessment tasks [16]. The methods applied often consider bag of words (BoW) models, topic models, lexicons, sentiment analysis tools, and readability and syntactical analysis features [5,15,17]. The most recent work with deep learning approaches consists of exploring sequence models encoding vector representations of terms known as word embeddings [3,11,15].

There are a few approaches that analyze features containing behavioral and relational information. Colombo et al [18] examined the connectivity and communication of users with suicidal ideation on Twitter based on the evaluation of retweets, whereas the work of De Choudhury et al [16] on Twitter defines an egocentric social graph for depression detection. However, these studies do not evaluate the performance of relational elements combined with image-based data.

Regarding the use of visual information from posts shared in social networks to address mental disorders, the closest approaches to our goal are methods for personality prediction [19], and a few recent approaches address self-harm, depression, and anxiety [20-22]. We believe that our study is the first image-based approach for suicide risk assessment on social media at the user level.

When we refer to suicide-related collections, to the extent of our knowledge, no previous publications have been reported on the generation of a suicide-related dataset built over Twitter where the annotation process involves expert clinicians labeling tweets according to the presence of suicidal thoughts focused on two types of control groups. Furthermore, our annotation was performed at the user level after having processed and analyzed sequences of posts within a period of a year. This allowed us to explore changes in time, as it has been proven to be useful for behavior analysis [23] and for the evaluation of early risk detection methods for depression [7,24].

Objectives

This study addressed the detection of mental health issues on social media, with suicide risk assessment as our use case. Our main contributions are listed as follows: (1) we defined a methodology to generate a reliable Twitter dataset for suicide risk assessment, which is also the first user-level dataset of this type dedicated to posts in Spanish; (2) we presented a method to obtain a subset of the user tweets related to a specific topic, in this case, *suicidal ideation*; (3) we generated models that explore the impact of not just relational and behavioral factors but also elements identified by specialists during consultations, which have been mapped to social networks; (4) we developed image-based predictive models to detect suicidal ideation; (5) we integrated the previous elements into a method that combines multimodal data to build predictive models that address the detection of mental health issues using cross-platform information (Reddit, Instagram, and Twitter); and (6) we refined the evaluation process of predictive models for mental health issues by considering 2 different types of control groups within the social media context: users with posts that might not use terms related to mental conditions (generic control cases) and users who make use of terms related to mental disorders (focused control group).

Methods

Data Collection

In this section, we described the methodology designed to generate a reliable Twitter dataset for suicide risk assessment. We selected Twitter as our main data source, as it has been proven to be suitable for analyzing mental disorders on social media [3,13,18,25-27], including suicidal ideation. We also highlighted the following aspects that this platform offers for our research: (1) the possibility of selecting posts in multiple languages; (2) the availability of relational and behavioral factors; and (3) the provision of a set of chronologically ordered posts from each user, which allowed us to do a cumulative analysis of the data referring to suicidal ideation.

XSL•FO
RenderX

Combining Reddit and Twitter for Gathering Suicide-Related Data

Before starting the crawling process on Twitter, we elaborated a list of suicide-related sentences to build a first filtering approach to obtain tweets coming from users at risk. In doing so, we started by collecting a sample of 500 titles of posts published in Reddit's Suicide Watch forum [28]. These posts were mostly written by users with suicidal ideations, so their titles can be considered to be suicide-related sentences.

These phrases were then translated to Spanish and reviewed by clinic psychologists. Some sentences were added, discarded, and/or adapted by the clinicians. We kept 454 phrases after the analysis by experts. Then, based on their relevance and link to suicide risk factors [29], a subset of 110 phrases were selected for use as search terms on Twitter. A total of 98,619 tweets containing these phrases were extracted for a year, that is, from December 21, 2017, to December 21, 2018. These tweets corresponded to 81,572 Twitter users, with 9559 users having more than one tweet matched with the search terms. At the same time, for all users, we extracted all their tweets posted within the same search period.

Two-Level Annotation Process

First Level

In total, 2 labeling stages were followed to identify users with suicidal ideation. As our intention was to follow a manual labeling process done by clinicians, we selected a random sample of 1200 users among those who had at least two tweets matching our search phrases. The user names were anonymized, and 3 tags were defined for labeling purposes: (1) *control*—defining users who on their tweets did not seem to manifest suicidal ideations, users who did not refer to their own conditions, and users who were reporting news or opinions regarding suicide; (2) *suicidal ideation risk*—labeling users who, judging by their writings, seemed to present suicidal ideation signs; and (3) *doubtful*—dedicated to cases where psychologists were not sure about labeling users within any of the other categories. At the first labeling stage, a clinician specializing in this topic was asked to classify users within these 3 categories based only on the tweets containing the suicide-related keywords. After the labeling process, 73.8% (885/1200) of users were classified as control cases, 9.6% (115/1200) were classified as suicidal ideation risk cases, and 16.7% (200/1200) fell within the doubtful category.

Second Level: Short Profile Version

A second labeling process was followed for the users tagged as suicidal ideation risk cases. We analyzed more of their collected tweets to confirm their labels. Annotators noticed that there was a high number of tweets that were not related to suicidal ideation and even sometimes no tweets related to suicide were caught in the sample. To address this issue, we developed a classifier at a tweet level to distinguish tweets containing signs of risk from those that were not related at all with suicide. Thus, we could provide the second annotator a summarized version of a user profile, which we call short profile version (SPV), that contained mainly tweets related to suicide and its risk factors.

We built a binary classifier distinguishing 2 classes: (1) *suicidal ideation–related tweet* and (2) *control tweet*. To train the model, we chose as instances for the *suicide tweet* class the tweets of users tagged as *suicidal ideation risk* cases (513 tweets) and 346 phrases evaluated by the clinicians. For the *control tweet* class, we selected an equally proportional set of random tweets related to other topics, using Twitter's Sample Tweets application programing interface (API) [30]. A BoW model with 1 to 5 grams was generated, and after applying principal component analysis and logistic regression analysis, we achieved F1=0.90, precision (Pr)=0.91, and recall (R)=0.89, with a 10-fold cross-validation procedure. This is considered as our *short profile version classifier* (SPVC).

The SPVC was applied to every tweet of the profile of all users labeled as *suicidal ideation risk* and, for each user, we selected the top 15 *suicide-related tweets* with the highest predicted probability values given by the SPVC. We considered these tweets as the sample to be evaluated by 2 additional annotators: a specialized clinician and a nonspecialized annotator. This second annotator was given detailed instructions and information on risk factors related to suicide. The annotators at this stage were asked to classify users into 2 categories: (1) *suicidal ideation risk* or (2) *control*, now having more information about each user. Once the second annotation process was completed, we only retained the positive cases (n=84) on which both annotators agreed, corresponding to 89% (84/94) of all the cases labeled as suicidal ideation by at least one of the annotators in this second stage. Thus, we ensured that the users labeled as *suicidal ideation risk* were classified under the agreement of 3 annotators at different stages.

We defined 2 different control groups with the same size as the *suicidal ideation risk* class:

- *Focused control group*: users writing suicide-related keywords in a non–suicidal ideation risk context, that is, users who trivialize about suicide, news reports, and information regarding the topic; or users who simply manifest their support or opinions to people at risk. Identifying these users is challenging for classification systems but is key in reducing false-positives. These users were chosen at random among the users labeled as control cases during the first annotation process.
- *Generic control group*: a set of Twitter users who might not necessarily use terms related to suicide. These users were selected randomly using the Sample Tweets API [30] and anonymized as had been done for the other groups.

For both the control groups, the second annotation process was followed to discard possible cases of users at risk within these samples.

Combining Multimodal Data for Detecting Suicide Risk

We proposed a method that given the profile of a user: (1) it uses a text-based model, described previously as the SPVC, which selects a subset of relevant tweets related to suicidal ideation. The set of tweets for which the SPVC provides a score over a given threshold is retained in SPV itself; (2) mostly from the outputted SPV, it extracts a set of relational, textual,

XSL•FO
RenderX

behavioral, lexical, statistical, suicidal ideation–related, and image-based features from the content and metadata of the tweets; and (3) it builds and evaluates different predictive models resulting from the combination of these features. Our features are organized into 3 different groups: (1) *BoW or n-grams and word embeddings* as a representation of textual features; (2) a set of features known as *social networks and psychological (SNPSY) features* containing a group of relational, behavioral, lexical, sentiment analysis, and statistical features, in addition to a set of features that attempt to map to the social media context certain signs and symptoms, which are usually considered by clinicians at the time of screening; and (3) an *image-based score*.

Features Definition

Generic Text-Based Features

Bag of Words and N-Grams

These are features that have been used to address similar tasks, such as depression detection and eating disorders screening [7,31]. These models represent terms or sequences of terms (n-grams) based on their frequencies on the documents analyzed. In our case, each user was represented by a document consisting of the concatenation of the text of all their tweets. Afterward, we used the *Scikit-learn* [32] Python library: *TfIdfVectorizer* to generate a *tf.idf* representation of 1 to 5 grams at the word level. A set of Spanish stop words were considered to build this representation [33]. These features are referred to as BoW features in further sections. We also used *ekphrasis* [34] as a text preprocessing tool to replace generic tag elements such as money, phone numbers, digits, hashtags, and emoticons. We also removed the n-grams that appeared in less than 5% of the documents to reduce the feature space. This is done considering that the features' number is given by the size of the vocabulary of all the writings of the users, and in Twitter, we found cases in which terms are misspelled or elongated (ie, *hellooooo* instead of *hello*); therefore, we avoided having multiple representations for the same term assuming that each spelling mistake and elongation is different and less likely to be repeated over most of the documents.

Word Embeddings

Word embeddings are representations of textual terms as vectors of real numbers. Words that are semantically related have a similar representation over the vector space. Most of the recent predictive models dealing with textual features using deep learning techniques make use of word embeddings to represent terms. The sequences of these representations are fed as inputs to train the predictive models. These types of representations have been recently used in state-of-the-art approaches to address suicide risk assessment [3,11]. We made use of word embeddings previously learned over a dataset with 2 million Spanish tweets [35].

Social Networks and Psychological Features

These consist of a group of features based on generic lexicons [36], statistics measured from the users' writings, information of interest for clinicians regarding the behavior of users in time, the users' social network (relational features) [16], and n-grams lexicons, which include terms referring to suicidal ideation or suicide risk factors (we referred to these features as suicide-related lexicon features). Each of these types of features is described in the next subsections.

Behavioral Features

These features are based on the information extracted from the metadata of tweets. Here, we measured the behavior of users based on their activity within certain periods, which are defined at different granularity levels. These features are detailed in Table 1.

The intention of the sleep time tweets ratio (STTR) and the daytime tweets ratio (DTTR) features is to identify the differences between control users and users at risk regarding the periods of the day on which they post. Considering that our data collection is delimited by language but not by location, that the posting time provided for a tweet is in coordinated universal time and not the time of the user location, and that not enough information from our data was found to automatically identify the location of all the users, we defined an approach to address this issue. As explained in Equation 1, a day was divided into 8 fixed time slots of 3 hours each. Afterward, we assumed that an average user had at least around 6 hours of sleep time, and within this 6-hour period, a smaller number of tweets would be created compared with the rest of the day, so we counted the number of tweets (t) created within each 3-hour time slot for all the tweets of the SPV of a user. Next, for each user, we calculated the sum of the number of tweets within each pair of continuous time slots and selected the minimum score obtained by all the pairs. We also assumed that the first and last slots can be continuous. Finally, this value was normalized according to the total number of tweets of the full profile of the user (T). This feature was considered as the STTR:



The DTTR was given by the difference between 1 and the sleep time ratio: that is, DTTR=1−STTR. It is important to recall that for the measurements that refer to a bigger granularity such as weekdays, weekends, and months, the impact of time difference is not as big as for features based on day periods.

Tweets Statistics

This group refers to 5 types of features that correspond to statistical measures calculated from the tweets of users. We considered elements such as the number of tweets created and their length and the number of tweets that were retained for each user at the SPV in relation to the total number of tweets posted. These features are described in Table 2.

Relational Features

These are informative features regarding the relationships and interactions between users. Elements such as the count of retweets and favorites received and given by the users can provide insight on the social support they receive, along with information regarding the number of followers and users followed, as previously considered for depression screening [16]. Table 3 describes the relational features extracted for our evaluation.

XSL•FO
RenderX

Table 1. Description of behavioral features.

Feature	Description	Source
Working week tweets count ratio	Total number of tweets on weekdays (Monday to Friday) normalized by the total amount of tweets	SPV[a] tweets
Weekend tweets count ratio	Total number of tweets on weekend days (Saturday and Sunday) normalized by the total amount of tweets	SPV tweets
Median time between tweets	Median of the time (in seconds) that passes between the publication of each tweet	SPV tweets
Sleep time tweets ratio	Ratio of tweets posted during the inferred sleep period of the user	Full profile tweets
Daytime tweets ratio	Ratio of tweets posted during the period the user is usually awake	Full profile tweets
Normalized tweet count per quarter (4 features)	Number of tweets posted by the user within each quarter of the year, normalized by the total amount of tweets generated by the user during the year	SPV tweets

[a]SPV: short profile version.

Table 2. Description of features based on tweet statistics.

Feature	Description	Source
Suicide-related tweets ratio	Ratio of tweets retained by the SPVC[a] over all the tweets of the full profile	SPV[b] and full profile tweets
Median SPVC score	Median of the scores obtained by the tweets that are part of the SPV after applying the SPVC	SPV tweets
Median tweet length	Median length of all the user tweets (word level)	SPV tweets
Number of SPV tweets	Number of tweets	SPV tweets
Number of user tweets	Number of tweets posted by the user since the creation of the account	Tweet metadata

[a]SPVC: short profile version classifier.

[b]SPV: short profile version.

Table 3. Description of relational features.

Feature	Description	Source
Followers number	Number of followers	Tweet metadata
Friends number	Number of accounts followed by the user	Tweet metadata
Favorites given	Total number of favorites given by the user	Tweet metadata
Median favorites count	Median of the favorites received by the user	SPV[a] tweets
Median retweets count	Median of the retweets received by the user	SPV tweets

[a]SPV: short profile version.

Lexicons and Suicide Risk Factors Vocabulary

The use of lexicons has been proven to be successful for tasks dedicated to screen mental disorders [9]. For our approach, we counted the frequency of words belonging to all the categories of the Linguistic Inquiry and Word Count (LIWC) 2007 Spanish dictionary [36,37] normalized by the size (in number of terms) of the concatenated writings of the users. It is important to recall that LIWC also contains categories that identify syntactical elements such as verbs, nouns, adverbs, and pronouns, among others. To this dictionary, a group of other categories was added containing vocabulary and up to 3-gram phrases that could be mapped to suicide-related terms and risk factors such as suicide methods; terms referring to self-injuries; explicit suicidal ideation references; self-loathing terms; words that might imply disdain, insomnia, and fear; and possible references to previous suicide attempts, experiencing racial or sexual discrimination, eating disorders, substance abuse, bullying, lack of social support, and family and money issues, along with vocabulary that might imply that some sort of discrimination or abuse has been suffered, that someone close has died from suicide, and even vocabulary regarding the lack of spiritual beliefs, as religion is considered to be a protective factor for screening tasks [29]. The terms and phrases selected for these categories were based on manually mapping common terms and phrases seen in a sample of tweets labeled as suicide related during the first labeling process with the assessment of clinicians. These features were calculated using SPV.

Sentiment Analysis

We obtained a score for each tweet in terms of its polarity. For this purpose, we used *senti-py* [38], trained on Spanish texts

XSL•FO
RenderX

from different sources, including Twitter. It is based on a BoW model with an intermediate feature selection process. To obtain a score per user, we calculated the median of the scores of all the tweets from the SPV.

Image-Based Feature

We followed the methodology proposed in the study by Rodriguez et al [39], where a method for inferring the personality under the OCEAN model was presented. In this sense, we created a classifier trained on images extracted from Instagram using a subset of the phrases and keywords used in the data collection process for Twitter. These images were considered suicide related, whereas a set of unrelated images was considered as our control cases. Afterward, this first model was applied to each of the images extracted from the users' tweets of our dataset. To create this model, we used 90,000 images for training and 60,000 images for validation. To obtain a single score per user (*images user score*), the average of the individual scores of the images of each user was considered as the user's aggregated score.

To define the image classifier, we used a convolutional neural network (CNN) [40] because CNNs are especially suited for image data. There are several variants of this type of network, but the most popular ones are based on the residual networks introduced by He et al [41]. They used skip connections between layers that force the gradient to flow directly between convolutional blocks. This makes backpropagation much more effective in deep architectures.

The training process is performed by minimizing the cross-entropy loss function through gradient descent. The problem when training CNNs is that they require many images (often in the range of millions) to successfully extract relevant features for the final classifier. As most of the time datasets are not large, networks are usually trained on a large dataset (eg, ImageNet [42]) and then fine-tuned on the target dataset. Fine-tuning refers to the process of using an already trained model and retraining it to fit a new distribution. This is much faster because the weights in the model are not randomly initialized, and one can often skip the training of the shallower layers, given the fact that they focus on detecting corners and edges, and thus the model converges faster. Note that the fully connected layer that acts as a classifier in the network must be trained from scratch because the target classes between datasets vary.

For our experiments, we used 101-layer ResneXt [43], which is a residual architecture that uses grouped convolution. This particular architecture uses convolution groups of size 32 with a dimensionality of 8 and a fully connected layer at the end that performs the actual classification. The output of the CNN is a vector that holds the scores for each of the classes; in our case, there were only 2 classes. The network was trained on ImageNet, and we fine-tuned it on our Instagram images. We trained for 8 epochs using stochastic gradient descent with warm restarts [44] with a weight decay of 0.001 and a learning rate of 0.0001 using Nesterov with a momentum of 0.99 on 2 GTX 1080 Ti. We used dropout (50%) to avoid overfitting at the training stage.

Classification Tasks

As we wanted to evaluate the change in the performance of models that use 2 different types of control groups, one constituted by users who make use of vocabulary related to suicide (focused control) and another group of users who might not make use of these terms at all (generic control), we created experiments for comparing (1) users at risk versus focused control users (task 1) and (2) users at risk versus generic control users (task 2). These were selected as our 2 supervised predictive tasks. Our instances and their features for the predictive models were previously defined following the process described in the Combining Multimodal Data for Detecting Suicide Risk section.

Baselines

We defined as baselines 2 models exclusively based on generic text representations. These models were generated using the previously extracted features and representations from the users: (1) full profile and (2) from their SPV. The first one is a BoW model trained with 1 to 5 grams, and the second one consists of a deep learning model defined by a CNN architecture that has been proven to be successful for text classification [45] and has been used in a similar task that addresses suicide risk assessment on Reddit users [11].

We adopted the approach of Shing et al [11] to define our user-level instances. Therefore, given a user represented by a set of sequential posts, we concatenated all these posts and represented each post as a concatenation of words, where each word is represented by a vector (word embedding), as described in the *Features Definition* section. As in the study by Coppersmith et al [3] and as it has been proven successful on similar tasks, we used a set of word embeddings previously learned on Twitter [35] to define the starting weights for our embedding layer and performed further fine-tuning to learn over the training set and adapt the representations to the task domain.

We considered the 2 models previously described as state-of-the-art approaches for the creation of generic and exclusively text-based models for the task, as it is one of the purposes of our work to analyze the contribution of the additional feature types defined. We therefore defined 4 baseline models. Baselines 1 and 3 correspond to the BoW model generated over the full profile tweets sample and the SPV, respectively. Baselines 2 and 4 correspond to the deep learning model built over the same data samples (full profile and SPV).

Classifiers

With the intention of evaluating the individual contribution of the types of features defined, along with their combinations toward a classification/detection task, we considered 4 types of classification algorithms and a deep learning model. We evaluated the performance of random forest, multilayer perceptron, logistic regression, and support vector machines as classifiers. For each feature combination approach, models were built for all these classifiers using the *Scikit-learn* [32] library's implementation, with a grid search for the best parameters. We used a CNN architecture for the embedding models.

XSL•FO
RenderX

Approaches for Combining Features

We evaluated several ways of combining our 3 main feature types defined: generic text-based features, SNPSY features, and the image-based feature (*image user score*). As can be seen in Table 4, we first generated individual models using exclusively all the features corresponding to the *BoW model*, the *embedding's model*, and the *SNPSY model*, with features mainly obtained from the users' SPV, as described in Tables 1-3. Afterward, we explored the combination of our different feature types using the BoW model to represent text-based features. Our first approach involves combining the BoW features with the SNPSY features. In this case, given the large number of BoW features and their sparsity, we opted to use the BoW model–predicted probabilities as values for a single feature, denoted as the *BoW outputted feature*, to be added to the SNPSY set of features. This is described in Table 4 as the *BoW+SNPSY model*. Subsequently, we evaluated the combination of the BoW features with the image feature. For this case, we simply added to the BoW set of features the *image user score* as another attribute; this combination is described by the *Images+BoW model*. Afterward, to combine the SNPSY features with the image feature, we used the *image user score* as a new feature in addition to the SNPSY feature set, which is the *Images+SNPSY model*.

Finally, to combine the 3 feature types, we defined 2 approaches. The first approach is an ensemble model where we consider the outputs (predicted probability scores) of the BoW model (*BoW outputted feature)* and SNPSY model (*SNPSY outputted feature*) along with the *image user score*. This approach corresponds to the *Images+BoW+SNPSY model 1* with 3 attributes based on the combination of the 3 independent models with all their features. The second approach consists of using all the features of the SNPSY type as attributes in addition to the *BoW output feature* and the *image user score*, which lead to the definition of *Images+BoW+SNPSY model 2*. It is necessary to recall that the predicted probability scores from the BoW and SNPSY individual models that were used for some of the feature combination approaches at the training stage correspond to the outputs of the classifiers on the test folds during the cross-validation process executed on the training set. This was done to avoid overfitting.

In addition to the combination approaches described, we created 2 other models over which we performed a feature selection procedure over all the feature types. We chose the features with statistically significant differences among the suicide and control groups to evaluate their contribution exclusively to a predictive model. We presented 2 models with features selected based on the *P* values obtained after performing a Mann-Whitney *U* test to compare the samples of each class. This is a feature selection method that has been previously used in medical applications [46]. In addition, we took into account the efficiency of this feature selection approach, given the large feature space considered (Table 4). These models are defined as the *selected features model 1* with the features where $P<.05$, when comparing the suicidal ideation risk and control groups; and the *selected features model 2*, where $P<.001$. The number of features obtained for each model is also given in Table 4.

Table 4. Models and features.

Model	Features	Number of features	
		Task 1	Task 2
BoW[a] model	BoW features generated with the Tf.Idf vectorizer with 1- to 5-gram features	24,645	24,336
Embeddings model	Word embeddings representations as input for a text-based convolutional neural network model	200	200
SNPSY[b] model	SNPSY features=behavioral+relational+tweets statistics+lexicons+suicide risk factors vocabulary+sentiment analysis features	112	112
BoW+SNPSY model	BoW outputted feature+SNPSY features	24,757	24,448
Images+BoW model	Images user score+BoW features	24,646	24,337
Images+SNPSY model	Images user score+SNPSY features	113	113
Images+BoW+SNPSY model 1	Ensemble model=images user score+BoW outputted feature+SNPSY outputted feature	24,758	24,449
Images+BoW+SNPSY model 2	SNPSY features+images user score+BoW outputted feature	114	114
Selected features model 1	Selected features from all the feature types with $P<.05$	5807	14,882
Selected features model 2	Selected features from all the feature types with $P<.001$	522	3250

[a]BoW: bag of words.

[b]SNPSY: social networks and psychological features.

Experimental Setup

Dataset Description

Following the description of the Data Collection section, to evaluate our approach, we selected a sample of 252 users with a total of 1,214,474 tweets and 305,637 images, from which up to 1000 images per user were selected for our experiments. We selected a balanced sample of 84 users presenting signs of suicidal ideation (users at risk), 84 focused control users, and 84 generic control users, who were classified within these groups by clinicians after seeing samples of their posts. Table 5 shows

XSL•FO
RenderX

the statistics regarding the users belonging to each of the defined groups. We can notice that the median tweet length (in words) is lower for users at risk than for the control and generic cases. We also saw that generic control users produce lower amounts of tweets compared with other types of users.

Table 5. Full dataset labeled group statistics.

Description	Suicidal ideation risk group	Focused control group	Generic control group
Number of users	84	84	84
Number of tweets	313,791	766,437	134,246
Median number of tweets per user	2797.5	2984	716
Median tweet length	11	19	14
Number of images	37,801	251,830	16,006

Detection Tasks and Evaluation Framework

We considered 3 different aspects to analyze: (1) the utility of having defined the SPV, as we believed that this would allow us to focus on the topic we are analyzing by getting rid of the noise provided by tweets that make no reference to our subject of interest; (2) the individual and combined contribution of the different aspects we analyzed: textual, relational, behavioral, and image-based information; and (3) the change in the performance of models that use 2 different types of control groups, one constituted by users that make use of vocabulary related to suicide (focused control) and another group of generic users who might not make use of these terms at all.

All posts from the full profile of the user were considered for baseline 1 and 2 models, whereas most of the features for our proposed models and combinations were extracted exclusively using the SPV, except for some elements extracted from the user's tweets metadata and features such as the STTR and DTTR, which required the usage of the posts from the full profile. For each task, 70% of all the instances were retained for training, and the remaining 30% (around 25 users per class) were left for testing purposes as unseen cases. To keep balanced instances from each class, we used stratification for these sets. In addition to these test sets, we also evaluated our best models over a sample of 200 users labeled as doubtful cases. This is done to verify if, as the human annotator, the models are capable of identifying most of these cases as users that are likely to be at risk.

The *PowerTransformer* class from Python's S*cikit-learn* library was used to transform the feature values to a normal distribution-like representation using Yeo-Johnson's [32]. To choose the best classifier, a 10-fold cross-validation process was followed over the training set with all the algorithms to evaluate. Afterward, the ones with the best performance were selected to perform a second 5-fold cross-validation along with a grid search to find the most suitable parameters for the classifier chosen.

We considered the Pr, R, F1 score (F1), accuracy, and area under the receiver operating characteristics curve (AUC-ROC) score denoted as AUC, which was the measure on which we based the parameter optimization of the grid search. The values for Pr, R, F1, and AUC corresponded to the suicidal ideation risk class, as it is our main class of interest. We reported on accuracy to analyze the performance of both classes. The results obtained by certain classifiers such as the CNNs were averaged results of multiple runs because of the randomness they can add.

Results

Statistical Analysis

We performed an analysis of the features extracted to identify significant differences between the samples of users at risk and our control groups. For each feature extracted for the groups analyzed, we conducted an independent 2-sample Mann-Whitney *U* test among the suicidal ideation group of users and the different control groups. We also conducted this test to compare both of our control groups (focused and generic control groups). We performed a nonparametric test considering that our features do not follow a normal distribution and that there was no homogeneity of variance for most of them.

When comparing the suicidal ideation risk and focused control groups at the SNPSY features, we found significant differences with $P<.001$ among the following features: overall ratio, median time between tweets, verbs, verbs conjugated in singular of the first person ("I"+verb), cognitive mechanisms, anxiety-related terms, usage of personal pronouns, usage of the pronoun "I," negations, terms to express feelings, and coursing terms. Regarding suicide-related lexicons, the usage of suicide explicit terms, depression-related terms, self-loathing, substance abuse, self-injuries, and terms expressing lack of social support also presented an important significance ($P<.001$). Regarding the features from the BoW model, after conducting the same test, we found significant differences with $P<.001$ for n-grams such as *I feel*, *sad*, *kill myself*, *cry/crying*, *depression*, *to die*, *horrible*, *anxiety*, *die*, *pills*, among others. Considering all the features used (24,758), a total of 522 features were significant for distinguishing the groups according to these tests with $P<.001$. Table 6 shows the medians and the distributions overlapping index [47] for both groups on a sample of relevant features.

When repeating the independent two-sample Mann-Whitney *U* test to compare the suicidal ideation risk group with the generic control set of users regarding the SNPSY features, among the ones with $P<.001$, we found the median classifier score, the number of tweets generated, and the median time between tweets to be different among both groups (suicidal ideation risk vs generic control). We identified differences in discussion topics such as money and work, about which the generic control users

XSL•FO
RenderX

seem to discuss more, whereas the members of the suicidal ideation risk group use terms more related to health and biological aspects. As in the previous case, the use of self-references was higher in the suicidal ideation risk group. Within the significant n-grams from the BoW model, we found terms such as *feel, to die, songs, someone, cry/crying, anxiety, life, breath, bad, and fear*. This is shown in Table 7, which displays the median value and overlapping index of the distributions of the groups in terms of some of the attributes mentioned. Again, taking into account all the features used (24,449), 3250 were significant for distinguishing between the suicidal ideation risk and the generic control group in terms of this test with $P<.001$.

Regarding other features explored, considering a 95% CI, for *task 1* (suicidal ideation risk vs focused control), the number of friends ($P=.04$) and median tweet length ($P=.04$) were significantly different. For these cases, the median number of friends for a focused control user (578.5) was higher than the median number of friends at risk (372.0). The same was true for the median tweet length, based on the SPV, which was higher for focused control users with 16 words against 13 of the suicidal ideation risk users. In addition, there were significant differences in the STTR ($P=.049$) and weekday count ratio ($P=.01$). Under the same CI, for *task 2*, the weekday count ratio ($P=.001$), the STTR ($P=.004$), along with the number of followers ($P=.05$), and the total amount of favorites given ($P=.006$) showed significant differences. In this sense, generic control users appeared to tweet more on weekdays (Monday to Friday) as well as focused control users, whereas the opposite behavior was found for suicidal ideation risk users. Regarding the median STTR, generic control users obtained an STTR value of 0.02, whereas users at risk obtained an STTR value of 0.04, meaning that users at risk seemed to tweet more at night compared with the generic and focused control users as well.

The image scores were also significantly different according to the test with $P=.002$ for the comparison between the suicidal ideation risk and generic control groups, considering a 95% CI. Curiously, for the comparison of the image scores between the suicidal ideation risk group and the focused control group, the test scores were different, with $P=.05$. This can be explained by the fact that users providing information or news about suicide make use of similar images, which characterize the condition, making it difficult to find a significant difference only judging by pictures. As can be seen in Table 8, for both the control groups and the suicidal ideation risk group, the median image scores were slightly higher for the suicidal ideation risk group.

Finally, to compare our control groups (focused and generic control groups), we performed the same test (Mann-Whitney *U* test) and found significant differences between some of these groups' features (n=181) with $P<.001$. Among these features, we found mainly suicide-related lexicons, such as suicide methods, suicide explicit terms, bullying, discrimination, and substance abuse–related terms. We also found differences ($P<.001$) in other textual, relational, and behavioral attributes, such as the number of tweets, number of friends, number of followers, median favorites and retweet counts, overall ratio, polarity score, median time between tweets, and STTR, among others. These differences confirmed our previous assumptions regarding the differences among the control groups.

Table 6. Medians and Distribution Overlapping Index for some of the attributes with the most significant differences between the Suicidal ideation and Focused control groups.

Attribute	Suicidal ideation median	Focused control median	Overlapping index
Anxiety	10.94	0	0.25
Coursing terms	21.52	7.68	0.43
To die (self-reference)	5.45	0	0.25
I feel	46.25	6.71	0.32
Self-loathing	0.03	0	0.35
Verb I	22.66	12.11	0.41

Table 7. Medians and Overlapping Index for some of the attributes with the most significant differences between the Suicidal ideation and Generic control groups.

Attribute	Suicidal ideation median	Generic control median	Overlapping index
Median classifier score	0.72	0.65	0.46
To die	19.5	0	0.25
Number of user tweets	2076.5	453	0.38
Health	17.19	8.18	0.44
Work	35.46	49.59	0.44
I	41.32	9.60	0.23

XSL•FO
RenderX

Table 8. Medians and Overlapping Index for the images score between the suicidal ideation, focused control and generic control classes.

Attribute	Group	Median value	Overlapping index
Images score	Suicidal ideation	0.24	0.64
	Focused control	0.23	
	Suicidal ideation	0.24	0.52
	Generic control	0.23	

Classification Task Results

In this section, we reported the results of our experiments. Table 9 presents the evaluation measure results for each task on the test sets. We reported the results for the best models, as described in Table 4, along with the baselines.

Table 9. Predictive task results.

Model	Suicidal ideation versus focused control group						Suicidal ideation versus generic control group					
	Pr[a]	R[b]	F1[c]	Ac[d]	AUC[e]	Classifier	Pr	R	F1	Ac	AUC	Classifier
BoW[f] model—full profile (baseline 1)	0.78	0.81	0.79	0.78	0.81	MLP[g]	0.79	0.85	0.81	0.80	0.91	MLP
Embeddings model—full profile (baseline 2)	0.76	0.81	0.79	0.77	0.82	CNN[h]	0.78	0.87	0.82	0.80	0.84	CNN
BoW model—SPV[i] (baseline 3)	0.81	0.85	0.83	0.82	0.85	LR[j]	0.80	0.92[k]	0.86	0.84	0.89	MLP
Embeddings model—SPV (baseline 4)	0.79	0.85	0.82	0.80	0.83	CNN	0.77	0.87	0.82	0.80	0.82	CNN
SNPSY[l] model	0.85	0.85	0.85	0.84	0.86	SVM[m]	0.85	0.88	0.87	0.86	0.94	LR
BoW+SNPSY model	0.82	0.88[k]	0.85	0.84	0.89	RF[n]	0.85	0.88	0.87	0.86	0.94	LR
Images+BoW model	0.79	0.88[k]	0.84	0.82	0.86	MLP	0.82	0.88	0.85	0.84	0.90	LR
Images+SNPSY model	0.88[k]	0.85	0.86[k]	0.86[k]	0.91	SVM	0.88	0.88	0.88[k]	0.88[k]	0.94	LR
Images+BoW+SNPSY model 1	0.85	0.85	0.85	0.83	0.87	LR	0.85	0.92[k]	0.88[k]	0.88[k]	0.92	MLP
Images+BoW+SNPSY model 2	0.88[k]	0.81	0.84	0.84	0.92[k]	SVM	0.85	0.88	0.87	0.86	0.94	LR
Selected features model 1 (P<.05)	0.85	0.85	0.85	0.84	0.90	MLP	0.91[k]	0.77	0.83	0.84	0.94	SVM
Selected features model 2 (P<.001)	0.83	0.77	0.80	0.80	0.92[k]	SVM	0.91[k]	0.81	0.86	0.86	0.95[k]	SVM

[a]Pr: precision.
[b]R: recall.
[c]F1: F1 score.
[d]Ac: accuracy.
[e]AUC: area under the curve.
[f]BoW: bag of words.
[g]MLP: multilayer perceptron.
[h]CNN: convolutional neural network.
[i]SPV: short profile version.
[j]LR: logistic regression.
[k]The best results for each of the evaluation measures.
[l]SNPSY: Social networks and psychological features.
[m]SVM: support vector machine.
[n]RF: random forest.

Short Profile Version Definition Results

As can be seen in Table 9, the definition of the SPVC is successful as the first filter for both the predictive tasks. Indeed, the BoW models trained exclusively on the SPV (baselines 3 and 4) outperformed baselines 1 and 2 for most of the measures on both tasks. For these representations, tweets unrelated to the topic seem to introduce noise, as they generate a bigger feature space. In contrast, setting a high decision threshold for the classifier implies reducing the vocabulary for the BoW model, which might reduce the performance of the model with the test data.

XSL•FO
RenderX

Regarding the CNN embedding models trained exclusively on the SPV, we can see that the model of task 1 obtains slightly better results compared with the baseline 2 model, whereas the results do not differ much for task 2. In general, we observed a better performance with the SPV for BoW models. Therefore, the combinations evaluated take into account these text-based representations (BoW).

It is important to recall that for the focused control cases, after applying the SPVC with a decision threshold over 0.5, 4 users were left without an SPV because any of their tweets obtained a predicted probability over the threshold. Considering that with higher thresholds, more focused control and generic users could be lost for training our next classifier, this is the threshold we kept for our further experiments (0.5). However, these results also showed that using SPVC reduces the number of control users with an SPV as the threshold value rises and that the definition of the SPV is useful for discarding users who do not present tweets similar to those of the users at risk. Initially, we found that focused control users were more easily discarded than generic users. However, this could be explained by the fact that the control users discarded might correspond to informative accounts such as newspapers, which we assumed to make use of certain terms referring to suicide in a way that does not make use of terms that imply a personal reference or opinion; therefore, the first classifier might find it easier to discard. In any case, this is a supposition as we did not have further access to the writings of users after the annotation.

Combining Model Results

Regarding the methods considered for combining the types of features extracted, we can observe that when these types are evaluated independently from each other, each has a good accuracy, with the SNPSY model obtaining the best results. For the combinations reported in Table 4 for the suicidal ideation risk versus focused control groups, we can observe that the models that use the 3 types of features do not significantly improve the results obtained by the SNPSY model. However, for Images+BoW+SNPSY model 2, we can see a 7% and 11% increase in the AUC score compared with baseline 3 and baseline 1, respectively, for the suicidal ideation versus focused control cases. The AUC difference of their ROC curves using the Delong method was P=.04, which is statistically significant, considering a 95% CI.

For task 2, the Images+BoW+SNPSY combination obtained results that improved baseline 1 for the suicidal ideation versus generic control task. For the Images+BoW+SNPSY model 1, we noticed a 4% increase in accuracy compared with baseline 3, and it increased to 8% compared with baseline 1. There was also an increase in the AUC value of up to 4% with the selected features model 2. We also noticed the same measured results between the SNPSY model, the BoW+SNPSY model, and the Images+BoW+SNPSY model 1, implying that we might not improve the performance of the SNPSY model by adding other feature types. In fact, after conducting a Delong test to compare the ROC curves of these models with the baseline 1 model, we could not find significant differences, implying that their performance was not significantly different from the baseline in terms of the AUC measure for this task. However, this also implied that the use of the SNPSY features alone allowed us to have a model with a reduced number of features that perform as well as the BoW model with thousands of features.

Regarding the role of the images, we can see that when they are individually combined either with the BoW features or the SNPSY features, either the F score or the AUC score increases minimally compared with baseline 3. As part of the experiments for this approach, it is necessary to mention that as some image scores were missing for a few users (up to 4 for each task), the approach considered to address this issue was to replace the scores by the mean of all the users except for the model where only a single score for each feature type was considered; for this case, the instances with missing values were removed.

In reference to the models with a set of selected features, we can notice that these models also outperform baselines 1 and 2 in terms of F1, accuracy, and AUC. The selected features model 1 for both tasks outperformed baselines 3 and 4 on F1 and AUC. It should be noted that these models consider a reduced number of features compared with the baseline models, and the Images+BoW model, as they attempted to reduce the overfitting that the usage of thousands of features might imply.

It is important to mention that for the models evaluated; we did not address the definition of specialized decision thresholds for the classifiers. Therefore, for future work, this should be considered because false-negative predictions should be minimized, and a threshold should be defined to maximize the R value of the suicidal ideation class without leaving aside the tradeoff with Pr. Regardless of this, we can see that most of our models outperformed the baselines in terms of the AUC score, which implies that our models are capable of distinguishing between the classes successfully. In this sense, the definition of a better decision threshold can only contribute to improving the performance of our models.

Comparative Results of Tasks

When comparing the results of both tasks, we saw that the results obtained by the models to distinguish between users at risk from generic control users were not that different from those trained over focused control users. However, we noticed always higher levels of certainty for the models trained to compare users at risk and generic control users. This can be observed when comparing the AUC scores, which are always higher for the models of task 2. In fact, for this task, we can see that a high AUC score is already obtained by the baseline models, and it does not improve significantly with other models. This differs from task 1, where the feature combination is relevant for improving the certainty of the models compared with the baseline.

Figure 1 shows the top 10 most correlated features with the class for each task considering the features of Images+BoW+SNPSY model 2. The most correlated features were given by textual elements such as the BoW model scores and lexicons. It is interesting to see that a behavioral feature as the median time between tweets is relevant for task 2. We can also notice that for both tasks, self-references are relevant and that the usage of explicit suicide terms and health-related terms

XSL•FO
RenderX

is relevant for task 2, as generic control users are not characterized by the usage of terms related to suicide.

Referring to the features that were more predictive for the models generated, we considered the random forest's feature importance function, which is based on its measure of impurity. In this sense, we can see how much each feature decreases the impurity. The more a feature decreases the impurity, the more important is the feature. In this case, because random forest uses multiple trees, the impurity decrease from each feature was averaged across all trees to determine the final importance of the variable. The most important features based on this approach, considering the features of Images+BoW+SNPSY model 2 is shown in Figure 2. For this case, we confirmed that for task 2, the usage of terms related to work and health is distinctive for both classes, as mentioned in the *Statistical Analysis* section.

For both approaches, we can see that the image scores do not appear within the features more relevant for the tasks, implying that textual and behavioral features can be more relevant. Regardless of this, the scores given by certain feature combinations showed that the inclusion of the image scores improves minimally the results of these predictive tasks. Some of the images that most activated neurons at the image processing stage involve pictures of people crying and images containing the term suicide written within them.

We also evaluated the *selected features model 2*, as one of our models with the best results for AUC for both tasks, over a sample of 200 users who were initially labeled as *doubtful* cases. We evaluated 2 models, one trained with the data of task 1 (selected features model 2—task 1) and another trained with the data of task 2 (selected features model 2—task 2). For the first model, we predicted 65% of the doubtful cases as positive (risk), whereas for the second model, 73% of the doubtful cases were found to be at risk. This indicates that our models detected signs of suicidal ideation in more than half of the doubtful users, which is in concordance with the criteria of the first annotator.

Finally, we evaluated the *selected features model 2—Task 1* over a test set of suicidal ideation and generic control users to evaluate the performance of this model over users who do not use a suicide-related vocabulary. We obtained the following results: Pr=0.91, R=0.77, F1=0.83, accuracy=0.84, and AUC=0.95. These results showed that the model obtains better results than generic control users in comparison with its performance over focused control users. Similarly, we evaluated *the selected features model 2—task 2* over a test set of suicidal ideation and focused control users obtaining Pr=0.83, R=0.80, F1=0.82, accuracy=0.82, and AUC=0.91. The performance of this model was worse than that of generic control users and was consistent with the fact that distinguishing these 2 cases is much harder.

Figure 1. Features more correlated with the class to predict for both tasks: Suicidal ideation risk vs Focused control (left), and Suicidal ideation risk vs Generic control (right).

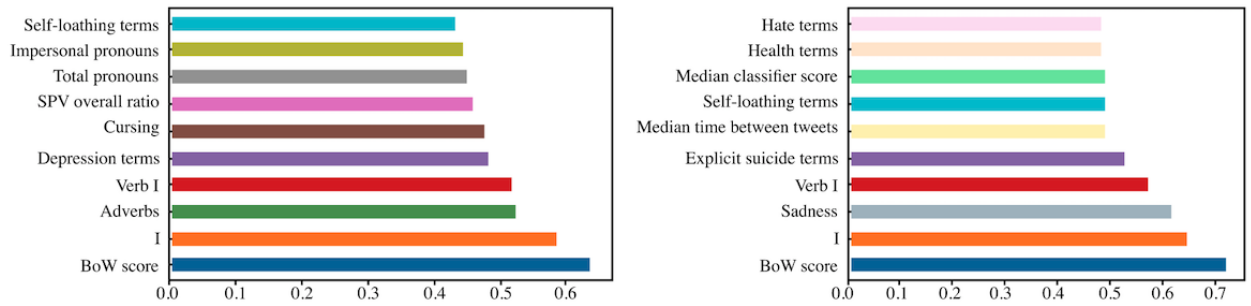
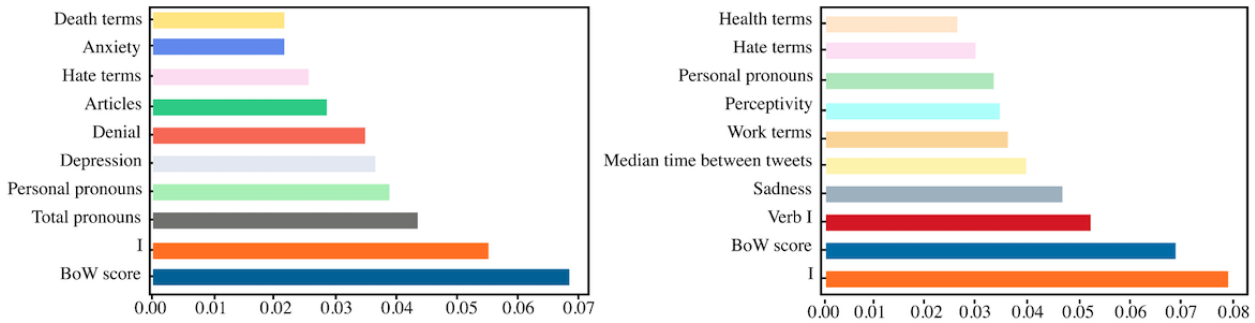


Figure 2. Most predictive features for both tasks: Suicidal ideation risk vs Focused control (left), and Suicidal ideation risk vs Generic control (right).



Discussion

Principal Findings

This study presented a methodology for suicide risk assessment on social media. We extracted information from multimodal data to build statistical and deep learning–based predictive models. Our models consider a set of features based on BoW and n-grams, lexicons, relational, statistical, and behavioral information, in addition to an image analysis. To the best of our knowledge, this is the first approach that addresses the

XSL•FO
RenderX

combination of all these types of features for suicide risk assessment at the user level. Moreover, we highlighted the usefulness of discarding the noise of writings not related to the topic of study through the definition of an SPV, which outperforms the baseline given by the analysis of the full profile of the user, with an increase in accuracy and F1.

We also compared the performance of predictive methods trained on different control groups with the goal of making a more specialized classifier capable of distinguishing users at risk from control cases, even when the discussed topic is similar. Better results are achieved in terms of AUC-ROC when using generic control users instead of users who make use of suicidal vocabulary. We also highlighted the importance of the interpretability of our features, considering elements that can be understood by clinicians and mapped to their screening practice. The results of our experiments showed that within the types of features analyzed, there are multiple significant features that may lead to the detection of risk situations, the most relevant ones are based on the identification of textual and behavioral elements such as self-references, the number of tweets posted, and the time that passes between each post ($P<.001$).

Text-based features were the most relevant for our model; however, their combination with image-based scores, along with relational and behavioral aspects, allowed us to obtain models that outperform the results provided by an exclusively text-based model.

Limitations

Our study presents some limitations, given mainly by the fact that this is an observational study where there is no access to the personal and medical information that is often considered in risk assessment studies. Limitations are given by the accuracy of the methods applied to infer some of these aspects based on the vocabulary and behavioral patterns of the users. For instance, the weekend tweets count ratio is because of the difference in the posting time according to the user's time zone and the sleeping time tweets ratio, for which we assume that the sleeping time is the period in a day when the user has less activity.

A representativeness analysis regarding age, gender, and location of the users analyzed was not performed, considering that this type of information is not available on Twitter. However, the biases existing in our data samples, regarding these aspects, follow similar gender and age biases for suicides [48], as most Twitter users are male and middle aged [14]. In addition to this, the fact that the lexicon used to extract data was obtained from a different source such as Reddit might imply biases introduced by the type of vocabulary used in Reddit, along with the limitations that the translation to Spanish might introduce even if the terms and phrases used were thoughtfully inspected and verified by specialized clinicians.

Finally, there are also limitations given by the nature of the users who post on Twitter, as they might differ from users at risk who do not choose to make their profiles public or even from users at risk who do not have a Twitter account. In addition, it is not guaranteed that the users annotated as users at risk are actually at risk because the annotation was performed just from reading a few tweets.

Reproducibility and Ethical Concerns

The analysis of data provided by social networks to detect health problems and assist clinicians is an open issue, not uncontroversial. The aim of our proposal, however, is to shed light on the real capabilities of these systems in a specific theoretical application: suicide risk prevention. Before such systems become available, a careful risk-benefit assessment along with a proper analysis of applicable legal framework compliance and the potential threats to users' privacy and civil liberties shall be conducted [12,26,49].

On the reproducibility of this work, we should respect Twitter's policies on the distribution of the data collected through its API. Taking into account the restrictions of sharing any information that can lead to identifying the users of our study [12], only the calculated features of our experiments along with the code and parameters used on the classifiers will be available under request with a proper explanation of the usage intended for the models. Finally, it is also necessary to address the potential scenarios of misusage of tools based on these models. Guntuku et al [9] mentioned hypothetical cases where these types of screening tools can be used by employers or insurance companies against the interests of people with mental disorders; therefore, policies against these applications shall be defined before their release.

Future Work

By observing the performance of image-based predictive models, we believe that our results can be improved by enhancing the contribution of the textual and relational features. This can be done by exploring other text representation methods and analyzing deeply the network and interactions between users. In addition, we believe that these methods can be extended and applied to different use cases related to mental disorders, such as depression, anxiety, or eating disorders.

We believe that the approach presented in this study can evolve into a real-time system that emits alerts when users at risk are found. It is necessary to recall that given that scenario, the idea of such a tool is to be the first filter to assist clinicians and does not intend by any means to replace their work. In fact, a readable subset of tweets from the SPV, with the top k tweets ordered according to the SPVC, can be an output of the system for clinicians to evaluate and proceed with future screening steps, if allowed. Indeed, legal and ethical issues for the deployment of such systems should be analyzed before. However, we believe that a tool of this kind can be the starting point for the development of noninvasive interventions where specialists and social media community members can contribute to the recovery and prevention of mental health issues and suicide. Again, the appropriate protocols and procedures should be defined for this instance. In particular, there should be a deep analysis of the implications of a potential intervention taking into account the legal boundaries set for the treatment of personal data in this context.

XSL•FO
RenderX

Acknowledgments

This study was supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Program (MDM-2015-0502). The authors acknowledge the funding received by the Spanish project TIN2015-65464-R (MINECO/FEDER). The authors also acknowledge the support of NVIDIA Corporation for the donation of a Tesla K40 GPU and a GTX TITAN GPU, used for this research.

Conflicts of Interest

None declared.

References

1. World Health Organization. Preventing Suicide: A Global Imperative. Geneva: World Health Organization; 2014.
2. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology. 2015 Presented at: CLPsych'15; June 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1204]
3. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
4. De Choudhury M. Anorexia on Tumblr: A Characterization Study. In: Proceedings 5th International Conference on Digital Health. USA: ACM; 2015 Presented at: DH'15; May 18 - 20, 2015; Florence, Italy p. 43-50. [doi: 10.1145/2750511.2750515]
5. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
6. Losada DE, Crestani F, Parapar J. eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In: Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages. 2017 Presented at: CLEF'17; September 11-14, 2017; Dublin, Ireland p. 343-360. [doi: 10.1007/978-3-319-98932-7_30]
7. Losada DE, Crestani F. A test collection for research on depression and language use. In: Fuhr N, Quaresma P, Gonçalves T, Larsen B, Balog K, Macdonald C, et al, editors. Experimental IR Meets Multilinguality, Multimodality, and Interaction. Cham: Springer; 2016:28-39.
8. De Choudhury M, Counts S, Horvitz E, Hoff A. Characterizing and Predicting Postpartum Depression from Shared Facebook Data. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. USA: ACM; 2014 Presented at: CSCW'14; February 15 - 19, 2014; Maryland, Baltimore, USA p. 626-638. [doi: 10.1145/2531602.2531675]
9. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
10. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
11. Shing HC, Nair S, Zirikly A, Friedenberg M, Daumé III H, Resnik P. Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic.: Association for Computational Linguistics; 2018 Presented at: CLPsych'18; June 5, 2018; New Orleans, LA p. 25-36. [doi: 10.18653/v1/w18-0603]
12. Chancellor S, Birnbaum M, Caine E, Silenzio V, De Choudhury M. A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. USA: ACM; 2019 Presented at: FAT*'19; January 29 - 31, 2019; GA, Atlanta, USA p. 79-88. [doi: 10.1145/3287560.3287587]
13. Abboute A, Boudjeriou Y, Entringer G, Azé J, Bringay S, Poncelet P. Mining Twitter for Suicide Prevention. In: Proceedings of the International Conference on Applications of Natural Language to Data Bases/Information Systems.: Springer International Publishing; 2014 Presented at: NLDB'14; June 18-20, 2014; Montpellier, France p. 250-253. [doi: 10.1007/978-3-319-07983-7_36]
14. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
15. Zirikly A, Resnik P, Uzuner Ö, Hollingshead K. CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts. In: Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology.: Association for Computational Linguistics; 2019 Presented at: CLPsych'19; June 6, 2019; Minneapolis, Minnesota p. 24-33. [doi: 10.18653/v1/w19-3003]
16. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: ICWSM'13; July 8–11, 2013; Boston, Massachusetts p. 128-137.

XSL•FO
RenderX

17. Desmet B, Hoste V. Recognising Suicidal Messages in Dutch Social Media. In: Ninth International Conference on Language Resources and Evaluation. 2014 Presented at: LREC'14; May 26-31, 2014; Reykjavik, Iceland p. 830-835.
18. Colombo GB, Burnap P, Hodorog A, Scourfield J. Analysing the connectivity and communication of suicidal users on Twitter. Comput Commun 2016 Jan 1;73(Pt B):291-300 [FREE Full text] [doi: 10.1016/j.comcom.2015.07.018] [Medline: 26973360]
19. Buraya K, Farseev A, Filchenkov A. Multi-View Personality Profiling Based on Longitudinal Data. In: Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages. 2018 Presented at: CLEF'18; September 10-14, 2018; Avignon, France p. 15-27. [doi: 10.1007/978-3-319-98932-7_2]
20. Scherr S, Arendt F, Frissen T, Oramas J. Detecting Intentional Self-Harm on Instagram: Development, Testing, and Validation of an Automatic Image-Recognition Algorithm to Discover Cutting-Related Posts. Social Science Computer Review 2019 Mar 18. [doi: 10.1177/0894439319836389]
21. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017;6(1):15. [doi: 10.1140/epjds/s13688-017-0110-z]
22. Guntuku C, Preotiuc-Pietro D, Eichstaedt J, Ungar L. What Twitter Profile and Posted Images Reveal about Depression and Anxiety. In: Proceedings of the 2019 International AAAI Conference on Web and Social Media. 2019 Presented at: ICWSM'19; June 12-14, 2019; Munich, Germany p. 236-246.
23. Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. J Med Internet Res 2019 Jun 27;21(6):e14199 [FREE Full text] [doi: 10.2196/14199] [Medline: 31250832]
24. Cacheda F, Fernandez D, Novoa FJ, Carneiro V. Early detection of depression: social network analysis and random forest techniques. J Med Internet Res 2019 Jun 10;21(6):e12554 [FREE Full text] [doi: 10.2196/12554] [Medline: 31199323]
25. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018 Sep 9;2018:1-10. [doi: 10.1155/2018/6157249]
26. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. JMIR Ment Health 2018 Dec 13;5(4):e11483 [FREE Full text] [doi: 10.2196/11483] [Medline: 30545811]
27. Vioules MJ, Moulahi B, Aze J, Bringay S. Detection of suicide-related posts in Twitter data streams. IBM J Res Dev 2018;62(1):7:1-7:12. [doi: 10.1147/jrd.2017.2768678]
28. Reddit. Suicide Watch Subreddit URL: https://www.reddit.com/r/SuicideWatch/ [accessed 2017-12-01]
29. Fowler JC. Suicide risk assessment in clinical practice: pragmatic guidelines for imperfect assessments. Psychotherapy (Chic) 2012 Mar;49(1):81-90. [doi: 10.1037/a0026148] [Medline: 22369082]
30. Twitter Developer. Sample realtime Tweets URL: https://developer.twitter.com/en/docs/tweets/sample-realtime/api-reference/get-statuses-sample [accessed 2019-01-02]
31. Ramírez-Cifuentes D, Mayans M, Freire A. Early Risk Detection of Anorexia on Social Media. In: Proceedings of the 2018 International Conference on Internet Science. In; 2018 Presented at: INSCI'18; October 24-26, 2018; St. Petersburg, Russia p. 3-14. [doi: 10.1007/978-3-030-01437-7_1]
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
33. Savand A. The Python Package Index. 2014. stop-words 2018.7.23 URL: https://pypi.org/project/stop-words/ [accessed 2019-03-05]
34. Baziotis C, Pelekis N, Doulkeridis C. DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation. 2017 Presented at: SemEval'17; August 3-4, 2017; Vancouver, Canada p. 747-754. [doi: 10.18653/v1/s17-2126]
35. Deriu J, Lucchi A, De Luca V, Severyn A, Müller S, Cieliebak M, et al. Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3 - 7, 2017; Perth, Australia p. 1045-1052. [doi: 10.1145/3038912.3052611]
36. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2010;29(1):24-54. [doi: 10.1177/0261927X09351676]
37. Ramírez-Esparza N, Pennebaker J, García F, Suriá R. The Psychology of Words Usage: A Computer Guided Program to Analyze Texts in Spanish. Mexican Psychology Journal 2007;24(1):85-99 [FREE Full text]
38. Hofman E. GitHub. 2018. senti-py URL: https://github.com/aylliote/senti-py [accessed 2019-03-05]
39. Rodriguez P, Gonzàlez J, Gonfaus JM, Roca FX. Integrating vision and language in social networks for identifying visual patterns of personality traits. Int J Soc Sci Humanit 2019;9(1):6-12. [doi: 10.18178/ijssh.2019.v9.981]
40. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278-2324. [doi: 10.1109/5.726791]
41. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016 Presented at: CVPR'16; June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.90]
42. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]

XSL•FO
RenderX

43. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017 Presented at: CVPR'17; July 21-26, 2017; Honolulu, HI p. 1492-1500. [doi: 10.1109/cvpr.2017.634]
44. Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts. In: Proceedings of the 2017 International Conference on Learning Representations. 2017 Presented at: ICLR'17; April 24 - 26, 2017; Toulon, France URL: https://openreview.net/pdf?id=Skq89Scxx
45. Kim Y. Convolutional Neural Networks for Sentence Classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25–29, 2014; Doha, Qatar p. 1746-1751. [doi: 10.3115/v1/d14-1181]
46. Pérez NP, López MA, Silva A, Ramos I. Improving the Mann-Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography. Artif Intell Med 2015 Jan;63(1):19-31. [doi: 10.1016/j.artmed.2014.12.004] [Medline: 25555756]
47. Pastore M, Calcagnì A. Measuring distribution similarities between samples: a distribution-free overlapping index. Front Psychol 2019;10:1089 [FREE Full text] [doi: 10.3389/fpsyg.2019.01089] [Medline: 31231264]
48. World Health Organization. 2019. Suicide in the World: Global Health Estimates URL: https://www.who.int/publications-detail/suicide-in-the-world [accessed 2020-05-27]
49. Zuluaga AF. Ethics and suicidal behaviors. Rev Colomb Psiquiatr 2001;30(4):369-382 [FREE Full text]

Abbreviations

API: application programing interface
AUC: area under the receiver operating characteristics
AUC-ROC: area under the receiver operating characteristics curve
BoW: bag of words
CNN: convolutional neural network
DTTR: daytime tweets ratio
LIWC: Linguistic Inquiry and Word Count
SNPSY: social networks and psychological features
SPV: short profile version
SPVC: short profile version classifier
STTR: sleep time tweets ratio

Edited by G Eysenbach; submitted 15.01.20; peer-reviewed by M Conway, D Reidenberg; comments to author 13.02.20; revised version received 28.03.20; accepted 28.03.20; published 07.07.20.

Please cite as:
Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Puntí J, Medina-Bravo P, Velazquez DA, Gonfaus JM, Gonzàlez J
Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis
J Med Internet Res 2020;22(7):e17758
URL: https://www.jmir.org/2020/7/e17758
doi:10.2196/17758
PMID:32673256

©Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Alejandro Velazquez, Josep Maria Gonfaus, Jordi Gonzàlez. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Instagram-Based Study to Understand Betel Nut Use Culture in Micronesia: Exploratory Content Analysis

Wayne Buente[1], PhD; Francis Dalisay[2], PhD; Pallav Pokhrel[3], PhD; Hanae Kurihara Kramer[1], PhD; Ian Pagano[3], PhD

[1]School of Communications, University of Hawaii at Manoa, Honolulu, HI, United States

[2]Communication and Fine Arts Division, College of Liberal Arts & Social Sciences, University of Guam, Mangilao, Guam

[3]Population Sciences in the Pacific Program (Cancer Prevention in the Pacific), University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, United States

Corresponding Author:
Wayne Buente, PhD
School of Communications
University of Hawaii at Manoa
2550 Campus Road
Crawford Hall
Honolulu, HI, 96822
United States
Phone: 1 8089563360
Email: wbuente@hawaii.edu

Abstract

Background: A 2012 World Health Organization report recognizes betel nut use as an urgent public health threat faced by the Western Pacific region. However, compared with other addictive substances, little is known about how betel nuts are depicted on social media platforms. In particular, image-based social media platforms can be powerful tools for health communication. Studying the content of substance use on visual social media may provide valuable insights into public health interventions.

Objective: This study aimed to explore and document the ways that betel nut is portrayed on the photo-sharing site Instagram. The analysis focuses on the hashtag #pugua, which refers to the local term for betel nut in Guam and other parts of Micronesia.

Methods: An exploratory content analysis of 242 Instagram posts tagged #pugua was conducted based on previous research on substance use and Instagram and betel nut practices in Micronesia. In addition, the study examined the social engagement of betel nut content on the image-based platform.

Results: The study findings revealed content themes referencing the betel nut or betel nut tree, betel nut preparation practices, and the unique social and cultural context surrounding betel nut activity in Guam and Micronesia. In addition, certain practices and cultural themes encouraged social engagement on Instagram.

Conclusions: The findings from this study emphasize the cultural relevance of betel nut use in Micronesia. These findings provide a basis for empirically testing hypotheses related to the etiological roles of cultural identity and pride in shaping betel nut use behavior among Micronesians, particularly youths and young adults. Such research is likely to inform the development of culturally relevant betel nut prevention and cessation programs.

(J Med Internet Res 2020;22(7):e13954) doi:10.2196/13954

KEYWORDS

betel nut; areca catechu; areca; cancer; health; Guam; Micronesia; Instagram; mobile phone; culture

Introduction

Background

Betel or areca nut is the fruit of a palm tree (*Areca catechu*) consumed as a psychostimulant by over 600 million people worldwide [1]. Betel nut chewing stretches from the African continent to the islands of Micronesia and has been prevalent since the dawn of recorded time [2]. Betel nut is enjoyed by both men and women as well as the young and old, often in combination with tobacco. Similar to other psychostimulants, betel nut chewing is known to cause a sense of euphoria, well-being, feelings of warmth, and increased capacity to work [3]. However, there are a number of health issues associated with betel nut consumption, most notably oral cancer [4].

XSL•FO
RenderX

Betel nut use is best understood as an orphan disease in the medical sense. It is consumed by non-Europeans in poor countries and under researched both in the West and in countries where it is endemic [5]. Increasingly, governments and health organizations are viewing betel nut chewing as a preventable threat to public health and a drain on medical resources.

In 2004, the International Agency for Research on Cancer classified the use of betel nut, with or without tobacco, as carcinogenic to human beings [6]. A 2012 World Health Organization report recognizes betel nut use as an urgent public health threat faced by the Western Pacific region, including Micronesia. Micronesia includes US territories such as the Northern Mariana Islands and Guam and island nations in free association with the United States, such as the Republic of Palau, Federated States of Micronesia, and the Republic of the Marshall Islands [7]. Cancer is 1 of the 2 leading causes of mortality in the US-Affiliated Pacific Islands (USAPI) [8,9]. Lung and oral cancer incidence and mortality rates are markedly higher in the USAPI than in the mainland United States [10].

Scholars have begun examining how social media platforms such as Facebook, Twitter, and YouTube contribute to alcohol use [11-13], smoking [14,15], and other health concerns [16,17]. Social media use in public health communication has been characterized as still in a *wild west* phase, where practitioners design and employ untested strategies [18], and the role of image-based social media platforms remains unexplored [19]. Systematic research is required to investigate how social media can be utilized to improve health communication.

As images are powerful tools of health communication [20], Instagram represents a unique social media platform that focuses on visual communication and the everyday ubiquity of smartphones. The use of betel nut in its various forms occurs throughout Southeast Asia and the Pacific Islands, where more than half of the world's mobile subscribers live [21]. A recent study found that betel nut content is trending on Instagram, with evidence of increasing user engagement [22]. Therefore, it is important to explore how betel nut use is depicted on an image-based social media platform such as Instagram.

Use of Social Media to Understand Betel Nut Use

Two theories that have been used to explain the uses and effects of social media include uses and gratifications (U&G theory) [23] and social cognitive theory (SCT) [24]. As originally conceived, the U&G theory suggests that individuals are driven to use media to satisfy certain social and psychological needs, such as, among others, wanting to integrate themselves into society and escape the real world [25]. Recent studies have identified factors that may motivate individuals to use Instagram, including wanting to interact with others, archiving or documenting one's experiences, expressing oneself, escaping, and peeking or surveillance [23,26]. Such motives might also explain why individuals may choose to post betel nut–related content on Instagram.

On the other hand, SCT [27] posits that individuals can learn to adapt behaviors by modeling what they observe in the media. This assumption of SCT is consistent with the findings of recent studies, which indicate that as individuals are exposed to depictions of the use of a particular substance on social media, they become more likely to use the substance [28-30]. Using SCT as a framework may help understand how substance use behaviors are promoted on social media platforms [31]. Similarly, this study seeks to examine distinctions of betel nut–related content on Instagram to see how the social media platform may be used to promote betel nut use.

With 800 million users and 500 million users using the site daily, Instagram is the leading photo-sharing social media platform [32]. The majority (71%) of Instagram users are aged between 18 and 24 years and visit Instagram daily, with 55% using the platform several times a day [33]. As a networked visual social media platform, Instagram's images, videos, and interactive content engagement can have a notable impact on people's knowledge, attitudes, and perceptions of betel nut use. As Instagram is primarily a smartphone app, it is a technology that takes advantage of the portability and habitual use of mobile media [34]. As a result, the omnipresent nature of smartphones could offer a candid examination of betel nut practices within the everyday lives of its users. Furthermore, Instagram images and videos are uploaded and shared directly from a smartphone where other users interact with the content through likes and comments. The power of Instagram likes and comments has health implications, particularly among teenagers and young adults [35,36].

Depicting Substance Use on Instagram

Instagram and similar image-sharing platforms have been the subject of several content analyses related to the promotion and use of addictive substances. For the most part, these studies favored visual over textual content utilizing topical hashtags to capture and document the context of individual use and marketing tactics undertaken on the platforms.

As a plant-based carcinogen, betel nut and betel quid have certain qualities that are similar to the marijuana plant. Instagram posts in marijuana-related hashtags showcased visual images of the marijuana plant [37]. Popular images included marijuana in its traditional forms (ie, buds and leaves) and nontraditional forms (ie, marijuana concentrates and marijuana-infused edibles). In addition, evidence of individuals using marijuana as well as marijuana-related advertisements demonstrated that marijuana content is prevalent on Instagram and helps to normalize and promote its use.

There is also evidence for the presence of hookah (water pipe) tobacco smoking content on visual social media platforms. Tumblr features prominent images of hookah pipes and smoke [38]. To a lesser degree, Tumblr also depicts the presence of waterpipe-related items such as hookah tobacco flavors, electronic hookahs, marijuana, alcohol, and tobacco products. Evidence of social gatherings, particularly between men and women, was also observed. Similarly, Instagram content using the hashtag *#hookah* found images featuring individual waterpipe use, paraphernalia, and social gathering [39]. The presence of promotional material was prominent on Instagram and often cross-promoted alcohol use. In this regard, the marketing power of Instagram [40] may be well-suited to promote waterpipe use and nightlife entertainment.

XSL•FO
RenderX

Recent studies have also documented electronic cigarettes (e-cigarettes) and vaping content on Instagram. In a dataset of 2208 Instagram images, the 3 most prominent themes were advertisements, products, and activity [41]. Advertisement themes explicitly promoted a commercial product and demonstrated the power of Instagram as a marketing tool. Activity themes showcased individuals exhaling aerosols, whereas product themes featured e-cigarettes or electronic juice (e-juice) bottles. An important finding is that likes and comments were found more with activity and product themes than with advertisement-themed images. Therefore, interactivity may prioritize particular Instagram content to larger audiences through social engagement on the platform [42]. A similar study considered e-cigarette visual content on both Instagram and Pinterest. Examining 1800 images from both image-based social media platforms, the popularity of marketing was evident, with 60% of Instagram posts devoted to marketing e-cigarettes. The second most common theme was customization, where users shared information about modifying e-cigarette devices for both functional and aesthetic purposes [43]. Product-related themes (ie, e-juice or flavors) were also evident. As Instagram is a very popular social media platform among teenagers, there is particular concern that Instagram postings may be especially attractive to youth, and that interest in flavors and juices will lure youth and others to try e-cigarettes.

Other health behaviors have been the subject of content analysis on Instagram. For example, a recent study examined the visual content of cigars and cigarillos on Instagram [44]. The findings indicated that Instagram promotes marijuana and tobacco use and promotion. In particular, Swisher products are used to promote the use of marijuana through blunt-making. Instagram images depicted individuals smoking little cigars and cigarillos as well as joints and blunts, though this was a less popular theme.

Social and Cultural Characteristics of Betel Nut in the Western Pacific

Chamorros inhabited the Mariana Islands, an archipelago located in the Western Pacific, when European explorers first arrived in 1521. The Chamorro word for betel nut is *pugua*. Although betel nut is known for its lexical diversity [5], it is the preferred term for betel nut among the various ethnicities of Guam. There are several ways to prepare betel nut for consumption, but 2 distinct practices have emerged in Guam and Micronesia. Betel nut may be chewed with slaked lime (*afok*) and betel pepper leaf (*pupulu*) in the form of a betel quid [45]. This is the most common and traditional preparation throughout the world [46]. Peoples across Micronesia consume betel nut (typically unripened) in this way and spit out the juices as well as masticated quids [47]. However, most Chamorro chewers prefer mature nuts without lime and betel pepper leaves. Moreover, they ingest the juices and the betel nut itself [47]. In fact, Guam is the only island in Micronesia where people chew the mature areca nut [7].

Social and cultural motivations are prominent in the use of betel nut and prove difficult for sustaining health interventions [48,49]. *Pugua* represents Guam's social glue, the bond that builds and maintains social networks among Chamorros and Micronesians [50]. It is shared among friends and family at weddings, anniversaries, and countless other social gatherings. Chewing betel nut is a habit that many Chamorros and Micronesians actively pass on to the next generation because it is a part of their heritage and is viewed as an important cultural identifier [47,51]. As a result, youth are continually exposed to betel nut practices leading to early onset of oral pathologies [46].

Health and the Native Hawaiian and Pacific Islander Populations

As an image-based social media platform with health implications, Instagram intersects with 3 important concerns for the Native Hawaiian and Pacific Islander (NHPI) population. In comparison with other groups, the NHPI population is young and increasing at a higher rate [52]. As teens and young adults increasingly move toward visual social media platforms [53-55], image-based platforms represent an important space for public health interventions that speak to the changing demographics in NHPI communities. Furthermore, adolescents and young adults consume betel nut, and social pressure is an important contributing factor for deciding to chew [56]. Second, few data are available for NHPI on health and wellness; however, there is much need to address inequalities in vital areas of NHPI community life [57]. All too often, NHPI population data are aggregated with other Asian American groups [52]. Finally, the NHPI population faces challenges in maintaining their unique languages and cultures. Previous research has shown that Pacific Islander communities value social media as a way to preserve culture and indigenous knowledge [58,59]. Betel nut use is closely tied to culture in Guam [47,51,60,61] and throughout the USAPI [62,63]. All these denote important aspects for exploring how social media can promote health and behavioral change [64].

The Goal of the Study

This study is an attempt to explore the ways in which betel nut is depicted on the photo-sharing site Instagram. As an image-based social media platform, little is known about what betel nut content is depicted on Instagram. This leads to the first research question:

- Research Question 1: How is betel nut portrayed and represented on Instagram in Micronesia?

Instagram interactivity through user content may shape user perceptions and choices undertaken on the site. For this study, social media engagement takes the form of *liking* and *commenting*. Thus, the second research question is as follows:

- Research Question 2: How do Instagram users engage with Instagram's betel nut images and videos they encounter?

Methods

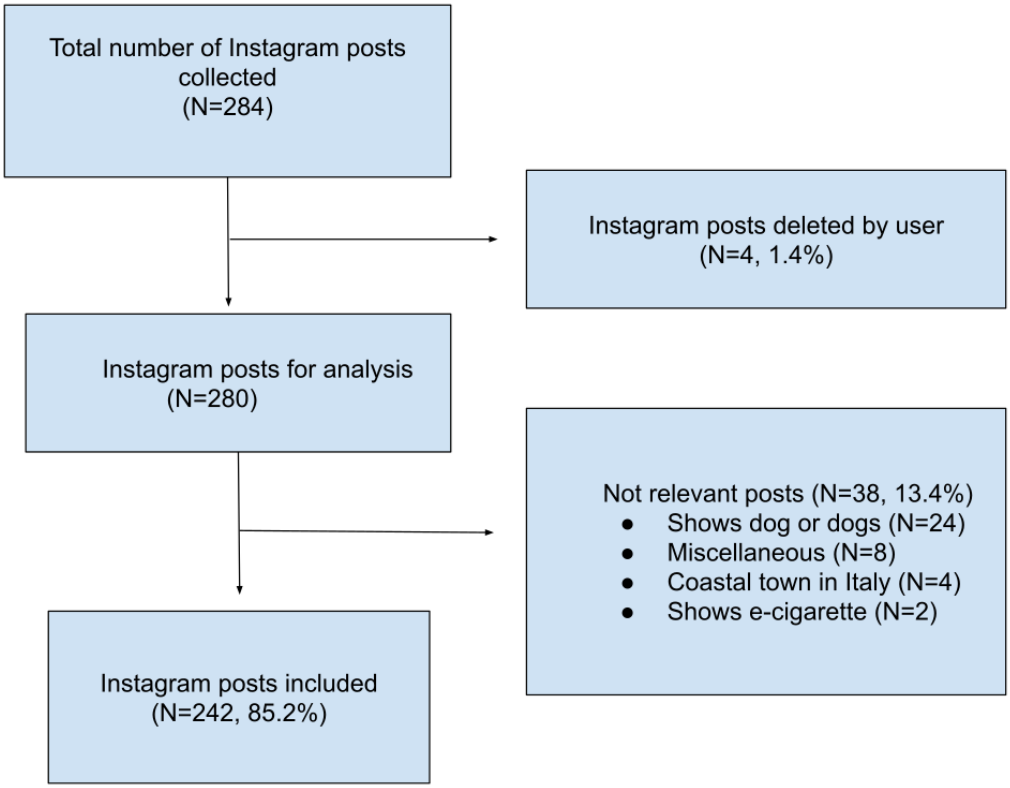
Data Collection and Analysis

To acquire Instagram content, Netlytic was employed as a cloud-based social media analytics tool [65]. Netlytic has been used to collect, analyze, and visualize social media data on a variety of research topics from political discussion on Twitter [66] to health care community practice [67]. On May 22, 2017,

XSL•FO
RenderX

public Instagram posts tagged *#pugua* were collected using the Instagram application programming interface (API) accessed through Netlytic. Data collection consisted of 284 Instagram posts dating back to 2011, uploaded by 180 unique Instagram users. A total of 4 posts were deleted from Instagram 1 month after data collection. All nonrelevant images and videos were also removed, resulting in a final dataset of 242 Instagram posts (234 images and 8 videos, 156 unique users). Figure 1 shows the number of posts identified and reasons for exclusion.

Figure 1. Inclusion of Instagram #pugua posts and reasons for exclusion. E-cigarette: electronic cigarette.



The unit of analysis was an Instagram post (metadata, image, caption, and comments). All images and videos were captured for data analysis. In addition, engagement was measured by adding the number of likes and comments for each post.

Content Analysis

A coding scheme was constructed based on prior content analysis studies on substance use and Instagram and the existing literature on betel nut use [37,39,47]. Images and videos were coded in content categories that addressed the type of betel nut depicted in the image and any paraphernalia present. This provided the category *betel nut type* that concerned the depiction of betel nut tree or leaves, betel nut paraphernalia, and the type of betel nut. The codebook accounts for 2 distinct practices between Chamorro betel nut preparation (mature nut) and non-Chamorro, Micronesian (Chuukese, Palauan, and Yapese) preparation (unripe nut with slaked lime and/or tobacco). Depictions of betel nut use formed an additional category, though later removed in pilot testing. Similar to marijuana, betel nut provides the medium for products and accessories such as jewelry and other consumer goods. Betel nut product advertisements were coded.

Drawing on the strong link between betel nut and Chamorro culture [47], the codebook accounted for the visual and textual depiction of Chamorro culture. In addition, posts that referenced island life in Guam were considered a separate category. Working through the data, *stereotype* accounted for the contentious use of betel nut by Micronesians in Guam [60]. All 3 categories were placed under an umbrella theme of *cultural identifier*.

Prior studies have shown that betel nut use has notable social characteristics, such as encouraging social acceptance and social meetings [47,51,68]. For example, a bag of betel nut can promote future social acceptance among peers through respect for tradition and social promotion [51]. In addition, betel nut is often brought as a gift or favor for celebrations and social meetings [47]. Two content categories documented *social acceptance and promotion* and *social gathering*. Social gathering was further divided into 2 subthemes for either *celebration* or *social meeting*.

Betel nut practices intertwine with familial relations. Thus, children and youth are exposed to betel nut at an early age. The initial codebook accounted for the presence of youth (ie, children, adolescents, or young adults) in image or video content. The last content category referenced traditional or independent media coverage of betel nut use.

For coding, images and videos can be placed in multiple content categories. For example, if an image depicted the betel nut tree ripe with betel nut, the image was coded to both *betel nut tree or leaves* and *Yapese style* (if Yapese-style betel nut was shown in the tree). If the image also included the presence of young

XSL•FO
RenderX

children, the image would receive an additional code for *children* under *youth presence*.

To judge the effectiveness of the codebook, 2 student coders assessed a sample (n=60) of the images and videos. A codebook with instructions and examples was created. After the first round of testing, codebook deficiencies led to improved content category definitions and the elimination of inadequate categorizations. For example, the depiction of betel nut use was difficult to detect in Instagram posts as it involves closed-mouth chewing. In addition, there were few perceived instances of betel nut chewing in the dataset, which were subsequently removed from the codebook.

A second round of interrater agreement was undertaken with a different set of more experienced coders. Utilizing their feedback on the coding instrument, a revised codebook was created, eliminating some content categories and improving the clarification of existing content categories. In all, 3 agreement measures demonstrated codebook validity: percentage concordance, Kappa α, and Krippendorff α. Given the multiple content categories and exploratory nature of the study, codebook concepts were considered reliable if percent agreement was ≥85%, and reliability coefficients were 0.50 or higher. A total of 6 primary content categories produced favorable levels of interrater agreement (Multimedia Appendix 1). Subthemes found in the betel nut type, cultural identifier, and youth presence categories yielded acceptable levels of interrater agreement ranging from 0.57 (betel nut tree or leaves, 85% agreement) to 1.00 (cultural stereotype, 100% agreement).

Results

Content Analysis

To answer the first research question, the 242 Instagram posts produced 386 top-level codebook references using NVivo 11. Of these, *betel nut type* was the most frequent primary content category, yielding 209 references or 54.1% (209/386) of the coded content. As shown in Table 1, betel palm tree or leaves and Chamorro, or Yapese-style betel nut, represented the majority of content references (182/209, 87.1%). Many of the Chamorro (56/64, 88%) or Yapese-style (34/48, 71%) betel nut were in a prepared state ready for consumption. Betel nut paraphernalia (mainly betel nut cutters) was displayed in 12.4% (26/209) of the referenced category content.

Table 1. Distribution of betel nut type categories (N=209).

Theme	References, n (%)[a]
Betel nut tree or leaves	70 (33.5)
Chamorro style nut	64 (30.6)
If Chamorro style nut, is it prepared?	56 (26.8)
Yapese style nut	48 (23.0)
If Yapese style nut, is it prepared?	34 (16.3)
Betel nut paraphernalia	26 (12.4)

[a]Percentages do not total to 100% because of the two conditional categories with their listed percentages.

The second most frequent content referenced on #pugua was images and videos addressing various aspects of cultural identification. Cultural identification references accounted for 22.3% (86/386) of the coded data (Table 2). Of these, posts referencing visual and/or textual aspects of Chamorro culture represented more than half (55/86, 64%) of the content category. A total of 23% (20/86) of cultural content displayed elements of island life in Guam. As spitting has become a sign of racial tension on the island, the remaining content (12/86, 14%) depicted stereotypical images of Micronesian use of betel nut. This included images of red spittle on the ground or public *no spitting* signs, which serve as blatant racialized markers of tension between Chamorros and Micronesians.

Table 2. Distribution of cultural identification categories (N=86).

Theme	References, n (%)
Chamorro culture	55 (64)
Island life	20 (23)
Stereotype	12 (14)

Social acceptance and promotion comprised approximately 10.1% (39/386) of the dataset. As shown in Table 3, social acceptance and promotion included a bag or bags of betel nut. Youth presence (N=24) and betel nut product advertising (N=23) represented 6% of the overall content. Youth presence was equally split between children (aged 0-10 years) and adolescents (aged 11-17 years). Some Instagram posts showed children or adolescents handling betel nut in a playful manner. A total of 23 image or video references attempted to use #pugua to sell or market betel nut products, clothing, or accessories. A total of 5 #pugua Instagram posts depicted media coverage referencing documentaries addressing health concerns or the practice of betel nut chewing.

XSL•FO
RenderX

Table 3. Distribution of social acceptance and promotion, youth presence, betel nut products and accessories, and media coverage categories (N=386).

Theme	References, n (%)
Social acceptance and promotion	39 (10.1)
Youth presence	24 (6.2)
Betel nut products and accessories	23 (6.0)
Media coverage	5 (1.3)

Social Engagement in #pugua

The average engagement score (likes + comments) was 43.33 (SD 48.47) and displayed an upward trend (Figure 2).

To determine if particular #pugua content related to social media engagement, engagement scores were classified into 3 categories: low (0-18), medium (19-49), and high (≥50). Matrix coding resulted in 68 low engagement posts, 111 medium engagement posts, and 62 high engagement posts. For each engagement group, we ranked the 3 most frequent categories (Table 4).

Figure 2. Social media engagement on #pugua (N=242).

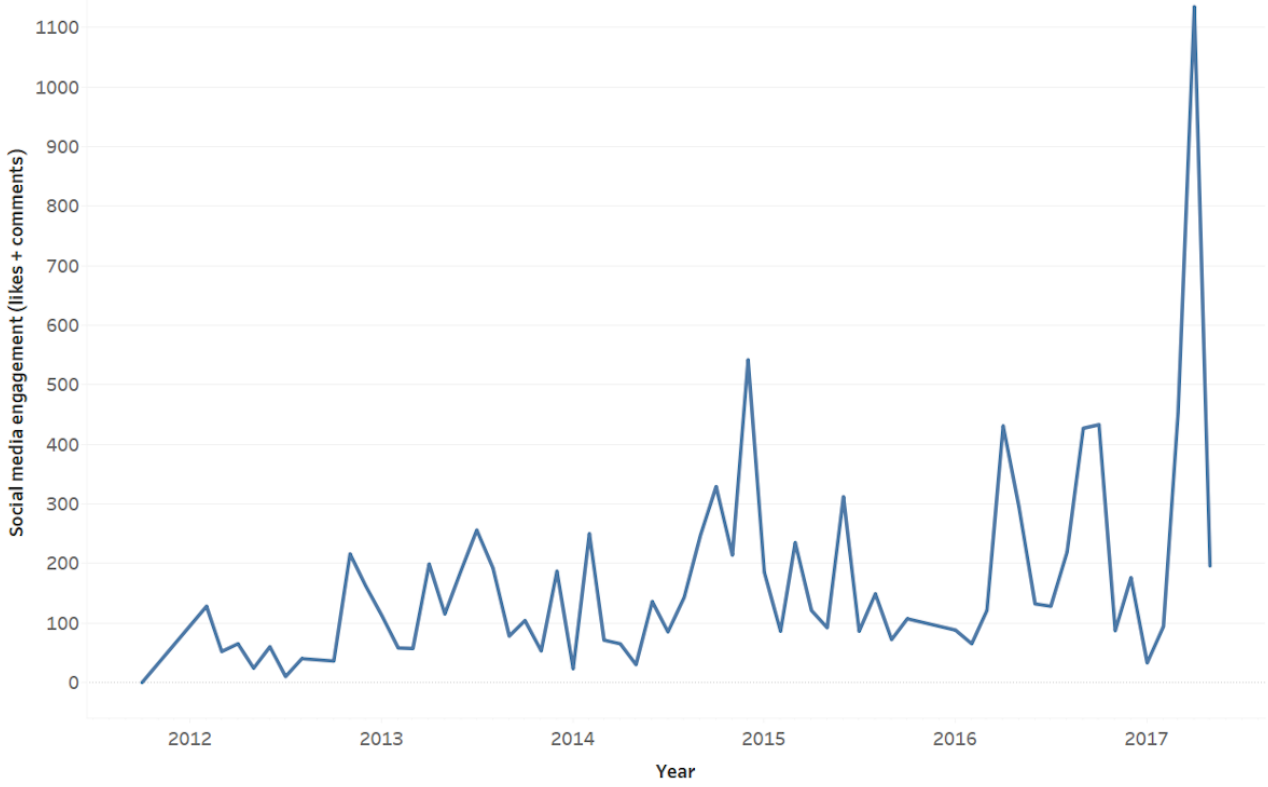


Table 4. Top 3 content categories by level of engagement.

Ranking	Low	Medium	High
1	Chamorro-style betel nut (22)[a]	Betel nut tree or leaves (35)	Chamorro culture (18)
2	Betel nut tree or leaves (20)	Chamorro culture (30)	Chamorro-style betel nut (16)
3	Yapese-style betel nut (17)	Chamorro-style betel nut (27)	Betel nut tree or leaves (15)

[a]Numbers in parentheses indicate the frequency of the content category appearing in the matrix cell.

For the second research question, Instagram posts featuring betel nut tree or leaves (Figure 3) and Chamorro-style betel nut (Figure 4) consistently appear regardless of the level of social engagement. When Chamorro-style betel nut is shown, it is often in a prepared state. At the medium and high engagement level, 95% (41/43) of the Chamorro-style betel nut posts showed betel nut in a Chamorro-prepared style (diced or cut). For low engagement, 68% (15/22) of Chamorro-style betel nut posts were in a prepared state (8/17, 47% for Yapese-style betel nut). For Instagram posts that generated medium and high engagement levels, visual or textual aspects of Chamorro culture were notably present. Highly engaged #pugua posts were more likely to feature Chamorro betel nut preparation or culture compared with other content categories.

XSL•FO
RenderX

Figure 3. Artist's impression of a typical high engagement #pugua image featuring a betel palm tree and mature betel nut. Illustration by Jessica K Sato.

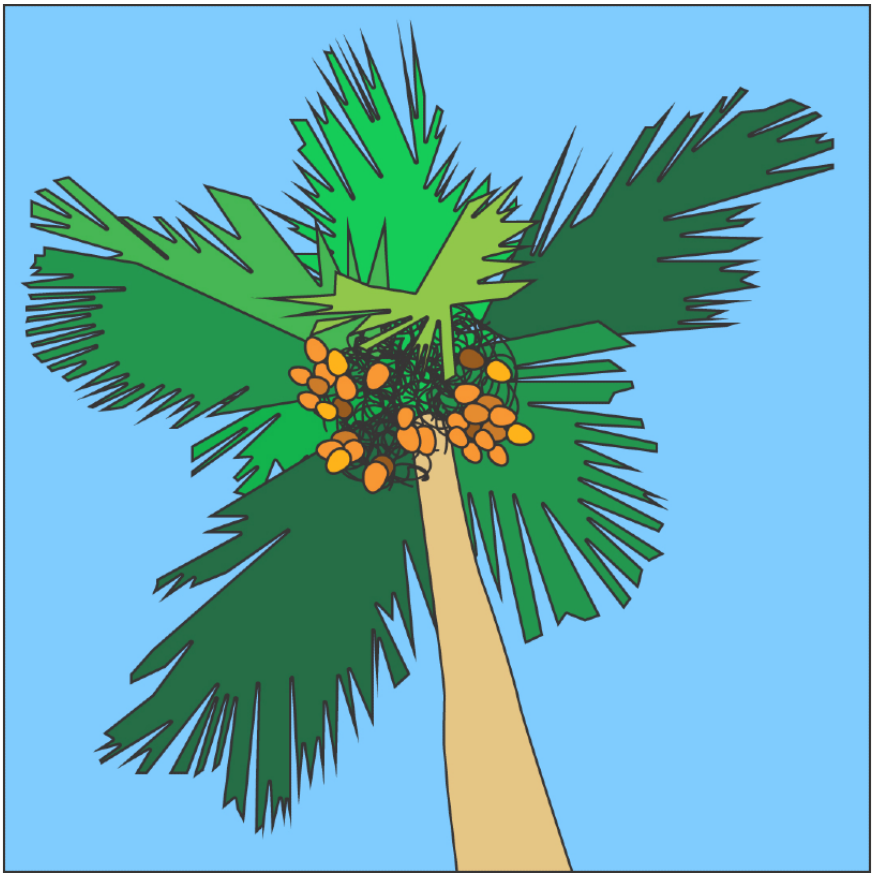
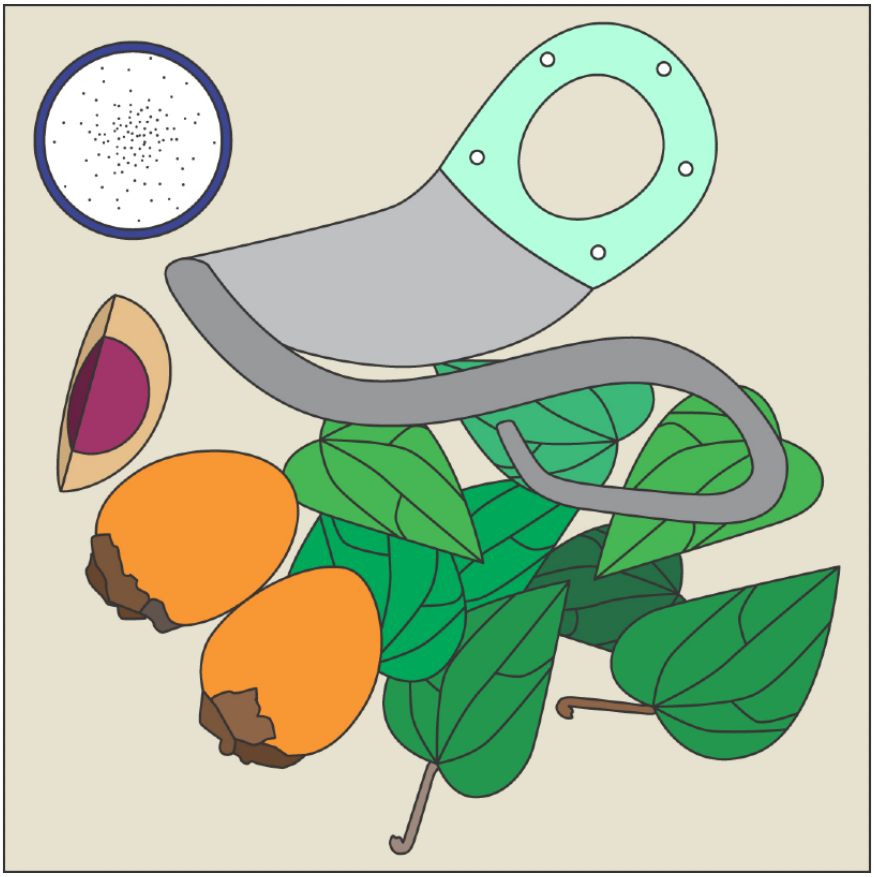


Figure 4. Artist's impression of a typical high engagement #pugua image featuring mature betel nut, betel pepper leaves, slaked lime, and betel nut cutter. Illustration by Jessica K Sato.



XSL•FO
RenderX

Discussion

Principal Findings

As shown in previous studies, betel nut use and practice vary by geographic region [69,70], and Guam is a unique case [47,51,68]. For Guam, betel nut chewing practices that are exhibited offline may replicate themselves on the web within the pugua hashtag. Prior evidence indicated 2 distinct betel nut chewing practices in Guam: Chamorro style and Yapese style [47]. Content analysis confirmed these 2 distinct preparation practices by noting the type of betel nut displayed (mature or unripe), as well as the distinct preparation practices before consumption. The Chamorro style prepares a mature betel nut that is often diced before consumption. In contrast, the Yapese style reflects Micronesian cultural preferences to consume an unripe nut with a piper betel leaf mixed with slaked lime or tobacco to form a betel quid. Aside from Instagram posts that featured betel nut tree or leaves, Chamorro- and Yapese-style betel nut and their respective preparation states were the most frequently occurring content in #pugua. In addition, approximately 10% of Instagram posts showed betel nut cutters that slice betel nut for consumption. As a result, it is plausible that #pugua users on Instagram have opportunities to observe how the betel nut moves from a whole nut to a chewable form. Further studies could determine if these visual depictions of betel nut cultural practices lead to greater uptake in betel nut use, especially among adolescents and young adults.

Approximately 1 out of every 5 #pugua posts referenced Chamorro cultural content either through image/video, text, or both. Pugua has been referenced as the social glue for Chamorro culture. In the exploration of #pugua, there were depictions of Chamorro culture such as dance, art, and festivals without the display of betel nut in the Instagram content. As prior work has demonstrated that Chamorros use social media to learn, capture, and preserve their culture [59], public health interventions should carefully consider how their visual messages link betel nut content on Instagram to Chamorro cultural identification.

#pugua and Social Engagement

As Instagram is a visual social media platform, the circulation of the image through likes and comments may have more value than the image content [71]. The findings from this study determined that certain Instagram #pugua posts had higher social engagement than others. Instagram #pugua posts with moderate or high amounts of social engagement centered around visual content privileging Chamorro culture, Chamorro-style betel nut, and betel nut tree and leaves. Instagram and the smartphone favor *in the moment* sharing often reflecting everyday life instances. As betel nut use is an underexamined research phenomenon, there is a need to understand who, where, when, how, and why people chew betel nut [72]. The exploratory results of this study provide insights into the who, how, and why people in Guam and Micronesia chew betel nut and confirm the existing limited research in this area. Thus, this study represents a first step toward understanding how image-based social media platforms capture and collate information on betel nut practices.

Comparison With Prior Work

As a plant-based carcinogen, the areca nut and betel pepper leaf share certain qualities with the marijuana plant. These psychoactive plants provide facilities for ingesting and supplementing the use of the product. For example, betel nut may be consumed within the betel pepper leaf. Similarly, marijuana buds or leaves may be consumed or form additional use practices. Both plant-based products supply the medium for additional products and accessories. As a result, betel nut practice and use on Instagram may closely resemble the practice and use of marijuana on Instagram. Comparing the study's findings with the large-scale work on marijuana and Instagram [37], there are some similarities, but also notable differences.

As indicated by previous research on marijuana and Instagram [37], the most frequent Instagram posts involved showing the betel palm tree and leaves, mature or unripe betel nut, and its various consumption practices. In this regard, the betel nut and marijuana plant share certain visual characteristics that appeal to their respective audiences on Instagram. Therefore, public health professionals should consider how to balance appealing visual content with helpful information for young people about the known risks of betel nut use.

The consumption and masticatory use of betel nut did not appear to be compelling content on Instagram. Very few Instagram posts featured chewing of betel nut or betel quid. If it was displayed, it was only through a closed mouth where a lump could be witnessed in the person's jaw. Previous studies have shown that betel nut chewing and spitting are bad habits that undermine public health [73]. On a visual communication medium such as Instagram, it is difficult to portray betel nut chewing as novel and interesting to facilitate social networking. This finding is in contrast to the work done on the use of marijuana in social media [37,74]. Previous research has shown that marijuana use on social media facilitates social networking about marijuana, particularly among young adults and underage youth. Ingesting marijuana in its traditional plant-based and novel forms (edibles and concentrates) is perhaps more visually appealing and also helps to normalize and promote marijuana use as a form of practice. Similarly, e-cigarette, hookah, and cigar/cigarillos Instagram content featured activity and use themes such as smoking, exhaling clouds, and smoke tricks [38,39,41,43,44]. The exploratory findings in this study did not observe the use of betel nut on an image-based social media platform.

Advertising and marketing tactics have been well documented for e-cigarette and marijuana use on Instagram [30,37,41]. However, there was little evidence of advertising and product promotion of betel nut in #pugua. This is most likely a reflection of the lack of a clear marketplace that sells specific accessories and products related to its consumption, such as vape shops and marijuana dispensaries. Betel nut products that were advertised tended toward clothing or jewelry accessories made from betel palm tree or pepper leaf fibers. These Instagram posts are often linked to specific references to Chamorro culture. Therefore, it is plausible that #pugua on Instagram functions less as a marketplace and more as a cultural space.

XSL•FO
RenderX

Limitations

There are limitations to this study. The dataset was small and limited the generalizability of the study findings. A paid Netlytic account captured all posts that met the inclusion criteria. Given the size of the dataset, all #pugua posts were analyzed, and there was no sampling required. However, data collection and sampling remain a challenge for capturing image-based social media content because of API restrictions and privacy concerns [19]. Algorithms on Instagram also influence the visibility of the #pugua content [75] and certainly contribute toward its social engagement. Nonetheless, caution should be exercised to avoid generalizing the findings of #pugua to all Instagram users and contexts. Future work should empirically test the study's implications through experiments documenting image content effects on Instagram betel nut users or pursue large-scale data collection of #betelnut and regional betel nut terms as hashtags on social media platforms.

Conclusions

This study analyzed the ways in which betel nut is depicted on the photo-sharing site Instagram. A content analysis of #pugua explored betel nut practices in #pugua and its engagement on the platform. The study findings suggest that betel nut preparation practices reflect offline practices in Guam and Micronesia. In addition, socially engaged content is more likely to reflect the areca nut, betel palm tree or leaves and Chamorro cultural sensibilities. Compared with prior work on substance use and Instagram, #pugua content demonstrated similarities to marijuana in terms of depicting the betel palm tree, areca nut, and quid ingredients. In addition, advertisements of betel nut clothing and accessories were observed. In contrast to prior work, there was little evidence of actual betel nut use and consumption, which is typically a prominent feature in marijuana, hookah, and vaping content on image-based social media.

The findings from this study emphasize the cultural relevance of betel nut use in Micronesia. The findings provide a basis for empirically testing hypotheses related to the etiological roles of cultural identity and pride in shaping betel nut use behavior among Micronesians, particularly youths and young adults. Such research is likely to inform the development of culturally relevant betel nut prevention and cessation programs.

Acknowledgments

The authors would like to acknowledge the U54 University of Guam (National Institutes of Health [NIH] Grant 1U54CA143728)/University of Hawaii Cancer Center Partnership (NIH Grant 1U54CA143727) for their support. The authors would also like to thank Dr Lala Hajibayova, Mallory McCorkhill, Hannah Mettias, and Lauren Muraoka for their valuable assistance with our research project. We are very grateful to freelance designer and strategist Jessica K Sato for her artist illustrations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reliability indicators for codebook categories.
[DOCX File , 18 KB - jmir_v22i7e13954_app1.docx]

References

1. Boucher BJ, Mannan N. Metabolic effects of the consumption of Areca catechu. Addict Biol 2002 Jan;7(1):103-110. [doi: 10.1080/13556210120091464] [Medline: 11900629]
2. Haddock RL, Hoffman JH, Williams WR. Betel nut chewing in Guam. Fiji Med J 1981;9(8/9):139-149.
3. Oliver D, Johnson L. Oceania: The Native Cultures of Australia and the Pacific Islands. Honolulu, HI: University of Hawaii Press; 1989.
4. Hernandez BY, Bordallo RA, Green MD, Haddock RL. Cancer in Guam and Hawaii: a comparison of two US island populations. Cancer Epidemiol 2017 Oct;50(Pt B):199-206 [FREE Full text] [doi: 10.1016/j.canep.2017.08.005] [Medline: 29120826]
5. Osborne PG, Ko Y, Wu M, Lee C. Intoxication and substance use disorder to Areca catechu nut containing betel quid: a review of epidemiological evidence, pharmacological basis and social factors influencing quitting strategies. Drug Alcohol Depend 2017 Oct 1;179:187-197. [doi: 10.1016/j.drugalcdep.2017.06.039] [Medline: 28787696]
6. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Betel-Quid and Areca-Nut Chewing and Some Areca-Nut-Derived Nitrosamines. Lyon, France: World Health Organization; 2004.
7. World Health Organization. Review of Areca (Betel) Nut and Tobacco Use in the Pacific: A Technical Report. Geneva, Switzerland: World Health Organization; 2012.
8. Guam Comprehensive Cancer Control Plan 2013-2017. Centers for Disease Control and Prevention. 2013. URL: ftp://ftp.cdc.gov/pub/Publications/Cancer/ccc/guam_ccc_plan.pdf [accessed 2019-03-01]

XSL•FO
RenderX

9. Cancer in the US Affiliated Pacific Islands 2007-2012. Pacific Cancer Programs. 2015. URL: http://www.pacificcancer.org/site-media/docpdfonwebpage/2015/PIJ%20Cancer%20Facts&%20Figures%202007-2012%20060115.pdf [accessed 2020-06-05]
10. Pacific Regional Comprehensive Cancer Control Plan 2007-2012. Centers for Disease Control and Prevention. 2007. URL: ftp://ftp.cdc.gov/pub/Publications/Cancer/ccc/pacific_regional_ccc_plan_2007_2012.pdf [accessed 2019-03-04]
11. Beullens K, Schepers A. Display of alcohol use on Facebook: a content analysis. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):497-503. [doi: 10.1089/cyber.2013.0044] [Medline: 23617225]
12. Erevik EK, Torsheim T, Vedaa O, Andreassen CS, Pallesen S. Sharing of alcohol-related content on social networking sites: frequency, content, and correlates. J Stud Alcohol Drugs 2017 May;78(4):608-616. [doi: 10.15288/jsad.2017.78.608] [Medline: 28728643]
13. Sampasa-Kanyinga H, Chaput J. Use of social networking sites and alcohol consumption among adolescents. Public Health 2016 Oct;139:88-95. [doi: 10.1016/j.puhe.2016.05.005] [Medline: 27311992]
14. Myslín M, Zhu S, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
15. Phua J. Participating in health issue-specific social networking sites to quit smoking: how does online social interconnectedness influence smoking cessation self-efficacy? J Commun 2013 Sep 16;63(5):933-952. [doi: 10.1111/jcom.12054]
16. Fung IC, Blankenship EB, Goff ME, Mullican LA, Chan KC, Saroha N, et al. Zika-virus-related photo sharing on Pinterest and Instagram. Disaster Med Public Health Prep 2017 Dec;11(6):656-659. [doi: 10.1017/dmp.2017.23] [Medline: 28330514]
17. Seltzer EK, Horst-Martz E, Lu M, Merchant RM. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
18. Ems L, Gonzales AL. Subculture-centered public health communication: a social media strategy. New Media Soc 2016 Jul 10;18(8):1750-1767. [doi: 10.1177/1461444815570294]
19. Fung IC, Blankenship EB, Ahweyevu JO, Cooper LK, Duke CH, Carswell SL, et al. Public health implications of image-based social media: a systematic review of Instagram, Pinterest, Tumblr, and Flickr. Perm J 2020;24:- [FREE Full text] [doi: 10.7812/TPP/18.307] [Medline: 31852039]
20. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
21. The Mobile Economy: Asia Pacific 2018. GMSA Intelligence. 2018. URL: https://www.gsma.com/asia-pacific/resources/the-mobile-economy-asia-pacific-2018/ [accessed 2019-05-31]
22. Buente W, Rathnayake C, Neo R, Dalisay F, Kramer H. Tradition gone mobile: an exploration of #betelnut on Instagram. Subst Use Misuse 2020;55(9):1483-1492. [doi: 10.1080/10826084.2020.1744657]
23. Sheldon P, Bryant K. Instagram: motives for its use and relationship to narcissism and contextual age. Comput Human Behav 2016 May;58:89-97. [doi: 10.1016/j.chb.2015.12.059]
24. Khang H, Han E, Ki E. Exploring influential social cognitive determinants of social media use. Comput Human Behav 2014 Jul;36:48-55. [doi: 10.1016/j.chb.2014.03.038]
25. Katz E, Haas H, Gurevitch M. On the use of the mass media for important things. Am Sociol Rev 1973 Apr;38(2):164. [doi: 10.2307/2094393]
26. Lee E, Lee J, Moon JH, Sung Y. Pictures speak louder than words: motivations for using Instagram. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):552-556. [doi: 10.1089/cyber.2015.0157] [Medline: 26348817]
27. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]
28. Depue JB, Southwell BG, Betzner AE, Walsh BM. Encoded exposure to tobacco use in social media predicts subsequent smoking behavior. Am J Health Promot 2015;29(4):259-261. [doi: 10.4278/ajhp.130214-ARB-69] [Medline: 24670071]
29. Nesi J, Rothenberg WA, Hussong AM, Jackson KM. Friends' alcohol-related social networking site activity predicts escalations in adolescent drinking: mediation by peer norms. J Adolesc Health 2017 Jun;60(6):641-647 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.01.009] [Medline: 28325545]
30. Pokhrel P, Fagan P, Herzog TA, Laestadius L, Buente W, Kawamoto CT, et al. Social media e-cigarette exposure and e-cigarette expectancies and use among young adults. Addict Behav 2018 Mar;78:51-58 [FREE Full text] [doi: 10.1016/j.addbeh.2017.10.017] [Medline: 29127784]
31. Kong G, LaVallee H, Rams A, Ramamurthi D, Krishnan-Sarin S. Promotion of vape tricks on YouTube: content analysis. J Med Internet Res 2019 Jun 18;21(6):e12709 [FREE Full text] [doi: 10.2196/12709] [Medline: 31215510]
32. Systrom K. Strengthening Our Commitment to Safety and Kindness for 800 Million. Instagram Press. 2017. URL: https://instagram-press.com/blog/2017/09/26/strengthening-our-commitment-to-safety-and-kindness-for-800-million/ [accessed 2019-02-25]
33. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center. 2018. URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-02-28]

XSL•FO
RenderX

34. Schrock A. Communicative affordances of mobile media: portability, availability, locatability, and multimediality. Int J Commun 2015;9:1229-1246 [FREE Full text]
35. Butkowski CP, Dixon TL, Weeks K. Body surveillance on Instagram: examining the role of selfie feedback investment in young adult women's body image concerns. Sex Roles 2019;81(5-6):385-397. [doi: 10.1007/s11199-018-0993-6]
36. Sherman LE, Payton AA, Hernandez LM, Greenfield PM, Dapretto M. The power of the like in adolescence: effects of peer influence on neural and behavioral responses to social media. Psychol Sci 2016 Jul;27(7):1027-1035 [FREE Full text] [doi: 10.1177/0956797616645673] [Medline: 27247125]
37. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. Prev Sci 2016 Aug;17(6):710-720 [FREE Full text] [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
38. Primack BA, Carroll MV, Shensa A, Davis W, Levine MD. Sex differences in hookah-related images posted on Tumblr: a content analysis. J Health Commun 2016;21(3):366-375 [FREE Full text] [doi: 10.1080/10810730.2015.1095814] [Medline: 26890733]
39. Allem J, Chu K, Cruz TB, Unger JB. Waterpipe promotion and use on Instagram: #hookah. Nicotine Tob Res 2017 Oct 1;19(10):1248-1252 [FREE Full text] [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
40. Carah N. Algorithmic brands: a decade of brand experiments with mobile and social media. New Media Soc 2017;19(3):384-400. [doi: 10.1177/1461444815605463]
41. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep;26(5):575-578 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
42. Carah N, Shaul M. Brands and Instagram: point, tap, swipe, glance. Mob Media Commun 2016;4(1):69-84. [doi: 10.1177/2050157915598180]
43. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat 2017 Jul;3:- [FREE Full text] [doi: 10.18332/tpc/74709] [Medline: 28815224]
44. Allem J, Escobedo P, Chu K, Cruz TB, Unger JB. Images of little cigars and cigarillos on Instagram identified by the hashtag #swisher: thematic analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
45. Cunningham L. Ancient Chamorro Society. Honolulu, HI: The Bess Press; 1992.
46. Winstock A. Areca nut, betel quids and associated products. In: Miller PM, editor. Principles of Addiction: Comprehensive Addictive Behaviors and Disorders. Volume 1. San Diego, CA: Academic Press; 2013:863-872.
47. Murphy KL, Herzog TA. Sociocultural factors that affect chewing behaviors among betel nut chewers and ex-chewers on Guam. Hawaii J Med Public Health 2015 Dec;74(12):406-411 [FREE Full text] [Medline: 26668772]
48. Little MA, Pokhrel P, Murphy KL, Kawamoto CT, Suguitan GS, Herzog TA. The reasons for betel-quid chewing scale: assessment of factor structure, reliability, and validity. BMC Oral Health 2014 Jun 3;14:62 [FREE Full text] [doi: 10.1186/1472-6831-14-62] [Medline: 24889863]
49. Sotto PP, Mendez AJ, Herzog TA, Cruz C, Chennaux JS, Legdesog C, et al. Barriers to quitting Areca nut consumption and joining a cessation program as perceived by chewer and nonchewer populations in Guam. Subst Use Misuse 2020;55(6):947-953. [doi: 10.1080/10826084.2020.1716012] [Medline: 31973623]
50. Diaz V. Simply Chamorro: tales of demise and survival in Guam. In: Borofsky R, editor. Remembrance of Pacific Pasts: An Invitation to Remake History. Honolulu: University of Hawaii Press; 2000:362-382.
51. Paulino YC, Novotny R, Miller MJ, Murphy SP. Areca (betel) nut chewing practices in micronesian populations. Hawaii J Public Health 2011 Mar;3(1):19-29 [FREE Full text] [Medline: 25678943]
52. Kana'iaupuni SM. Lots of aloha, little data: data and research on native Hawaiian and pacific islanders. AAPI Nexus 2011 Sep;9(1-2):207-211. [doi: 10.17953/appc.9.1-2.k154x5g4711j984m]
53. Anderson M, Jiang J. Teens, Social Media & Technology 2018. Pew Research Center. 2018. URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-06-01]
54. Sarkar U, Le GM, Lyles CR, Ramo D, Linos E, Bibbins-Domingo K. Using social media to target cancer prevention in young adults: viewpoint. J Med Internet Res 2018 Jun 5;20(6):e203 [FREE Full text] [doi: 10.2196/jmir.8882] [Medline: 29871850]
55. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. J Med Internet Res 2017 Jun 7;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
56. Pokhrel P, Dalisay F, Pagano I, Buente W, Guerrero E, Herzog TA. Adolescent tobacco and betel nut use in the US-affiliated pacific islands: evidence from Guam. Am J Health Promot 2019 Sep;33(7):1058-1062 [FREE Full text] [doi: 10.1177/0890117119847868] [Medline: 31117807]
57. Hune S, Kagawa-Singer M. Closing the research and data gap in order to serve Asian Americans and pacific islanders better. AAPI Nexus 2011 Sep;9(1-2):vii-xii. [doi: 10.17953/appc.9.1-2.qmmw65040n774036]
58. Diettrich B. Virtual Micronesia: performance and participation in a pacific Facebook community. Perfect Beat 2016 Mar 31;17(1):52-70. [doi: 10.1558/prbt.v17i1.28841]
59. San Nicolas-Rocca T, Parrish J. Capturing and conveying Chamorro cultural knowledge using social media. International Journal of Knowledge Management 2013;9(3):1-18. [doi: 10.4018/ijkm.2013070101]

XSL•FO
RenderX

60. Hattori AP. Betel mania, from culture to cancer: digestive and discursive uses of the betel nut (areca catechu) in Guam. Pac Stud 2018;41(3):199-223.
61. Moss J, Kawamoto C, Pokhrel P, Paulino Y, Herzog T. Developing a betel quid cessation program on the island of Guam. Pac Asia Inq 2015;6(1):144-150 [FREE Full text] [Medline: 27057560]
62. Griffin MT, Mott M, Burrell PM, Fitzpatrick JJ. Palauans who chew betel nut: social impact of oral disease. Int Nurs Rev 2014 Mar;61(1):148-155. [doi: 10.1111/inr.12082] [Medline: 24512263]
63. Paulino YC, Ettienne R, Novotny R, Wilkens LR, Shomour M, Sigrah C, et al. Areca (betel) nut chewing practices of adults and health behaviors of their children in the freely associated states, micronesia: findings from the children's healthy living (CHL) program. Cancer Epidemiol 2017 Oct;50(Pt B):234-240. [doi: 10.1016/j.canep.2017.07.009] [Medline: 29120830]
64. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
65. Gruzd A, Mai P, Kampen A. A how-to for using Netlytic to collect and analyze social media data: a case study of the use of Twitter during the 2014 Euromaidan revolution in Ukraine. In: Sloan L, Quan-Haase A, editors. The SAGE Handbook of Social Media Research Methods. Thousand Oaks, CA: SAGE Publications; 2016:513-529.
66. Harder A, Howard V, Sedo DR. Creating cohesive community through shared reading: a case study of one book Nova Scotia. Partnership 2015 Jun 25;10(1):-. [doi: 10.21083/partnership.v10i1.3098]
67. Gruzd A, Haythornthwaite C. Enabling community through social media. J Med Internet Res 2013 Oct 31;15(10):e248 [FREE Full text] [doi: 10.2196/jmir.2796] [Medline: 24176835]
68. Murphy KL, Liu M, Herzog TA. Confirmatory factor analysis and structural equation modeling of socio-cultural constructs among chamorro and non-chamorro micronesian betel nut chewers. Ethn Health 2019 Aug;24(6):724-735. [doi: 10.1080/13557858.2017.1346177] [Medline: 28678534]
69. Croucher R, Islam S. Socio-economic aspects of areca nut use. Addict Biol 2002 Jan;7(1):139-146. [doi: 10.1080/13556210120091518] [Medline: 11900634]
70. Gupta PC, Warnakulasuriya S. Global epidemiology of areca nut usage. Addict Biol 2002 Jan;7(1):77-83. [doi: 10.1080/13556210020091437] [Medline: 11900626]
71. Hand M. Visuality in social media: researching images, circulations, and practices. In: Sloan L, Quan-Haase A, editors. The SAGE Handbook of Social Media Research Methods. Thousand Oaks, CA: SAGE Publications; 2016:215-231.
72. Joo YJ, Newcombe D, Nosa V, Bullen C. Investigating Betel Nut Use, Antecedents and Consequences: A Review of Literature. Subst Use Misuse 2020;55(9):1422-1442. [doi: 10.1080/10826084.2019.1666144] [Medline: 31549570]
73. Warnakulasuriya S. Areca nut use following migration and its consequences. Addict Biol 2002 Jan;7(1):127-132. [doi: 10.1080/13556210120091491] [Medline: 11900632]
74. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
75. Cotter K. Playing the visibility game: how digital influencers and algorithms negotiate influence on Instagram. New Media Soc 2019;21(4):895-913. [doi: 10.1177/1461444818815684]

Abbreviations

API: application programming interface
e-cigarettes: electronic cigarettes
e-juice: electronic juice
NIH: National Institutes of Health
NHPI: Native Hawaiian and Pacific Islanders
SCT: social cognitive theory
U&G: uses and gratifications
USAPI: US-Affiliated Pacific Islands

Edited by G Eysenbach; submitted 07.03.19; peer-reviewed by JP Allem, MS Aslam; comments to author 09.12.19; revised version received 27.04.20; accepted 30.04.20; published 09.07.20.

Please cite as:
Buente W, Dalisay F, Pokhrel P, Kramer HK, Pagano I
An Instagram-Based Study to Understand Betel Nut Use Culture in Micronesia: Exploratory Content Analysis
J Med Internet Res 2020;22(7):e13954
URL: https://www.jmir.org/2020/7/e13954
doi:10.2196/13954
PMID:32673220

XSL•FO
RenderX

©Wayne Buente, Francis Dalisay, Pallav Pokhrel, Hanae Kurihara Kramer, Ian Pagano. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterizing Social Media Messages Related to Underage JUUL E-Cigarette Buying and Selling: Cross-Sectional Analysis of Reddit Subreddits

Hejing Liu[1,2], BE; Qiudan Li[1,3], PhD; Yongcheng Zhan[4], PhD; Zhu Zhang[1,3], PhD; Daniel D Zeng[1,2,3], PhD; Scott J Leischow[5], PhD

[1]The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3]Shenzhen Artificial Intelligence and Data Science Institute (Longhua), Shenzhen, China
[4]Orfalea College of Business, California Polytechnic State University, San Luis Obispo, CA, United States
[5]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

Corresponding Author:
Qiudan Li, PhD
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation
Chinese Academy of Sciences
95 Zhongguancun East Rd
Beijing, 100190
China
Phone: 86 10 82449802
Email: qiudan.li@ia.ac.cn

Abstract

Background: Stopping the epidemic of e-cigarette use among youth has become the common goal of both regulatory authorities and health departments. JUUL is currently the most popular e-cigarette brand on the market. Young people usually obtain and exchange information about JUUL with the help of social media platforms. Along with the rising prevalence of JUUL, posts about underage JUUL buying and selling have appeared on social media platforms such as Reddit, which sharply increase the risk of minors being exposed to JUUL.

Objective: This study aims to analyze Reddit messages about JUUL buying and selling among the users of the UnderageJuul subreddit, and to further summarize the characteristics of those messages. The findings and insights can contribute to a better understanding of the patterns of underage JUUL use, and help public health officials provide timely education and guidance to minors who have intentions of accessing JUUL.

Methods: We used a novel cross-subreddit method to analyze the Reddit messages on 2 subreddits. From July 9, 2017, to January 7, 2018, we collected data from the UnderageJuul subreddit, which was created for underage JUUL use discussion. The data set included 716 threads, 2935 comments, and 844 Reddit users (ie, Redditors). We collected our second data set, comprising 23,840 threads and 162,106 comments posted between July 9, 2017, and January 8, 2019, from the JUUL subreddit. We conducted analyses including the following: (1) annotation of users with buying/selling intention, (2) posting patterns discovery and topic comparison, and (3) posting activeness observation of discovered Redditors. Term frequency–inverse document frequency and regular expression-enhanced keyword search methods were applied during the content analysis to extract the posting patterns. The public posting records of the discovered users on the JUUL subreddit during the year after the UnderageJuul subreddit was shut down were analyzed to determine whether they were still active and interested in obtaining JUUL.

Results: Our study revealed the following: (1) Among the 716 threads on the UnderageJuul subreddit, there were 214 threads related to JUUL sale and 168 threads related to JUUL purchase, which accounted for 53.5% (382/714) of threads. (2) Among the 844 Redditors of the UnderageJuul subreddit, 23.82% (201/844) of users were annotated with buying intention, and 21.10% (178/844) of users were annotated with selling intention. There were 34 users with buying/selling intention that self-reported as being <21 years old. (3) The most common key phrases used in selling threads were “WTS,” “want to sell,” “for sale,” and “selling” (154/214, 72.0%). The most common key phrases used in buying threads were “look for/get JUUL/pods” (58/168, 34.5%) and “WTB” (53/168, 31.5%). (4) The most important concern that UnderageJuul Redditors had in obtaining JUULs was

XSL•FO
RenderX

the price (311/1306, 23.81%), followed by the delivery service (68/1306, 5.21%). (5) The most popular flavors among the users with buying/selling intention were mango, cucumber, and mint. The flavor preferences remained consistent on both subreddits. Adverse symptoms related to the mango flavor were reported by 3 users on the JUUL subreddit. (6) In total, 24.4% (49/201) of users wanted to buy JUULs and 46.6% (83/178) of users wanted to sell JUULs, including 11 self-reported underage users, who also participated in the discussions on the JUUL subreddit. (7) Within one year of the UnderageJuul subreddit shutting down, there were 40 users who continued to post 186 threads on the JUUL subreddit, including 10 threads indicating buying/selling willingness that were posted shortly after the UnderageJuul subreddit was closed.

Conclusions: There were overlapping users active in the JUUL and UnderageJuul subreddits. The buying/selling-related content appeared in multiple venues with certain posting patterns from July 9, 2017, to January 7, 2018. Such content might lead to a high risk of health problems for minors, such as nicotine addiction. Based on these findings, this study provided some insights and suggestions that might contribute to the decision-making processes of regulators and public health officials.

(J Med Internet Res 2020;22(7):e16962) doi:10.2196/16962

KEYWORDS

JUUL; e-cigarette; Reddit; cross-sectional analysis; electronic nicotine delivery system; underage JUUL use

Introduction

The use of Electronic Nicotine Delivery Systems (ENDS; also called “e-cigarettes”) has been increasing rapidly among youth. In 2019, a cross-sectional survey with 19,018 participants showed the prevalence of self-reported current e-cigarette use was 27.5% among US high school students and 10.5% among middle school students [1]. At present, JUUL (by JUUL Labs) is the most popular brand, with more than 70% of the market share of e-cigarettes [2]. The appealing flavors and its relaxing effect cause potential intention and actual use among nonsmoking teenagers [3-5]. In addition, because of the lax regulatory environment in the United States, unregulated constituents in e-liquids as well as high nicotine delivery have led to concerns about addiction potential and pulmonary risks from the use of ENDS products [6-8]. To protect the physical health of young people, in April 2018, the Food and Drug Administration (FDA) issued a statement on preventing teenagers from using and buying JUUL and similar products [9]. As the FDA indicated [9], JUUL was more difficult for parents and teachers to recognize and detect than other e-cigarette products. Therefore, increased efforts to strengthen the regulation of minors' access to JUUL have been implemented.

JUUL’s marketing heavily relied on social media platforms like Twitter and Instagram [10]. As a result, young people read and spread information about JUUL with the help of social media platforms [11]. Common JUUL discussions on social media included flavor preference [12-14], use experience [12-16], use location [12,14], purchasing methods [17], etc. Allem et al [12] found that JUUL use occurred in school locations, such as classrooms, bathrooms, libraries, and gyms. Brett et al [14] found teenagers used JUUL because the use of this product was part of a popular trend and to feel the “buzz” (ie, a feeling caused by nicotine). Multiple researchers have noted that adolescents prefer the mango and mint flavors [14,17,18]. Li et al [19] found that the fruit flavor was associated with multiple adverse symptoms, including cough, headache, and lung tightness, though the causal relationship is unclear. With the rise of youth exposure to JUUL, Zhan et al [17] analyzed their purchasing methods. They found that peer-to-peer purchase and shipping was the most common method for UnderageJuul subreddit users.

As an online forum that allows users to create communities, Reddit has multiple sections (also called “subreddits”) related to JUUL discussion. Previous research determined that underage Redditors post about their JUUL use on Reddit [20]. To better understand and surveil youth JUUL use, researchers have conducted analyses on Reddit data focusing on underage JUUL users and their posting content, such as topic discovery [14,15], age and location distribution [17], popular flavor discovery [15,17], and negative symptoms analysis [19].

Previously, there was a public subreddit on Reddit called UnderageJuul, which was dedicated to discussing underage JUUL use. This subreddit was created in July 2017 and had almost 1000 users at its peak. Right before it was shut down in January 2018, the UnderageJuul subreddit was gaining 8 new members per day [20,21]. Previous studies conducted qualitative analysis on the UnderageJuul subreddit and found that posts in this subreddit included content such as self-reported underage JUUL users seeking sellers, sale advertisements without age verification, and adult users buying for minors to earn money, researchers thought such trade-related content indicated that minors had increased access to JUUL, and there might be other similar communities on social media platforms [17,20,21]. However, the characteristics of the buying/selling-related messages among the users of the UnderageJuul subreddit have not been further identified.

This study aimed to analyze the content on the UnderageJuul and JUUL subreddits that indicated JUUL buying and selling, which provides a more comprehensive understanding of underage JUUL use. Specifically, this study was designed to address the following questions:

First, what are the posting patterns of messages related to JUUL buying and selling on the UnderageJuul subreddit? The “posting patterns” refer to the common key phrases, way of expression, variants, or abbreviations the Redditors used when publishing buying/selling-related posts. For instance, Redditors usually use “look for/need + juul/pods” or “WTB” (short for “want to buy”) to express their buying willingness. Discovering users’ posting patterns could enable researchers to extract the core

XSL•FO
RenderX

features of this kind of message, and help regulators continue to identify similar content on social media, thus contributing to the prevention of minors' access to JUULs.

Second, what are the concerns of the users of the UnderageJuul subreddit when they want to buy or sell JUULs? While UnderageJuul users had a willingness to buy and use JUULs, they also expressed concerns regarding JUULs, such as the choice of flavors. Given this subreddit was created to discuss underage JUUL use, identifying the concerns of its users has the potential to uncover specific characteristics of minors or others using that subreddit who intend to buy or sell JUUL. Understanding the topics and concerns discussed in this subreddit could be useful for health departments or educators responsible for the development and implementation of education programs on potential ENDS risk, as well as for regulatory agencies responsible for developing media and educational programs for young people.

Third, will users of the UnderageJuul subreddit cross-publish their comments, questions, and concerns on multiple subreddits at the same time? Will they continue to be active after the UnderageJuul subreddit is closed? We adopted the evaluation of "posting activeness" to address these questions, which involves counting the total number of different posts by users [22]. We chose the JUUL subreddit (ie, the general JUUL discussion subreddit) as a contrast. The posting activeness across subreddits will be compared during two time periods: the time period when UnderageJuul was accessible, and the year after UnderageJuul was shut down. This analysis could provide a deeper understanding of the buying and selling behavior of UnderageJuul users, including the extent to which shutting down one social media channel impacts information-sharing on other platforms, which could have significance for both understanding the breadth and scope of social media networks and the implementation of social media–based educational programs.

By answering the 3 questions above, we hope to provide valuable information to regulators to help them establish practices and policies to minimize adolescent e-cigarette use and safeguard their health. Note that the user analysis in this study was for the users who had public posting records on the UnderageJuul subreddit. The ages of underage users were based on their self-reported age information.

Methods

Overview

We applied a novel cross-subreddit method to analyze the content related to JUUL buying and selling, with a focus on comparing those using the UnderageJuul subreddit (which was originally designed for young people to discuss the use of JUULs, but soon became a platform to explore and share ways to illegally obtain JUULs) with those using the general JUUL subreddit (which is the main Reddit channel for discussing general JUUL-related topics). Figure 1 shows the framework for our cross-subreddit analyses. It consists of three components: data collection; cross-subreddit content and posting activeness analysis; and results.

Data Collection

The data used in this research consisted of two parts: data from the UnderageJuul subreddit and data from the JUUL subreddit.

UnderageJuul Subreddit Data Collection

Since the UnderageJuul subreddit was shut down by the time we began our analysis, we collected data from the UnderageJuul subreddit through the application programming interface (API) provided by pushshift.io [23], which is a website that stores all publicly available Reddit threads and comments. The UnderageJuul subreddit was created on July 9, 2017, and was shut down on January 8, 2018. We obtained the complete UnderageJuul subreddit posting and user information data set from that period, including 716 threads, 2935 comments, and 844 Redditors.

JUUL Subreddit Data Collection

We collected data from the JUUL subreddit via Reddit's API [24] from July 9, 2017, to January 8, 2019, which covered two continuous time periods: (1) the same time period of the UnderageJuul data set (from July 9, 2017, to January 7, 2018), and (2) the year after the end of UnderageJuul activities (from January 8, 2018, to January 8, 2019). The first data set (ie, the JUUL-overlap data set) contained 3978 threads, 31,589 comments, and 5785 Redditors. The second data set (ie, the JUUL-afterwards data set) contained 19,862 threads, 130,517 comments, and 18,460 Redditors. Table 1 shows a detailed description of these three data sets.

Note that the data we used in this research were all publicly available on Reddit, and no personal information (eg, account profile) was included. The usernames contained in these data sets are not presented in this paper, for the privacy protection of the Redditors involved in this study.

XSL•FO
RenderX

Figure 1. Framework for cross-subreddit analyses on Reddit. API: application programming interface.

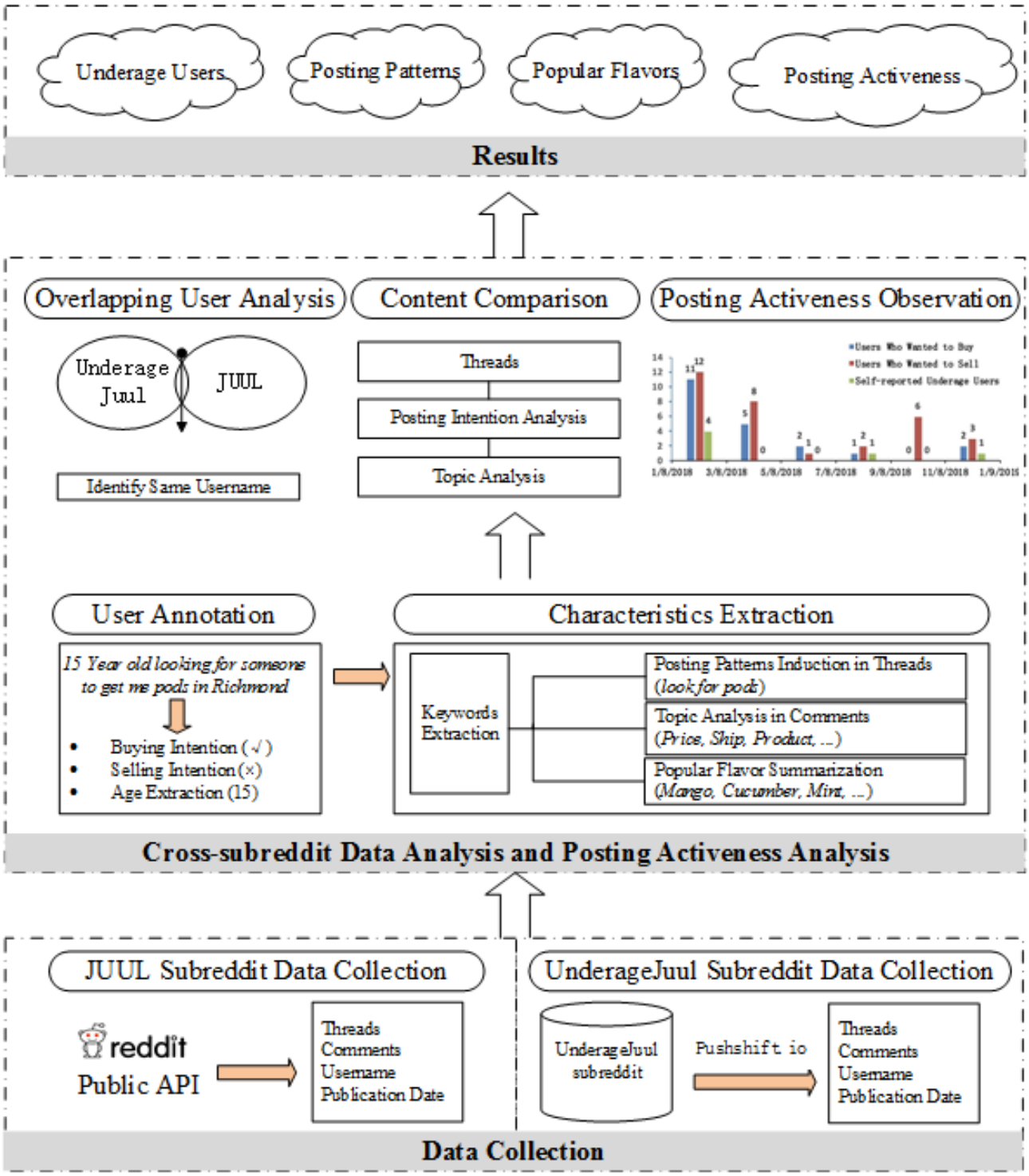


Table 1. Data set description.

Data set	Time span	Threads, n	Comments, n	Redditors, n	Threads per Redditor, n	Comments per Redditor, n
UnderageJuul	07/09/2017-01/07/2018	716	2935	844	1.95	4.11
JUUL-overlap	07/09/2017-01/07/2018	3978	31,589	5785	2.45	5.95
JUUL-afterwards	01/08/2018-01/08/2019	19,862	130,517	21,186	2.49	7.07

XSL•FO
RenderX

Data Analysis

Overview

The data analyses consisted of three steps. The first step was to annotate the users on the UnderageJuul subreddit that indicated buying/selling intention. Second, the text features were extracted to discover the posting patterns in their posting content. Finally, the activeness of the Reddit accounts was analyzed. We used a cross-subreddit analysis method in this study, thus the results of each step were all compared with the content on the JUUL subreddit.

User Analysis Across Subreddits

Overview

We first set the rules for annotation, then manually labeled Redditors who wanted to buy or sell JUUL on the UnderageJuul subreddit. Based on the results of user annotation on the UnderageJuul subreddit, the overlapping users were detected to discover how many of these users also participated in the JUUL subreddit during the same time period.

Step 1: User Annotation on the UnderageJuul Subreddit

First, two annotators labeled the users who wanted to buy or sell JUULs according to the descriptions in Table 2. Users interested in buying or selling usually post a specialized thread to attract sellers or to advertise the product to be sold, so we went through all threads on UnderageJuul subreddit. For example, when a Redditor posted a thread saying "[WTB] Looking to buy blue Juul and cucumber pods for under $75," he/she was labeled as a user who wanted to buy JUULs; when a Redditor posted a thread saying "I am selling Juul warranty codes, …, 25$ per code, 20$ per code (if you buy 3 or more at once)," he/she was labeled as a user who wanted to sell JUULs. The threads were also tagged with corresponding labels. A Redditor could be annotated as a user who wanted to buy JUULs as well as a user who wanted to sell JUULs, if he/she posted multiple threads indicating both buying and selling intentions. Then, we annotated all comments below these labeled threads to discover users who did not post themselves but replied to those threads and showed buying/selling willingness. The Cohen κ coefficient of the annotation from two annotators was 0.83. We next extracted age information from the posts of these users to discover underage JUUL users. Since the US federal minimum age for sale of tobacco products had been raised from 18 to 21 years (ie, the Tobacco 21 legislation [25]), we labeled Redditors who self-reported that they were under 21 years old as the self-reported underage users. We used regular expressions to extract phrases like "# years old" or "I'm #" to detect age information of corresponding users. Some users did not mention a specific age but had expressions like "I'm underage" in their posts. These users were also labeled as self-reported underage users.

Table 2. Annotation categories and description.

Label	Description
User who wanted to buy JUULs	User who wanted to purchase JUULs peer-to-peer or face-to-face, asked others to buy JUULs for him/her, or wanted to buy a JUUL replacement code to get JUULs.
User who wanted to sell JUULs	User who wanted to sell JUULs to others through shipping or face-to-face trade, buy JUUL for others for a fee, or sell JUUL replacement codes.
Self-reported underage user	User who wanted to buy or sell JUULs and had a self-report age under 21 years old or a description indicating underage.

Step 2: Overlapping User Analysis on the JUUL Subreddit

The username is the unique identifier for each Redditor [26]. We compared the public posting records of these users on the two subreddits UnderageJuul and JUUL (ie, the JUUL-overlap data set). Redditors with the same usernames were regarded as overlapping users.

Content Analyses and Comparison

Overview

The following content analyses consisted of two steps. First, we extracted the common key phrases and words from the user-generated content to summarize the regular posting patterns. Second, we compared the overlapping users' posts between the UnderageJuul subreddit and the JUUL subreddit.

Step 1: Posting Patterns Extraction on the UnderageJuul Subreddit

First, the textual features needed to be extracted from users' posts. To achieve this, we used term frequency–inverse document frequency to uncover important words in each post automatically. Based on the frequency of these important words, we adopted regular expression-enhanced keyword search [27] on each post to discover key phrases. The trigrams near the important words were extracted from the buying/selling-related threads. The key phrases were determined according to the frequency rank and whether the expression was complete. For example, the key phrase "need juuls/pods" was determined because it was a complete expression and had a top rank in the search results of the word "need." We manually classified the important words and key phrases into several posting patterns based on semantics and phrase structure.

Step 2: Content Comparison on the JUUL Subreddit

We compared the user-generated contents of overlapping users between the JUUL and UnderageJuul subreddits to discover whether these users also published trade-related content on the JUUL subreddit, and whether their concerns remained consistent. For the first perspective, each thread posted by the overlapping users on the JUUL subreddit was automatically labeled if it implied a willingness to buy or sell JUULs using the posting patterns discovered in Step 1. For the second perspective, we conducted a qualitative content analysis on users' posts on both subreddits. To acquire more detailed and

XSL•FO
RenderX

specific concerns, our analysis was applied to all the comments under trade-related threads on the UnderageJuul subreddit, as well as all the content they posted on the JUUL subreddit (ie, threads and comments). We used the inductive category development method [28] to generate the topics based on the posts' semantics and JUUL background information. For example, expressions like "v1" and "v2" represented versions of JUUL device. When they were identified in the posts, they were categorized with other words (eg, "leak," "refill") that were previously found to be related to JUUL devices, regarded as the "device description" topic. Because flavor is one of the most important factors in ENDS products that appeal to teenagers, we specifically counted the occurrences of JUUL's seven flavors (ie, mango, mint, cucumber, menthol, fruit, crème, and tobacco) among the discussions of these users to discover their preference.

Posting Activeness Observation

Finally, the online posting activeness of these users on the JUUL subreddit was observed during the year after the UnderageJuul subreddit was shut down. We analyzed the posting activeness of users to find whether they continued to express a willingness to buy or sell in other communities after the UnderageJuul subreddit was closed. Specifically, we searched through the JUUL-afterwards data set to count the number of threads and comments posted by these users, recorded their posting dates to get the posting trend changes, and analyzed their posts. Thus, we manually checked their content to determine whether their posts showed the intention to buy/sell JUUL.

Results

User Analyses Across Subreddits

Among 844 Redditors on the UnderageJuul subreddit, we discovered 201 users who wanted to buy JUULs, 178 users who wanted to sell JUULs, and 45 users with both intentions. Among the 334 discovered users, there were 34 users that were self-reported underage. Of these, 30 indicated a willingness to purchase and 4 were willing to sell JUUL. Table 3 shows the proportion of each self-reported age group. The "teenager, age unknown" group represents users that did not mention a specific age but used expressions like "I'm underage" in their posts.

Table 3. Groups of discovered users with self-reported age.

User age (years)	Number of discovered users
≤15	5
16-18	14
19-21	3
Teenager, age unknown	5

After analyzing the users across subreddits, we found 49 of 201 (24.4%) users who wanted to buy JUULs and 83 of 178 (46.6%) users who wanted to sell JUULs also participated in discussions on the JUUL subreddit. Among the overlapping users, 11 underage users (8 users with buying willingness and 3 users with selling willingness) were discovered. Table 4 shows the detailed statistics of the overlapping users.

We also counted the threads and comments posted by the discovered users on each subreddit. They posted 382 threads in the UnderageJuul subreddit and there were 1306 comments under their threads. In addition, they also posted 123 threads and 1600 comments in the JUUL subreddit. Table 5 shows the detailed statistics of their posts. We ranked the activeness of all overlapping users according to their posting amount. The most active self-reported underage user published 16 threads and 96 comments on the JUUL subreddit, as well as 1 thread and 44 comments on the UnderageJuul subreddit. There were 128 Redditors that posted 182 replies to this user.

Table 4. Statistics of the discovered users and overlapping users.

User types	UnderageJuul subreddit, n/N (%)	JUUL-overlap subreddit, n/N (%)
Users who wanted to buy JUULs	201/844 (23.8)	49/201 (24.4)
Users who wanted to sell JUULs	178/844 (21.1)	83/178 (46.6)
Self-reported underage users	34/334 (10.2)	11/34 (32.4)

Table 5. Statistics of posts from discovered users and overlapping users.

Data set	Users who wanted to buy JUULs			Users who wanted to sell JUULs		
	Threads, n/N	Comments, n/N	Total, n	Threads, n/N	Comments, n/N	Total, n
UnderageJuul	168/716	180/1306	348	214/716	238/1306	452
JUUL-overlap	39/3978	447/31,589	486	84/31,589	1153/31,589	1237

XSL•FO
RenderX

Content Analyses and Comparison

Posting Patterns Analysis

We summarized 10 patterns observed in the threads posted by the users discovered on the UnderageJuul subreddit. Table 6 presents each pattern and its ratios. Since one thread may contain multiple key phrases, the total post numbers of the key phrases groups are greater than the total number of posts mentioned in Table 5.

The posting patterns in the threads from users who wanted to buy JUULs all expressed buying willingness directly. They usually included "looking for juul/pods" (58/168, 34.5%) and "want to buy (WTB)" (53/168, 31.5%) in the title, then described the buying requests in the body, such as the flavor and the location. An example is "Need Pods In Orlando Area. Just an underaged kid looking for pods in the Orlando area any flavor." The body of these threads also contained the key phrases like "need juul/pods" or "need help for getting juul/pods" (36/168, 21.4%). Alternative expressions were also used to show buying willingness, such as "anyone selling juul/pods" and "anyone know seller in somewhere" (28/168, 16.7%).

Table 6. Regular posting patterns of discovered users on the UnderageJuul subreddit.

User type, topic number, and key phrases		Post, n (%)
Users who wanted to buy JUULs (n=168)		
1	WTB[a], want to buy, buy juul/pods	53 (31.5)
2	Need juul/pods, need help, need seller	36 (21.4)
3	Look for juul/pods, get juul/pods	58 (34.5)
4	Anyone selling, anyone know seller	28 (16.7)
5	Other	57 (34.0)
Users who wanted to sell JUULs (n=214)		
1	WTS[b], want to sell, for sale, selling	154 (72.0)
2	Ship, discreet	97 (45.3)
3	Charger, refill, starter	55 (25.7)
4	Sealed[c], unboxing, unopen, brand new	40 (18.7)
5	Other	39 (18.2)

[a]WTB: want to buy.

[b]WTS: want to sell.

[c]"Sealed" indicates that the product is new and has not been opened.

The posting patterns in the threads from users who wanted to sell JUULs included multiple kinds of information. They usually contained key phrases such as "WTS," and "something for sale" (Pattern 1, 154/214, 72.0%) in the title to draw attention. Pattern 2 (97/214, 45.3%) was about shipping services, such as the shipping scope and discretion. Pattern 3 (55/214, 25.7%) involved a description of the product, including the accessories. If the product was brand-new, that was emphasized (40/214, 18.7%). For instance, a user who wanted to sell JUULs posted the following thread: "[WTS] JUUL V3 Starter Kit $40 in BTC. Brand new, sealed starter kit. Will ship quickly and discreetly."

Based on the above posting patterns, we further analyzed the threads posted by the overlapping users on the JUUL subreddit and found 17 threads with buying purpose and 18 threads with selling purpose.

Content Comparison

We summarized the 15 topics discussed by the discovered users in the UnderageJuul and JUUL subreddits. Table 7 shows the 10 topics discussed in the UnderageJuul subreddit and the 5 topics discussed in the JUUL subreddit.

XSL•FO
RenderX

Table 7. Topic comparison of the UnderageJuul and JUUL subreddits.

Subreddit, topic number, and topic		Key phrases	Posts, n (%)
UnderageJuul subreddit (n=1306)			
1	Price	Price, how much, $number	311 (23.81)
2	Contact	PM[a], message	63 (4.82)
3	Ship	Ship, shipping	68 (5.21)
4	Scam	Scammer, scam, fake	83 (6.36)
5	Product description	v1, v2, v3[b], refill	53 (4.06)
6	Purchasing method	eBay, website, .com, store	20 (1.53)
7	Paying method	Gift card, PayPal, Bitcoin, BTC	40 (3.06)
8	Stock	Still available, still selling, how many, still have, SOLD	15 (1.15)
9	Flavor	Mango, mint, fruit, cucumber, crème, menthol, tobacco	27 (2.07)
10	Other	Posts did not contain the words above	626 (47.93)
JUUL subreddit (n=123)			
1	Buying/selling	This was annotated manually.	35 (28.46)
2	Flavor	Mango, mint, fruit, cucumber, crème, menthol, tobacco	16 (13.01)
3	Product description	Leak, refill, charged, help, real, fake, charger, battery	25 (20.33)
4	Experience sharing	Buzz, cough, lung	9 (7.32)
5	Other	Posts did not contain the words above	38 (30.90)

[a]PM: private message.

[b]v1, v2, and v3 represent the different versions of JUUL devices.

In the UnderageJuul subreddit, discussions were primarily focused on buying and selling JUULs. Topic 1 (311/1306, 23.81%) refers to users raising questions about the price. The replies usually contained specific prices in the form of $number. Topic 2 (63/1306, 4.82%) refers to posts in which authors would ask others to contact them through private messages if they did not want to directly answer questions in public replies. Topic 3 (68/1306, 5.21%) was about shipping services. Topic 4 (83/1306, 6.36%) includes posts in which some users commented under the threads with selling purposes that the post author is a scammer. Topic 5 (53/1306, 4.06%) refers to descriptions of the selling product, including the version of the product and whether it was refilled. Topic 6 (20/1306, 1.53%) and Topic 7 (40/1306, 3.06%) refer to the methods of purchasing and paying. Since some users self-claimed they were vendors from eBay or other websites, they accepted payment methods such as VISA gift cards and PayPal. Some users offered to buy JUULs from stores for underage users for a fee. Bitcoin was sometimes accepted as a form of payment. Topic 8 (15/1306, 1.15%) is related to users asking post authors about their stock. The post authors might update the threads with “SOLD” to indicate that they had sold out. Topic 9 (27/1306, 2.07%) represents the flavors discussed in the threads and comments; this topic is analyzed in detail in the next section.

In the JUUL subreddit, Topic 1 (35/123, 28.46%) represented the threads about buying and selling JUULs, which suggested the most popular topic discussed by these users did not change on this subreddit. Topic 2 (16/123, 13.01%) was about the flavors discussed by the overlapping users. The flavor topic is also analyzed individually in the next section. However, the focus of this topic was different from that on the UnderageJuul subreddit. Topic 3 (25/123, 20.33%) referred to the product descriptions. Unlike the product description topic (Topic 5) on the UnderageJuul subreddit, the discussions about this topic on the JUUL subreddit were about describing the device’s problems and seeking help. For instance, a post about this topic stated, “… All my official juul pods keep leaking out all over my juul. Idk what's going on.” Topic 4 (3/123, 2.44%) was about negative experience sharing. For example, one user said, “I used the juul to stop smoking but have noticed a chronic cough starting to begin.”

It can be seen that access to JUUL was a common topic for the discovered users in both subreddits. However, the content in the UnderageJuul subreddit included more details about JUUL transactions. The flavor topic and the product description topic appeared on both subreddits. The focal point in the UnderageJuul subreddit was users describing their buying/selling demands, while in the JUUL subreddit, users commonly shared their JUUL use experience.

We identified the 10 users who posted the most for each topic as the active users. In the UnderageJuul subreddit, there were 16 users who were active in multiple topics. There were 4 active users who were active in 3 topics at the same time. The most active user was in the top 10 among 6 topics. In the JUUL subreddit, there were 6 users active in multiple topics. The most active user posted about 4 topics at the same time. Notably, there were 7 users who were active in both subreddits, and the topics they promoted were related to JUUL buying and selling.

XSL•FO
RenderX

Flavor Analysis and Comparison

We counted the word frequency of the 7 flavors mentioned above and compared their popularity in the two subreddits. Among 382 threads and 1306 comments in the UnderageJuul subreddit, mango (n=50) was the most popular flavor, followed by cucumber (n=39) and mint (n=31). Fruit (n=9), tobacco (n=7), menthol (n=5), and crème (n=5) were not mentioned frequently. These flavors were mentioned when users described the selling products or the buying requests, as shown in the two examples given here:

> *Pods Available: I have 8 tobacco, 16 cucumber, 8 mint, 12 menthol, 8 mango, and 4 fruit. …*
>
> *… Looking for 2 packs of mango pods let me know if you have any for sale.*

Among the 123 threads in the JUUL subreddit, there were 16 threads that mentioned flavor. The most popular flavor was mango (n=8), followed by cucumber (n=7) and mint (n=7). These results were similar to those found in the UnderageJuul subreddit, which indicates that users' flavor preferences did not change between subreddits. There was not much discussion of the crème (n=2), menthol (n=2), tobacco (n=2), and fruit (n=1) flavors. One example of a flavor topic post is as follows:

> *Just got my hands on a couple packs of cool cucumber. …, it's really good and up there with mango.*

However, we found 3 reports of adverse symptoms among the 8 threads that mentioned the mango flavor in the JUUL subreddit. Two users reported mouth burning and one user reported a stomachache, as shown below:

> *…when ever I hit a mango pods my stomach almost immediately starts hurting.*

Posting Activeness Observation

We observed the posting activeness of all labeled users in the year following the shutdown of the UnderageJuul subreddit (January 8, 2018, to January 8, 2019). Table 8 shows the number of labeled users who had posted during this time period, and Table 9 shows the number of posts they made. For both tables, values were summed across consecutive 2-month periods.

Table 8. The number of users who posted messages in the year after the UnderageJuul subreddit was removed.

Time period	Users who wanted to buy	Users who wanted to sell	Self-reported underage users
1/8/2018-3/7/2018	11	12	4
3/8/2018-5/7/2018	5	8	0
5/8/2018-7/7/2018	2	1	0
7/8/2018-9/7/2018	1	2	1
9/9/2018-11/7/2018	0	6	0
11/8/2018-1/9/2019	2	3	1

Table 9. The number of posts by the labeled users in the year after the UnderageJuul subreddit was removed.

Time period	Posts by users who wanted to buy	Posts by users who wanted to sell	Posts by self-reported underage users
1/8/2018-3/7/2018	34	48	17
3/8/2018-5/7/2018	12	24	0
5/8/2018-7/7/2018	6	3	0
7/8/2018-9/7/2018	1	3	1
9/9/2018-11/7/2018	0	23	0
11/8/2018-1/9/2019	4	28	2

In the first 4 months after the UnderageJuul subreddit was shut down, the discovered users were still active in the JUUL subreddit. They posted 10 threads with the purpose of buying or selling. In the next 2 months, these users were less active and did not post threads with buying or selling willingness. One possible reason was that the administrators of the JUUL subreddit started to prohibit posts about trading JUULs. In the second half of the year, the discussions changed to flavor and device-related topics. There were 2 users who expressed the intention of quitting JUULs because of nicotine addiction and the high price.

Discussion

Principal Findings

This paper used a cross-subreddit method to analyze underage JUUL use. Based on our previous study [17], this paper further analyzed the content related to underage JUUL buying and selling on Reddit, and summarized its characteristics and patterns. Additionally, this paper analyzed the posting activeness of the discovered users in the JUUL subreddit after long-term observation.

In summary, we found 214 threads related to JUUL sale and 168 threads related to JUUL purchase on the UnderageJuul subreddit, which accounted for 53.5% (382/714) of the

XSL•FO
RenderX

UnderageJuul subreddit's threads. In addition, we found that these threads were posted with certain regular word-level patterns. The most common key phrases used in selling-related threads were "WTS," "want to sell," "for sale," and "selling" (154/214, 72.0%). The most common key phrases used in buying-related threads were "look for/get JUUL/pods" (58/168, 34.5%) and "WTB" (53/168, 31.5%). The FDA has already announced policies about preventing youth use of and access to JUUL e-cigarettes, and reducing the marketing and promotion of tobacco products toward minors [29,30]. Though the UnderageJuul subreddit has been removed, there might be other similar communities that support underage JUUL trading. The posting patterns could help discover this type of content among other communities on Reddit and different social media platforms such as Twitter and Instagram. Once such content is discovered, a link to the official educational page about the tobacco products' adverse health effects could be automatically added below such posts or sent to the content publisher via private message. This could have implications for interventions to reduce youth ENDS use.

We have demonstrated that discussions about obtaining and using JUUL e-cigarettes occur in multiple venues, and not just the subreddit dedicated to underage JUUL use (UnderageJuul). Among the 844 Redditors of the UnderageJuul subreddit, 23.82% (201/844) of users were annotated with buying intention, and 21.10% (178/844) of users were annotated with selling intention. We found that 24.38% (49/201) of users who wanted to buy JUULs and 46.63% (83/178) of users who wanted to sell JUULs had posting activeness on the JUUL subreddit. The results are consistent with our earlier research [17] regarding the approaches of illegally obtaining JUUL products. The buying/selling topics were common concerns for users on both subreddits during the same time span. This indicates the people who want to access to JUULs may not express their desires on only one platform and may seek multiple resources. When regulatory organizations take action to reduce youth's access to JUUL and similar products, they might need strategies on multiple social media platforms simultaneously. When selling ENDS products via online shopping sites and other vaping-related websites, necessary warning statements should be clearly presented alongside, including the health risks and nicotine strength [31].

In our cross-subreddit analysis, we developed a procedure to evaluate the effect that the discovered users had based on their posting activeness and user-generated content. The most active self-reported underage user we discovered had published content that at least 128 Redditors read about, so there was a social network effect that could impact youth use of ENDS products like JUUL. Evaluation of JUUL-related content posted by those active users on social media platforms, especially the underage ones, is highly relevant to educators and those involved in tobacco regulation. In addition, it is not clear from these analyses whether these active users are paid by vendors or companies for advertising tobacco products to minors, and this is an important analysis that should be conducted to understand the motivation of those who post about these products on social media.

Our findings suggest that price was what the UnderageJuul users most cared about, along with the delivery services. The mango, mint, and cucumber flavors were the favorite flavors among these users. In January 2018, the UnderageJuul subreddit was banned. We found there was a self-reported underage user that posted a message on the JUUL subreddit asking for an alternative to the mango flavor, such as mint. In addition, we discovered negative health symptoms were reported on the JUUL subreddit after the use of the mango flavor. These popular flavors are still available in many countries and young people continue to discuss ways to acquire ENDS products like JUUL on Twitter, Reddit, etc. Despite regulatory and company efforts to reduce access to flavored ENDS products [32], considerable research is still needed on the role of flavors regarding the addiction potential of ENDS products [13,18,19,33]. In particular, since social media can provide timely and rapid information on what is being discussed about a particular ENDS product, continued analysis of social media is an essential early warning surveillance method for identifying addiction and the health risks of JUUL and other products (eg, the NJOY Ace product that has steadily increased in sales).

In our posting activeness observation, we found that in the year after the UnderageJuul subreddit was removed, 40 users (including 4 self-reported underage users) continued to post 186 threads in the JUUL subreddit. There were 10 threads with buying/selling willingness during the first 2 months. This result indicates that shortly after the UnderageJuul subreddit was closed, users transferred to other subreddits to have discussions, and continued to publish messages about buying/selling JUUL pods through social media platforms.

In conclusion, this study highlights the need for continued research on how social media can become a more fundamental component of the tobacco and nicotine surveillance system. In particular, rapid and extensive analyses of social media can serve as an early warning system of the rapid rise in the use of new products and health problems that are occurring as a result of using those products.

Limitations and Future Work

Since the data we used for analysis was user-generated public information (ie, post, publication date, and username), there were many limitations for us regarding determining the authenticity and legitimacy of users' claims like self-reported age, their stock of JUUL devices, and the completion of trading. As the UnderageJuul subreddit was built specifically for discussing underage JUUL use, we thought the age distribution analysis was valuable and necessary. In addition, we noticed the scam topic (ie, Topic 4 on the UnderageJuul subreddit on Table 7) during the content analysis, which indicates some Redditors who claimed to have JUULs to sell might post such content to scam other users, rather than to trade JUULs. This might cause discrepancies for regulators when trying to identify actual buyers and sellers.

In addition, the conclusions and insights of our study were based on content analysis of the two specific subreddits during the specific time span. Our findings might not be generalizable to other subreddits or other time periods, considering the different barrier policies being carried out. However, future studies could

XSL•FO
RenderX

be extended to other subreddits that are about JUULs and other social media platforms (eg, Twitter) with the same methodology and mechanism.

Conclusion

This is the first study to investigate the patterns of buying and selling among underage JUUL users by utilizing information from multiple connected online forums. We hope these findings can be conducive to timely guidance and education of underage JUUL users, thus protecting the health of young people.

Acknowledgments

This work was partially supported by the National Key R&D Program of China under grant number 2016QY02D0305, the National Institutes of Health (NIH) grant 5R01DA037378-05; the National Natural Science Foundation of China under grant numbers 61671450, 71621002, 71902179, and 71702181; the Key Research Program of the Chinese Academy of Sciences under grant number ZDRW-XH-2017-3, and the Early Career Development Award of the State Key Laboratory for Management and Control of Complex Systems (SKL-MCCS) under grants 20190212 and 20190204. Research reported here was supported by the National Center Institute (NCI) and FDA Center for Tobacco Products (CTP). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the FDA.

Authors' Contributions

HL, QL, YZ, ZZ, DZ, and SL conceived the idea for this study. HL and QL designed the study. HL, QL, and YZ collected the data, conducted the data analysis, and drafted the manuscript. ZZ, DZ, and SL provided critical feedback and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

SL is the co-owner and editor-in-chief of the journal Tobacco Regulatory Science. All other authors declare no conflicts of interest.

References

1. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. e-Cigarette Use Among Youth in the United States, 2019. JAMA 2019 Nov 05;322(21):2095 [FREE Full text] [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
2. King BA, Gammon DG, Marynak KL, Rogers T. Electronic Cigarette Sales in the United States, 2013-2017. JAMA 2018 Oct 02;320(13):1379-1380 [FREE Full text] [doi: 10.1001/jama.2018.10488] [Medline: 30285167]
3. Vaporized: Youth and young adult exposure to e-cigarette marketing. Truth Initiative. 2019. URL: https://truthinitiative.org/sites/default/files/media/files/2019/03/Vaporized-Youth-Exposure-To-E-Cigarette-Marketing.pdf [accessed 2020-01-28]
4. Alexander JP, Williams P, Lee YO. Youth who use e-cigarettes regularly: A qualitative study of behavior, attitudes, and familial norms. Prev Med Rep 2019 Mar;13:93-97 [FREE Full text] [doi: 10.1016/j.pmedr.2018.11.011] [Medline: 30568866]
5. Pepper JK, Ribisl KM, Brewer NT. Adolescents' interest in trying flavoured e-cigarettes. Tob Control 2016 Nov 15;25(Suppl 2):ii62-ii66 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053174] [Medline: 27633762]
6. Barrington-Trimis JL, Samet JM, McConnell R. Flavorings in electronic cigarettes: an unrecognized respiratory health hazard? JAMA 2014 Dec 17;312(23):2493-2494 [FREE Full text] [doi: 10.1001/jama.2014.14830] [Medline: 25383564]
7. Centers for Disease Control and Prevention. Severe Pulmonary Disease Associated with Using E-Cigarette Products. Health Alert Network. 2019. URL: https://emergency.cdc.gov/han/han00421.asp [accessed 2019-09-30]
8. Flower M, Nandakumar L, Singh M, Wyld D, Windsor M, Fielding D. Respiratory bronchiolitis-associated interstitial lung disease secondary to electronic nicotine delivery system use confirmed with open lung biopsy. Respirology Case Reports 2017 Apr 03;5(3):e00230. [doi: 10.1002/rcr2.230]
9. FDA. Statement from FDA Commissioner Scott Gottlieb, M.D., on new enforcement actions and a Youth Tobacco Prevention Plan to stop youth use of, and access to, JUUL and other e-cigarettes. 2018. URL: https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-enforcement-actions-and-youth-tobacco-prevention [accessed 2019-09-30]
10. Huang J, Duan Z, Kwok J, Binns S, Vera LE, Kim Y, et al. Vaping versus JUULing: how the extraordinary growth and marketing of JUUL transformed the US retail e-cigarette market. Tob Control 2019 Mar 31;28(2):146-151 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054382] [Medline: 29853561]
11. Chu K, Colditz JB, Primack BA, Shensa A, Allem J, Miller E, et al. JUUL: Spreading Online and Offline. J Adolesc Health 2018 Nov;63(5):582-586 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.08.002] [Medline: 30348280]
12. Allem J, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Sep 01;190:1-5 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
13. Wang L, Zhan Y, Li Q, Zeng D, Leischow S, Okamoto J. An Examination of Electronic Cigarette Content on Social Media: Analysis of E-Cigarette Flavor Content on Reddit. Int J Environ Res Public Health 2015 Nov 20;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]

XSL•FO
RenderX

14. Brett EI, Stevens EM, Wagener TL, Leavens EL, Morgan TL, Cotton WD, et al. A content analysis of JUUL discussions on social media: Using Reddit to understand patterns and perceptions of JUUL use. Drug Alcohol Depend 2019 Jan 01;194:358-362. [doi: 10.1016/j.drugalcdep.2018.10.014] [Medline: 30472576]
15. Liu H, Li Q, Yao R. Analyzing Topics of JUUL Discussions on Social Media Using a Semantics-assisted NMF model. In: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). 2019 Presented at: 2019 IEEE International Conference on Intelligence and Security Informatics (ISI); July 1-3 2019; Shenzhen, China p. 212-214. [doi: 10.1109/isi.2019.8823541]
16. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying Topics for E-Cigarette User-Generated Contents: A Case Study From Multiple Social Media Platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
17. Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL Use Patterns: Content Analysis of Reddit Messages. J Med Internet Res 2019 Sep 09;21(9):e13038 [FREE Full text] [doi: 10.2196/13038] [Medline: 31502542]
18. Leventhal AM, Miech R, Barrington-Trimis J, Johnston LD, O'Malley PM, Patrick ME. Flavors of e-Cigarettes Used by Youths in the United States. JAMA 2019 Nov 05;322(21):2132 [FREE Full text] [doi: 10.1001/jama.2019.17968] [Medline: 31688891]
19. Li Q, Zhan Y, Wang L, Leischow SJ, Zeng DD. Analysis of symptoms and their potential associations with e-liquids' components: a social media study. BMC Public Health 2016 Jul 30;16(1):674 [FREE Full text] [doi: 10.1186/s12889-016-3326-0] [Medline: 27475060]
20. Kavuluru R, Han S, Hahn EJ. On the popularity of the USB flash drive-shaped electronic cigarette Juul. Tob Control 2019 Jan 13;28(1):110-112 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054259] [Medline: 29654121]
21. McKelvey K, Baiocchi M, Halpern-Felsher B. Adolescents' and Young Adults' Use and Perceptions of Pod-Based Electronic Cigarettes. JAMA Netw Open 2018 Oct 05;1(6):e183535 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3535] [Medline: 30646249]
22. Shi J, Salmon CT. Identifying Opinion Leaders to Promote Organ Donation on Social Media: Network Study. J Med Internet Res 2018 Jan 09;20(1):e7 [FREE Full text] [doi: 10.2196/jmir.7643] [Medline: 29317384]
23. Reddit Statistics. pushshift.io. URL: https://pushshift.io/ [accessed 2019-08-18]
24. Reddit. API documentation. 2019. URL: https://www.reddit.com/dev/api/ [accessed 2019-09-30]
25. Food and Drug Administration. Tobacco 21. 2019. URL: https://www.fda.gov/tobacco-products/retail-sales-tobacco-products/tobacco-21 [accessed 2020-02-12]
26. Reddit. 2018. Reddit Privacy Policy. URL: https://www.redditinc.com/policies/privacy-policy [accessed 2019-09-30]
27. Thompson K. Programming Techniques: Regular expression search algorithm. Commun ACM 1968;11(6):419-422. [doi: 10.1145/363347.363387]
28. Mayring P. Qualitative content analysis: theoretical foundation, basic procedures and software solution. Social Science Open Access Repository. 2014. URL: https://www.ssoar.info/ssoar/handle/document/39517 [accessed 2019-09-30]
29. Food and Drug Administration. Warning Letters and Civil Money Penalties Issued to Retailers for Selling JUUL and Other E-Cigarettes to Minors. 2018. URL: https://www.fda.gov/tobacco-products/ctp-newsroom/warning-letters-and-civil-money-penalties-issued-retailers-selling-juul-and-other-e-cigarettes [accessed 2020-02-09]
30. Food and Drug Administration. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. 2020. URL: https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children [accessed 2020-01-04]
31. Food and Drug Administration. Selling Tobacco Products in Retail Stores. 2020. URL: https://www.fda.gov/tobacco-products/retail-sales-tobacco-products/selling-tobacco-products-retail-stores [accessed 2020-06-19]
32. 2019. JUUL Labs stops the sale of mint JUULpods in the United States. URL: https://newsroom.juul.com/juul-labs-stops-the-sale-of-mint-juulpods-in-the-united-states/ [accessed 2020-02-01]
33. Cwalina SN, Leventhal AM, Barrington-Trimis JL. E-cigarette flavour enhancers: Flavoured pod attachments compatible with JUUL and other pod-based devices. Tob Control 2020 Apr 09:2020. [doi: 10.1136/tobaccocontrol-2020-055618] [Medline: 32273434]

Abbreviations

API: application programming interface
ENDS: Electronic Nicotine Delivery Systems
FDA: Food and Drug Administration

XSL•FO
RenderX

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by J Cappella, A Majmundar, N O'Kane; comments to author 06.01.20; revised version received 26.05.20; accepted 25.06.20; published 20.07.20.

Please cite as:
Liu H, Li Q, Zhan Y, Zhang Z, Zeng DD, Leischow SJ
Characterizing Social Media Messages Related to Underage JUUL E-Cigarette Buying and Selling: Cross-Sectional Analysis of Reddit Subreddits
J Med Internet Res 2020;22(7):e16962
URL: http://www.jmir.org/2020/7/e16962/
doi:10.2196/16962
PMID:32706661

©Hejing Liu, Qiudan Li, Yongcheng Zhan, Zhu Zhang, Daniel D Zeng, Scott J Leischow. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Development of a Smart Health Awareness Message Framework Based on the Use of Social Media: Quantitative Study

Elaf Ali Alsisi[1], BSc, MSc, PhD; Ahmed Al-Ashaab[1], PhD; Wadhah Ahmed Abualfaraa[1], BSc, MSc, PhD

Cranfield University, Cranfield, United Kingdom

Corresponding Author:
Elaf Ali Alsisi, BSc, MSc, PhD
Cranfield University
Bedfordshire, MK43 0AL
Cranfield
United Kingdom
Phone: 44 1234750111 ext 5622
Email: e.alsisi85@gmail.com

Abstract

Background: Social media has recently provided a remarkable means of delivering health information broadly and in a cost-effective way. Despite its benefits, some difficulties are encountered in attempting to influence the public to change their behavior in response to social media health messages.

Objective: This study aimed to explore the factors that affect individuals' acceptance of using social media as a tool for receiving health awareness messages and adapting such content accordingly by developing a smart health awareness message framework.

Methods: A quantitative method was adapted to validate the hypotheses and proposed framework through the development of a survey based on the technology acceptance model with the extension of other constructs. The survey was distributed on the web to 701 participants from different countries via Qualtrics software; it generated 391 completed questionnaires, and the response rate was 55.8% (391/701).

Results: Of the 391 respondents, 121 (30.9%) used social media platforms often during the week, and 27 participants (6.9%) did not use social media. In addition, 24.0% (94/391) of the respondents used these platforms to seek health information. On the basis of the results, perceived usefulness (β=.37; *P*<.001), gain-framed message (β=.04; *P*<.001), and loss-framed message (β=.08; *P*<.001) were seen to positively and significantly influence people's intention to use social media as a means to spread information about health promotion. The proposed smart health awareness message framework identifies 64.2% of the variance in intention to use, 55.4% of the variance of perceived usefulness, and 26.2% of the variance of perceived ease of use.

Conclusions: This study sheds light on the factors that are associated with people's intention to use and adopt social media in the health promotion domain. The findings reveal that the intention of using social media for health awareness purposes is positively impacted by the perception of usefulness of social media and the design of health messages. Future research might seek to explore other factors that relate to people's behavior. This point of view will assist health organizations in developing their health messages more effectively and to be patient friendly.

(J Med Internet Res 2020;22(7):e16212) doi:10.2196/16212

KEYWORDS

social media; health promotion and social media; health awareness; health promotion; eHealth; technology acceptance theory

Introduction

Background

The advent of the internet has become a fundamental avenue for gaining health information [1] and for the provision of interventions to enhance healthy behaviors [2]. Rapid and easy access to the internet has driven individuals to go on the web to seek health information [3]. The internet has contributed to the evolution of a new era of communication, known as social media. The phenomenon of social media is defined as a collection of web-based social networking apps that enable individuals or groups of people to communicate and interact with each other, share information, collaborate, and exchange content [4,5]. The advancement of the internet, web-based health information, and social media has driven the emergence of electronic health (eHealth). According to the World Health Organization [6], eHealth can be defined as "the use of

XSL•FO
RenderX

information and communication technologies (ICT)," which involves the development of spreading health information through digital methods [7,8].

One aspect of public health communication, which has received increasing attention, is the media channels through which health messages could be successfully conveyed to a wide range of relevant audiences. Several studies have found that mass media (eg, television, radio, newspaper, leaflets, and posters) have a positive impact on health promotion [9-12]. However, others have marked the ineffectiveness of this impact [13], unlike social media, which has greater potential in health promotion for individuals and health care professionals because of its ability to deliver meaningful health content [14] in different formats such as text, images, and videos [15]. It can be argued, therefore, that social media has largely proved to be an effective and influential instrument in spreading health awareness messages [16] because of its easy access by all socioeconomic strata and its cost effectiveness [3]. Due to the evolution of eHealth and social media tools, health organizations reinforce practitioners in providing health-related information to increase health awareness and obtain better health outcomes [17,18].

Despite the increasing utilization of social media by health organizations in disseminating health awareness, the actual impact of social media interventions demands further research to explore the factors that may affect users' acceptance of this technology and adoption of the content [15,19]. These factors include the frame of the message, trust of the content, and the degree of technology acceptance [20-22]. This study is motivated by the need to take into consideration such key factors that lead to effective acceptance of social media as a means to receive, read, and apply health awareness messages.

The paper provides the results of the smart health awareness message framework development and, in turn, ensures spreading health awareness messages effectively on a faster and wider scale through social media. The focus of this paper presents the identification of the factors influencing an individual's intention to use social media as a means to receiving health awareness messages and following its instructions for the well-being of the individual by using the technology acceptance model (TAM) [23], task technology fit (TTF) [24], and prospect theory [25].

The research approach starts with a review of related literature concerning health awareness messages and the use of social media in spreading such messages to a wider community. The second stage involved developing a conceptual framework of the factors influencing an individual's intention to use social media for health promotion. The effectiveness of the proposed framework was evaluated based on the hypotheses developed in this study. To validate these hypotheses, public opinion was analyzed based on a web-based survey using the Qualtrics software with 391 participants.

Such a random sample size would be a good representative because it reflects the characteristics of the population from which it has been drawn (ie, from a wide range of countries) and different opinions that were relatively close to each other.

The remainder of this paper is structured as follows. The first section includes an introduction that presents the research motivation, research approach, and the aim of this paper. The second section presents a review of the related literature. The third section presents the conceptual framework along with the proposed hypotheses. In the fourth section, methods of data collection and measurement development are presented. The section following the fourth section presents some public perspectives of the smart health awareness message framework through data analysis, including testing hypotheses. Finally, the authors conclude with a discussion of the research limitations and future work.

Literature Review

Public health communication has emerged as a modern strategy to change public behavior by raising awareness of risk diseases. Public health communication refers to "the scientific development, strategic dissemination, and critical evaluation of relevant, accurate, accessible, and understandable health information communicated to and from intended audiences to advance the health of the public" [26]. Therefore, health promotion encompasses the development of approaches that supply health knowledge to individuals, motivating them to adopt the healthy behaviors and change their current ones [27]. Traditionally, mass media has been used as a tool for public health promotion, which has involved a variety of forms including television, newspapers, radio, booklets, billboards, leaflets, and posters [9,12,28]. Each format varies according to the level of effectiveness and drawbacks. For example, numerous studies have explored the efficacy of using television campaigns to promote smoking cessation [29,30]. However, exposure to such campaigns has been found to be expensive in comparison with radio broadcasts [12,30].

Although several studies have highlighted the effectiveness of promoting health awareness via leaflets and posters [31-33], the reality is that they are still an expensive media to be published. This is due to the long process and expense of publishing paper-based media and also the factors such as time and labor consumption, limited information being given to the audience [34], poor health content [35], and overlapping information [36]. These reservations also include editing the health content, graphical design, printing, and distributing. Updating any of these printed media requires a further long loop of modification.

Social media has a great potential in public health communication, as it provides patients and the public with the best opportunity by delivering meaningful health content. Ba and Wang [14] found that online social groups have an essential role to play in an individual's routine in terms of encouraging them to adopt a healthy lifestyle through observing their daily diet. Previous research has focused on customized digital health interventions that help individuals to control chronic disease and make proper decisions accordingly [37]. Roland et al [38] developed an online community represented by #FOAMed on Twitter for the purpose of sharing medical knowledge. Similarly, Diddi and Lundy [39] indicated the usage of Twitter to spread breast cancer awareness by 4 different health organizations, presenting different factors of the health belief model in the content of the message. A previous study has supported diabetic

XSL•FO
RenderX

people by offering a forum for sharing personal experience and providing feedback on performance by physicians [40].

So far, few research studies have examined the influential factors that affect people's intention to use social media in the health promotion context [15,20]. However, understanding these factors is important for designing health promotion messages that incorporate content strategy and simplicity [41]. To fill this research gap, a conceptual smart health awareness message framework was developed based on the TAM, TTF, and prospect theory, as presented in the following section.

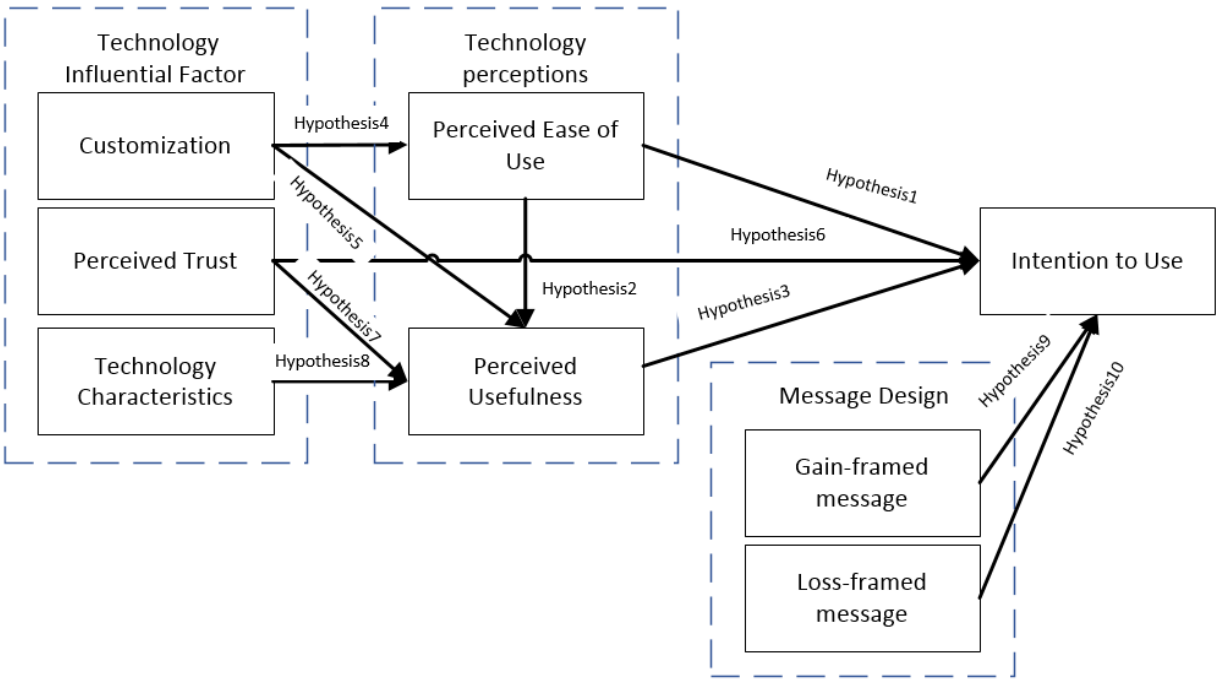
The TAM assumes that the extent to which the technology is accepted and used by an individual is predicted by 2 main constructs (factors): perceived usefulness and perceived ease of use [23]. TTF focuses mainly on the features that the technology offers, and thus, it believes that technology must match the task it supports to have performance impact [24]. TTF has 4 key constructs (elements), one of which is technology characteristics. Prospect theory postulates that health communication messages can be designed to shed light on the benefits (gain) or the consequences (loss) of performing a specific behavior [25].

Smart Health Awareness Message Framework and Hypothesis Development

The smart health awareness message framework includes different elements, which are called constructs, and each construct represents the key factor of a different adapted theory. Thus, this study investigates the impact of such constructs that influence an individual's acceptance of using social media as a tool for receiving health awareness messages and consequently following its instructions for the individual's well-being. The authors adapted the key constructs of 3 theoretical foundations: (1) TAM, (2) TTF, and (3) prospect theory. The TAM serves as a concrete base to develop the conceptual framework. The TTF offers a key element of social media characteristics, whereas the prospect theory provides a theoretical framework for designing such messages. The proposed framework, therefore, will help in designing health messages that will be spread via social media apps.

Figure 1 illustrates smart health awareness message framework, where the authors proposed different hypotheses that provide a statement based on the feature extracted from the intended theory to represent a specific state of an individual's beliefs. This is to be used in a survey to obtain public perspectives on the use of social media technology in receiving health messages. First, they hypothesized that intention to use is influenced by perceived ease of use, perceived usefulness, perceived trust, gain-framed message, and loss-framed message (hypotheses H1, H3, H6, H9, and H10). Second, the authors hypothesized that perceived usefulness is impacted by perceived ease of use, customization, perceived trust, and technology characteristics (H2, H5, H7, and H8). Finally, it was hypothesized that perceived ease of use is influenced by customization (H4). Each defined hypothesis supports the relationships among the constructs of the framework. The following subsections present in detail the constructs of smart health awareness message framework.

Figure 1. Smart Health Awareness Message Framework.



Technology Perceptions

Technology perceptions include the key elements of the TAM, namely, perceived ease of use and perceived usefulness.

Perceived ease of use, as proposed by Davis [23], alludes to the extent to which the user of the technology will think that the use of a certain tool will be easy or free of effort. As social media supports easy access, easy engagement with its interactions (ie, reply, like, retweet, and repost) [42,43], and

XSL•FO
RenderX

easy navigation, people's intention will be enhanced to receive and apply health messages accordingly [44]. Exposure to such interactions is generally dynamic, resulting in encouraging health care professionals to create online communities where medical knowledge can be easily shared and freely accessed [38].

A positive association was supported between perceived ease of use and usefulness of technology usage that involved different contexts [45-47]. Therefore, H1 states that perceived ease of use of social media positively influences people's intention to receive and follow health awareness messages. H2 states that perceived ease of use positively influences perceived usefulness of social media to receive and follow health awareness messages.

Perceived usefulness is widely defined as ''the degree to which an individual believes that using a particular system would enhance his/her job performance'' [23]. In this study, perceived usefulness refers to the degree to which an individual considers that the benefits of social media will enhance his or her intention to receive and then follow health awareness messages. Some individuals may perceive social media as a personal digital assistant with the purpose of improving medical usage [48].

The link between the usefulness of social media and the intention to adopt such technology as a means to acquire and share health information has been explored by a number of studies [47,49]. Moreover, Deng et al [50] investigated the association between perceived usefulness and the individual's intention to adapt to mobile health. Therefore, H3 states that perceived usefulness of using social media will positively influence people's intention to receive and follow health awareness messages.

Technology-Influencing Factors

This section includes 3 technology factors, customization, perceived trust, and technology characteristics, which influence the overall perceptions of social media to receive and follow health awareness messages.

Message customization means reaching target people with individualized health messages that work well to engage with the messages effectively [51]. Patrick et al [52] found that people who were exposed to customized text messages with the purpose of promoting dietary behaviors were more likely to achieve weight loss compared with those related to printed materials. Customization in the conceptual framework refers to the empowerment that enables intended systems to understand its users' demographics and interest topics and then tailor their health messages accordingly, for example, the preferable time, the frequency of the messages, the type of disease, and the type of social media platforms. Social media customization provides a number of features that encourage users to prioritize particular accounts to view and act accordingly. On Twitter, Instagram, Facebook, and WhatsApp, by selecting get notifications, a user will automatically be notified every time these accounts post. Thus, such features allow users to easily access and efficiently follow social media posts. By leveraging social media customization efficiently, such technology can be harnessed to be most instrumental and useful in practice.

Customization correlates to perceived usefulness, as evidenced by Ho [53], who found that customized information technology services offer considerable benefits to customers that involve producing right content and format at the proper time for usage motivation purposes. In addition, customization in web-based interfaces has optimized web-based shopping due to its ease of use [54]. Therefore, H4 states that customization will have a positive impact on perceived ease of use of social media platforms to receive and follow health awareness messages. H5 states that customization will have a positive impact on perceived usefulness of social media platforms to receive and follow health awareness messages.

McAllister [55] defined trust as "the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another." He studied the trust between patients and the source of health-related information. Undoubtedly, the trust that individuals place in web-based health knowledge varies with the source of such knowledge. Thus, trust is a significant factor that affects people's adoption of health awareness messages [56,57]. Perceived trust in the framework will evaluate the contextual part of health-related information, and perceived usefulness will evaluate the practical part of social media use for passing health messages.

The association between trust and perceived usefulness has been discussed in several studies, confirming that the more the user perceives the technology to be useful, the greater the likelihood of trusting the content of such technology and therefore their intention to use it [58-60]. Thus, H6 states that perceived trust will positively influence people's intention to use social media to receive and follow health awareness messages. H7 states that perceived trust will positively impact perceived usefulness of social media to receive and follow health awareness messages.

Technology characteristics constitute a key element of the TTF model identified by Goodhe and Thompson [24], which refers to the extent to which a technology fits when the required tasks are met. The more individuals perceive that the technology suitably fits the intended tasks, the greater the likelihood that they will use that technology [61]. TTF has been used to measure social media appropriateness in many topics, such as sharing information about flood anticipation [62]. A previous study proposed that sharing is a fundamental characteristic that social media provides [63], which refers to the extent to which content can be exchanged among users [64]. For Facebook, Twitter, and Instagram, qualities that correspond to sharing are the message itself and media representation, which involves photographs and videos in terms of photograph quantity and video length [22].

The suitability of TTF depends on the user selection of the technology, which is based on technology characteristics that perfectly correspond to the task's attributes. Hence, this research demonstrates that social media features are capable of boosting the adaptation of this technology in viewing health awareness messages. Earlier studies have investigated the relationship between TTF constructs and perceived usefulness of using SMS for health awareness purposes [65]. Therefore, H8 states that technology characteristics will positively impact perceived usefulness of social media to view health awareness messages.

XSL•FO
RenderX

Message Design

This section presents a technique that aids in designing health awareness messages through prospect theory.

In loss- and gain-framed message design, health messages that aim at a particular behavior in terms of its benefits (gains) or costs (losses) play a significant role in health communication [57,66]. Health messages, therefore, might be designed either to emphasize the benefits of complying with the message content or the consequences of failure to comply with it [67,68]. For instance, a gain-framed message targeting increasing water intake could be *drinking a lot of water daily can help you lose weight*. In contrast, a loss-framed message could be *not drinking enough water causes you to gain weight* [69]. According to the prospect theory, the associated persuasion of gain- and loss-framed appeals is linked to the level of risk involved in the relevant actions [25]. The more individuals believe that they are at risk, the more motivated they are to the loss-framed message [57]. Therefore, the effectiveness of gain- and loss-framed messages varies depending on the goal of the message, either preventing or detecting health problems [66,70]. Therefore, positive messages are manifested to be more powerful in disease prevention messages [68], such as skin cancer prevention [66], whereas loss-framed messages are more likely to be useful in disease detection [71]. A number of studies have examined the effectiveness of negative messages in designing persuasive health messages. Meyerowitz and Chaiken [71] investigated the issue of women's breast self-assessment, indicating that female students were more encouraged to perform the assessment through passive messages rather than positive ones. Levin et al [72] developed a framing effect-based typology to check the influence of negative- and positive-framed information on decision makers. They concluded that passive goal framing was more convincing and influential than positive ones.

Message frame is believed to have a significant relationship with an individual's intention to adapt to technological invention [73]. Hence, H9 states that the positive effect of gain-framed messages on consumers 'intention to use social media for health awareness purposes would be stronger. On the other hand, hypothesis 10 states that the positive effect of loss-framed messages would be stronger.

Methods

Data Collection

The authors developed the questionnaire items based on an understanding of the literature, as presented in Multimedia Appendix 1. Before conducting the survey, it was validated by 5 experts in different industrial and research fields, including community medicine consultants, family medicine consultants, oncologists, and public health specialists based in Saudi Arabia, the United Kingdom, and the United Arab Emirates. Including qualified experts' opinions will assure that items are clarified, accurate, and free of confusion. The survey was distributed on the web through Qualtrics over a 2-month period in 2019, and it produced 701 responses from different countries, with 391 completed surveys.

Development of Questionnaire Items

The questionnaire included 3 parts: the first part presented the survey's introduction and consent form, the second part focused on the participant's demographics, as shown in Table 1, and the third part included 27 items. Each item is a statement that has been adapted from the literature to measure the opinion of the end user regarding the 7 identified constructs of the conceptual framework. List of items are presented in Multimedia Appendix 1; both items and scales were adapted from previous studies with some modifications to fit the research context. Perceived ease of use and intention to use social media items were adapted from Hong et al [74]. The items of perceived usefulness on the intention to use social media were borrowed from El-Wajeeh et al [75]. Items on customization were adapted from Bandyopadhyay et al [65]. The items on the perceived trust construct were adapted from El-Wajeeh et al [75], and items on technology characteristics were derived from Bandyopadhyay et al [65] and Zaini et al [76]. All questionnaire items were rated using a 5-point Likert scale, ranging from *strongly disagree* (1) to *strongly agree* (5), in which participants were required to choose the most suitable answer.

XSL•FO
RenderX

Table 1. Respondent demographics (N=391).

Measure	Values, n (%)
Gender	
Male	154 (39.4)
Female	237 (60.6[a])
Age (years)	
20-29	91 (23.3)
30-39	142 (36.3[a])
40-49	64 (16.4)
50-59	52 (13.3)
≥60	42 (10.7)
Level of education	
Secondary school	9 (2.3)
Bachelor's degree	125 (32.0)
Master's degree or above	227 (58.1[a])
Others	30 (7.7)
Job	
Governmental employee	184 (47.1[a])
Private employee	85 (21.7)
Self-employed	21 (5.4)
I do not work	101 (25.8)
Frequency of using social media during the week	
Always	74 (18.9)
Very often	75 (19.2)
Often	121 (31.0[a])
Hardly often	94 (24.0)
Never	27 (6.9)
Frequency of using social media for seeking health information (years)	
<2	121 (31.0[a])
2 to <4	84 (21.5)
4-6	94 (24.0)
>6	92 (23.5)

[a]Indicates the highest percentage.

Respondent Profile and Descriptive Statistics

The respondents' demographics are illustrated in Table 1. Of the 391 participants, 121 used social media platforms often during the week, with a percentage of 30.9% (Table 1). In total, 18.9% (74/391) and 19.2% (75/391) of the participants used them always and very often during the week, respectively (Table 1). Conversely, 6.9% (27/391) of the participants never used social media (Table 1). In addition, 24.0% (94/391) of the respondents utilized these platforms to seek health information (Table 1). This is due to several reasons, including easy and free access to social media, with no physical existence requirement as with health care centers, and no storage capacity is needed as with printed media.

Results

Data Analysis

Smart health awareness message framework has been proposed to elicit the opinion of the end user about different constructs, and the results of the survey require a range of statistical methods. First, SPSS (version 25; IBM Corp) was used to acquire respondents' descriptive statistics. Then, data were analyzed using the IBM SPSS Analysis of a Moment Structures

XSL•FO
RenderX

(AMOS) version 25, which requires 2 stages of assessment: measurement model assessment and structural equation modeling (SEM) assessment. The measurement model was assessed to confirm that the survey items reflected the corresponding constructs of the conceptual framework [77]. SEM was used to test hypothesized relationships among the constructs after conducting confirmatory factor analysis (CFA). The following subsections present the results of exploratory factor analysis (EFA), which includes a measurement model followed by SEM.

Measurement Model

In the first stage, an EFA was conducted to determine the correlation among observed variables or items being tested. A correlation matrix presented in Multimedia Appendix 2 shows the internal correlations between variables, which are higher than ±0.3, and not exceeding the cut-off threshold, ±0.8, refers to the absence of multicollinearity [77]. Then, EFA was conducted and provided a factor structure of 27 items (Multimedia Appendix 1). These variables are grouped into 7 factors: perceived ease of use, perceived usefulness, customization, perceived trust, technology characteristics, gain- and loss-framed message, and intention to use. The factor analysis results are illustrated in Table 2 using maximum likelihood with a promax rotation of data. This analysis shows a clean factor loading pattern, no major cross loading, where values ranged between 0.3 and 0.8, cut-off criteria [78].

Another issue to be considered in EFA is the appropriateness of the data set that has been verified using the Kaiser-Meyer-Olkin (KMO) statistics and Bartlett test of sphericity. According to Kaiser [79], the KMO value is recommended to be greater than 0.7 to obtain meaningful and good EFA. To assure the factorability of the correlation matrix among variables, the Bartlett test value should be significant (P<.01) [80]. The KMO yielded data adequacy with a value of 0.80, and the sphericity test showed a statistically significant χ^2_{190}=2467.0 (P<.01; Multimedia Appendix 3). Thus, it is evident that the factorability of the correlation matrix is adequate. Then, construct reliability was measured by Cronbach alpha (CA), composite reliability (CR), and average variance extracted (AVE). CA was .893 for the total items; thus, the value was greater than the recommended .7 [78]. Table 3 presents CA for each construct, ranging between .733 and .826, leading to fit reliabilities of the data. Convergent validity can be assessed by calculating the average variance extracted and CR where the values should be greater than 0.5 and 0.7, respectively [81]. The results in Table 3 reveal that the AVE and CR applied such criteria. Although the AVE of technology characteristics is below the recommended value, Fornell and Larcker [82] confirmed that a researcher may conclude that the convergent validity of the construct is adequate, as CR is higher than the acceptable range.

Discriminant validity refers to the extent to which the constructs are varied from each other, which can be assessed using the Fornell-Larcker criterion [82]. In this method, the square root of AVE is compared with the correlation of constructs or variables. The variance between constructs and their items should exceed the variance explained with other constructs [82]. Table 4 illustrates that all diagonal square roots of the AVEs were higher than the off-diagonal values, which present constructs' correlations. Given the adequate reliability and acceptable convergent and discriminant validities, it is concluded that the measurement model is satisfactory.

XSL•FO
RenderX

Table 2. Promax matrix showing factor analysis results.

Factor[a,b]	1	2	3	4	5	6	7
	PU [c]	PEU[d]	PT[e]	TECH[f]	CUST[g]	INT[h]	Message[i]
PU1	0.406	N/A[j]	N/A	N/A	N/A	N/A	N/A
PU2	0.512	N/A	N/A	N/A	N/A	N/A	N/A
PEU1	N/A	0.789	N/A	N/A	N/A	N/A	N/A
PEU2	N/A	0.738	N/A	N/A	N/A	N/A	N/A
PEU3	N/A	0.644	N/A	N/A	N/A	N/A	N/A
PEU4	N/A	0.562	N/A	N/A	N/A	N/A	N/A
PT1	N/A	N/A	0.596	N/A	N/A	N/A	N/A
PT2	N/A	N/A	0.839	N/A	N/A	N/A	N/A
TECH1	N/A	N/A	N/A	0.379	N/A	N/A	N/A
TECH2	N/A	N/A	N/A	0.791	N/A	N/A	N/A
TECH3	N/A	N/A	N/A	0.769	N/A	N/A	N/A
TECH4	N/A	N/A	N/A	0.379	N/A	N/A	N/A
TECH5	N/A	N/A	N/A	0.720	N/A	N/A	N/A
TECH6	N/A	N/A	N/A	0.764	N/A	N/A	N/A
TECH7	N/A	N/A	N/A	0.725	N/A	N/A	N/A
CUST1	N/A	N/A	N/A	N/A	0.821	N/A	N/A
CUST2	N/A	N/A	N/A	N/A	0.845	N/A	N/A
CUST3	N/A	N/A	N/A	N/A	0.411	N/A	N/A
CUST4	N/A	N/A	N/A	N/A	0.301	N/A	N/A
INT1	N/A	N/A	N/A	N/A	N/A	0.752	N/A
INT2	N/A	N/A	N/A	N/A	N/A	0.783	N/A
INT3	N/A	N/A	N/A	N/A	N/A	0.596	N/A
Message1	N/A	N/A	N/A	N/A	N/A	N/A	0.723
Message2	N/A	N/A	N/A	N/A	N/A	N/A	0.735
Message3	N/A	N/A	N/A	N/A	N/A	N/A	0.583
Message4	N/A	N/A	N/A	N/A	N/A	N/A	0.536
Message5	N/A	N/A	N/A	N/A	N/A	N/A	0.500

[a]Rotation converged in 7 iterations.

[b]Extraction method: maximum likelihood; rotation method: Promax with Kaiser normalization.

[c]PU: perceived usefulness.

[d]PEU: perceived ease of use.

[e]PT: perceived trust.

[f]TECH: technology characteristics.

[g]CUST: customization.

[h]INT: intention to use.

[i]Message: gain- and loss- framed message.

[j]N/A: not applicable.

XSL•FO
RenderX

Table 3. Cronbach alpha, composite reliability, and average variance extracted for the constructs.

Constructs and items	CA[a]	CR[b]	AVE[c]	Factor loading
PEU[d]	**.83**	**0.69**	**0.53**	
PEU1				0.76
PEU2				0.69
PEU3				0.80
PEU4				0.78
PU[e]	**.80**	**0.80**	**0.66**	
PU1				0.80
PU2				0.83
CUST[f]	**.82**	**0.82**	**0.70**	
CUST1				0.78
CUST2				0.89
CUST3				0.49
CUST4				0.44
PT[g]	**.76**	**0.71**	**0.55**	
PT1				0.78
PT2				0.69
TECH[h]	**.75**	**0.71**	**0.38**	
TECH1				0.60
TECH2				0.63
TECH3				0.68
TECH4				0.55
TECH5				0.63
TECH6				0.55
TECH7				0.43
Message[i]	**.73**	**0.75**	**0.50**	
Message1				0.81
Message2				0.72
Message3				0.57
Message4				0.53
Message5				0.48
INT[j]	**.76**	**0.77**	**0.62**	
INT1				0.82
INT2				0.76
INT3				0.57

[a]CA: Cronbach alpha.

[b]CR: composite reliability.

[c]AVE: average variance extracted.

[d]PEU: perceived ease of use.

[e]PU: perceived usefulness.

[f]CUST: customization.

[g]PT: perceived trust.

[h]TECH: technology characteristics.

XSL•FO
RenderX

[i]Message: gain-loss framed message.

[j]INT: intention to use.

Table 4. Discriminant validity.

Factors[a]	PU[b]	PEU[c]	PT[d]	TECH[e]	CUST[f]	INT[g]	Message[h]
PU	*0.81*	N/A[i]	N/A	N/A	N/A	N/A	N/A
PEU	0.72[j]	*0.73*	N/A	N/A	N/A	N/A	N/A
PT	0.50[j]	0.46[j]	*0.74*	N/A	N/A	N/A	N/A
TECH	0.57[j]	0.66[j]	0.59[j]	*0.62*	N/A	N/A	N/A
CUST	0.29[j]	0.27[j]	0.36[j]	0.32[j]	*0.84*	N/A	N/A
INT	0.74[j]	0.50[j]	0.41[j]	0.58[j]	0.27[j]	*0.79*	N/A
Message	−0.09	−0.11	−0.03	−0.07	−0.001	0.02	*0.71*

[a]Off-diagonal elements are correlations, and diagonal elements are square roots of the average variance extracted.

[b]PU: perceived usefulness.

[c]PEU: perceived ease of use.

[d]PT: perceived trust.

[e]TECH: technology characteristics.

[f]CUST: customization.

[g]INT: intention to use.

[h]Message: gain-loss framed message.

[i]N/A: not applicable.

[j]0.27: significance of correlations $P<.001$.

Structural Equation Modeling

In the second stage, CFA was conducted before testing the hypothesized relationships among the constructs in smart health awareness message framework using SEM [83]. To proceed with CFA, standardized loadings for each item were obtained, in which these values should be at least 0.5 or ideally 0.7 or higher [78]. As shown in Table 5, of the 27 items, CUST3, CUST4, TECH7, and Message5 are attributed to deletion from the research model because of their lower loadings, whereas others are well related to their associated constructs. Given the significant standardized residual covariances, which means the largest values (in absolute value) for items PEU3, PEU4, TECH5, TECH6, Message4, and INT3, they require removal as they affect the goodness fit of the model [84] (Multimedia Appendix 4). The analysis illustrated in Multimedia Appendix 5 confirmed that the linear regression model is adequately fit, with χ^2_{104} value of 299.0 and $P<.001$.

In the second step of the CFA, model fit indexes were measured: χ^2 divided by *df*, root mean square error of approximation (RMSEA), normed fit index (NFI), incremental fit index (IFI), comparative fit index (CFI), and Tucker-Lewis Index (TLI) [83]. The CFA results showed an acceptable fit model ($\chi^2_{97}=145.8$; χ^2 divided by $df=1.503$; RMSEA=0.036; NFI=0.937; IFI=0.978; CFI=0.978; TLI=0.969; Multimedia Appendix 6). The results also confirm that the linear and covariance fit models meet the standards, thus emphasizing the acceptance of model fit (Multimedia Appendices 7 and 8, respectively).

The next step is measuring the path coefficient, coefficient of determination, and *t* value using SEM. A path coefficient or path analysis indicates the relationships among the constructs. The coefficient of determination (R2) is a measure of the percentage of the total variation of the dependent variable that is explained or predicted by the independent variable(s) or predictor(s) [78]. The larger the value of the coefficient of determination, the greater the prediction of the dependent variable. Table 5 illustrates the path analysis and hypotheses testing. The results show that the coefficient of determination (R2) is 0.642 for the *intention to use* construct. This means that the 4 constructs (perceived ease of use, perceived usefulness, perceived trust, gain-framed message, and loss-framed message) moderately explain 64.2% of the variance in intention to use social media. Perceived ease of use, together with customization, perceived trust, and technology characteristics, explain 55.4% of the variance in perceived usefulness. Finally, customization explains 26.2% of the variance in perceived ease of use.

XSL•FO
RenderX

Table 5. Summary of testing hypotheses.

Hypothesis	Hypothesized path	Beta[a]	*P* value	Result
H1	PEU[b]-INT[c]	.05	.43	Not supported
H2	PEU-PU[d]	.37	<.001	Supported
H3	PU-INT	.43	<.001	Supported
H4	CUST[e]-PEU	.12	.12	Not supported
H5	CUST-PU	.16	.05	Supported
H6	PT[f]-INT	.11	.08	Not supported
H7	PT-PU	.07	<.001	Supported
H8	TECH[g]-PU	.12	<.001	Supported
H9	Gain-framed message-INT	.04	<.001	Supported
H10	Loss-framed message-INT	.08	<.001	Supported

[a]Beta is standardized.

[b]PEU: perceived ease of use.

[c]INT: intention to use.

[d]PU: perceived usefulness.

[e]CUST: customization.

[f]PT: perceived trust.

[g]TECH: technology characteristics.

Path analysis results also reveal that perceived ease of usefulness has little effect on intention to use (β=.05; *P*=.43), unlike the significant effect on perceived usefulness (β=.37; *P*<.001). Thus, H1 is not supported, whereas H2 is supported. The impact of perceived usefulness on intention to use is significant (β=.43; *P*<.001), supporting H3. However, the results indicate that customization has no significant impact on perceived ease of use (β=.12; *P*=.12), whereas it has a significant impact on perceived usefulness (β=.16; *P*=.05). Hence, H4 is not supported, and H5 is supported. The results also indicate that there is no significant impact between perceived trust and intention to use (β=.11; *P*=.07). In contrast, perceived trust significantly impacts perceived usefulness (β=.07; *P*<.001). Therefore, H6 is not supported, whereas H7 is supported. Technology characteristics are considered to be related and have a significant impact on perceived usefulness (β=.12; *P*<.001), lending support to H8. Finally, it was found that gain-framed messages and loss-framed messages are significantly and positively related to intention to use social media (β=.04; *P*<.001) and (β=.08; *P*<.001), respectively. Thus, H9 and H10 are supported.

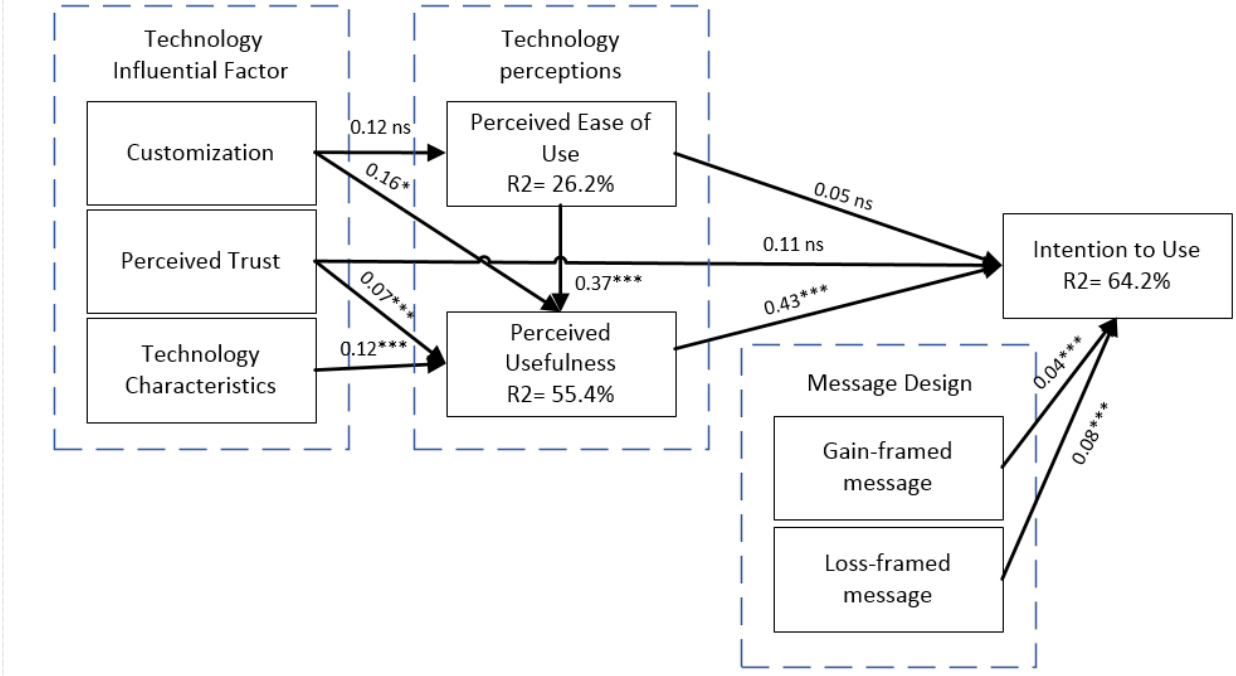
Discussion

Principal Findings

Nowadays, social media plays a considerable role in an individual's daily routine, as it provides different features that encourage people to adapt it for a range of uses, including health promotion. Therefore, the motivation of this paper was to examine the factors that affect people's intention to use social media as a way of receiving health awareness messages, which, in turn, will help them to maintain their diet and reduce the incidence of diseases. In turn, the challenges that arise from printed media, involving paper and power consumption, storage capacity, and labor intensity, will be reduced. The results in Table 1 show that 69.0% (270/ 391) of the public surveyed used social media always, often, or very often, and 31.0% (121/391) of them hardly or never used social media. This indicates that eHealth that involves using social media to convey health messages has the potential to reach about 70% of the public. Findings (shown in Figure 2) confirm that perceived usefulness and message design relating to health message frames (positive vs negative) are the leading predictors of people's intention to use social media in the health promotion context. Loss-framed messages have been examined in previous studies [85,86] to be a motivating factor that influences people to engage and comply with health behavior on social media. A recent study has come to an opposing conclusion where people are encouraged to acquire and trust health information on social media when they are exposed to gain-framed messages [87]. Regarding social media usefulness, the results of this study are consistent with those of Lin and Ho [49], indicating that perceived usefulness significantly affects people toward social media adaptation in sharing health information.

XSL•FO
RenderX

Figure 2. Hypothesized Smart Health Awareness Message Framework. *P<.05, **P<.01, ***P<.001; ns: nonsignificant.



In addition, health message customization encourages the prediction of perceived usefulness, whereas it has no effect on perceived ease of use of social media. Thus, it can be indicated that social media users perceive the acquired benefits from social media when they receive health messages tailored to their preferences [51]. Regarding the relationship between technology characteristics and perceived usefulness (β=.12; P<.001), it can be concluded that the higher the characteristics offered by social media, the greater the perceptions of the benefits of adopting health messages received via social media. These characteristics involve hyperlinks and hashtags provision, which, in turn, generate higher engagement with messages [88]. Moreover, posting photographs and videos that enhance the message by being more visual encourage individuals to adopt social media for health promotion [15]. The results reveal that the design of health messages plays a significant role in people's intention to use social media. Consequently, the key factors specified are essential for health organizations to promote eHealth by developing and spreading health messages effectively, which, in turn, will enhance people's health.

Conclusions

The study's results demonstrate the use of social media in health promotion purposes, which will enhance the outcomes of an individual's well-being. This paper aimed to investigate the influential factors that affect people's intention to adopt such technology in health communication campaigns. Undoubtedly, high levels of health message success cannot be achieved without emotions embedded in the content of health messages [89]. The study's findings indicate that health message frames would be efficacious in improving public health communication toward social media adaptation. Furthermore, perceived usefulness has an impact on people's intention to adapt to social media to acquire health awareness information. These results can be explained by the adaptation of the TAM and the prospect theory.

Given the findings of smart health awareness message framework, designing health awareness messages to include loss- or gain-framed content to evoke high emotions might contribute to boosting the effectiveness of health promotion interventions. Hence, this study offers implications for health awareness message developers that guide them to establish materials that are more patient friendly and technologically outstanding by adapting social media as a delivery method. Accordingly, this strategy will encourage individuals to exchange these messages among social media users.

Limitations and Future Work

This study has several limitations and indicates several directions for future work. First, for the construct of message design, there are few studies associated with the prospect theory that examine the public perspective in terms of their preferences. Thus, the authors developed a number of items, validated by experts, and adapted in this study to ensure the validity of the construct. Future works might examine this construct more broadly to determine the extent to which the public might receive this message in a more positive or negative manner. Second, although the study involved 391 respondents from different countries, in which sample size is convenient for testing the framework, future studies with larger samples are needed to reinforce the generalization of results. In addition, the participants were English speakers, and findings related to a particular language might restrict generalization to others. Thus, future research might duplicate this study with different languages.

Smart health awareness message framework will also be used to define the right content and format of the health awareness messages to be spread via a software system that is integrated with different social media platforms. Furthermore, a computer-based knowledge framework based on the use of social media apps will be developed to spread health awareness messages. Finally, a specific statistical technique will be used

XSL•FO
RenderX

to validate the impact of the health awareness message on recipients.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Items of the study's constructs.
[DOCX File , 17 KB - jmir_v22i7e16212_app1.docx]

Multimedia Appendix 2
Items correlation matrix.
[DOCX File , 158 KB - jmir_v22i7e16212_app2.docx]

Multimedia Appendix 3
Kaiser-Meyer-Olkin and Bartlett test.
[DOCX File , 12 KB - jmir_v22i7e16212_app3.docx]

Multimedia Appendix 4
Standardized residual covariances for deleted items.
[DOCX File , 31 KB - jmir_v22i7e16212_app4.docx]

Multimedia Appendix 5
Standardized estimate of linear regression.
[DOCX File , 174 KB - jmir_v22i7e16212_app5.docx]

Multimedia Appendix 6
Model fit summary.
[DOCX File , 13 KB - jmir_v22i7e16212_app6.docx]

Multimedia Appendix 7
Covariance fit model.
[DOCX File , 14 KB - jmir_v22i7e16212_app7.docx]

Multimedia Appendix 8
Standardized regression weights among items.
[DOCX File , 13 KB - jmir_v22i7e16212_app8.docx]

References

1. Ritterband LM, Borowitz S, Cox DJ, Kovatchev B, Walker LS, Lucas V, et al. Using the internet to provide information prescriptions. Pediatrics 2005 Nov;116(5):e643-e647. [doi: 10.1542/peds.2005-0404] [Medline: 16263978]
2. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
3. Syn SY, Kim SU. College students' health information activities on Facebook: investigating the impacts of health topic sensitivity, information sources, and demographics. J Health Commun 2016 Jul;21(7):743-754. [doi: 10.1080/10810730.2016.1157652] [Medline: 27220029]
4. Salem J, Borgmann H, Murphy DG. Integrating social media into urologic health care: what can we learn from other disciplines? Curr Urol Rep 2016 Feb;17(13):13. [doi: 10.1007/s11934-015-0570-2] [Medline: 26757907]
5. Brusilovskiy E, Townley G, Snethen G, Salzer MS. Social media use, community participation and psychological well-being among individuals with serious mental illnesses. Comput Human Behav 2016 Dec;65:232-240. [doi: 10.1016/j.chb.2016.08.036]
6. World Health Organization. National EHealth Strategy Toolkit. Geneva, Switzerland: WHO Publications; 2012.
7. Albalawi YA, Sixsmith J. Exploring the diffusion of tweets designed to raise the road safety agenda in Saudi Arabia. Glob Health Promot 2017 Dec;24(4):5-13. [doi: 10.1177/1757975915626111] [Medline: 27251329]

XSL•FO
RenderX

8. Thabrew H, Stasiak K, Hetrick SE, Wong S, Huss JH, Merry SN. E-health interventions for anxiety and depression in children and adolescents with long-term physical conditions. Cochrane Database Syst Rev 2018 Aug 15;8:CD012489 [FREE Full text] [doi: 10.1002/14651858.CD012489.pub2] [Medline: 30110718]
9. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]
10. Boles M, Adams A, Gredler A, Manhas S. Ability of a mass media campaign to influence knowledge, attitudes, and behaviors about sugary drinks and obesity. Prev Med 2014 Oct;67(Suppl 1):S40-S45 [FREE Full text] [doi: 10.1016/j.ypmed.2014.07.023] [Medline: 25066020]
11. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
12. Austin LS, Husted K. Cost-effectiveness of television, radio, and print media programs for public mental health education. Psychiatr Serv 1998 Jun;49(6):808-811. [doi: 10.1176/ps.49.6.808] [Medline: 9634162]
13. Randolph W, Viswanath K. Lessons learned from public health mass media campaigns: marketing health in a crowded media world. Annu Rev Public Health 2004;25(1):419-437. [doi: 10.1146/annurev.publhealth.25.101802.123046] [Medline: 15015928]
14. Ba S, Wang L. Digital health communities: the effect of their motivation mechanisms. Decis Support Syst 2013 Nov;55(4):941-947. [doi: 10.1016/j.dss.2013.01.003]
15. Li Y, Wang X, Lin X, Hajli M. Seeking and sharing health information on social media: a net valence model and cross-cultural comparison. Technol Forecast Soc Chang 2018 Jan;126:28-40. [doi: 10.1016/j.techfore.2016.07.021]
16. Lyson HC, Le GM, Zhang J, Rivadeneira N, Lyles C, Radcliffe K, et al. Social media as a tool to promote health awareness: results from an online cervical cancer prevention study. J Cancer Educ 2019 Aug;34(4):819-822. [doi: 10.1007/s13187-018-1379-8] [Medline: 29948924]
17. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 7;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
19. Mouakket S. Factors influencing continuance intention to use social network sites: the Facebook case. Comput Human Behav 2015 Dec;53:102-110. [doi: 10.1016/j.chb.2015.06.045]
20. Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. J Med Internet Res 2015 Jun 11;17(6):e141 [FREE Full text] [doi: 10.2196/jmir.3662] [Medline: 26068087]
21. Welch V, Petkovic J, Pardo JP, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
22. Huo C, Zhang M, Ma F. Factors influencing people's health knowledge adoption in social media: the mediating effect of trust and the moderating effect of health threat. Libr Hi Tech 2018 Mar 19;36(1):129-151. [doi: 10.1108/lht-04-2017-0074]
23. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-313. [doi: 10.2307/249008]
24. Goodhue DL, Thompson RL. A mucoprotein derived from human urine which reacts with influenza, mumps, and Newcastle disease viruses. MIS Q 1995;19(2):71-97. [Medline: 14907962]
25. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
26. Bernhardt JM. Communication at the core of effective public health. Am J Public Health 2004 Dec;94(12):2051-2053. [doi: 10.2105/ajph.94.12.2051] [Medline: 15569948]
27. Frati A, Luzi AM, Colucci A. Communication for health promotion: history and identification of effective methods. Ann Ist Super Sanita 2010;46(4):422-426 [FREE Full text] [doi: 10.4415/ANN_10_04_10] [Medline: 21169674]
28. Catalán-Matamoros D. The role of mass media in public health. Am J Public Health Nations Health 1960 Apr;50:515-523. [doi: 10.2105/ajph.50.4.515] [Medline: 13829314]
29. Sims M, Langley T, Lewis S, Richardson S, Szatkowski L, McNeill A, et al. Effectiveness of tobacco control television advertisements with different types of emotional content on tobacco use in England, 2004-2010. Tob Control 2016 Jan;25(1):21-26 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051454] [Medline: 25037155]
30. Farrelly MC, Hussin A, Bauer UE. Effectiveness and cost effectiveness of television, radio and print advertisements in promoting the New York smokers' quitline. Tob Control 2007 Dec;16(Suppl 1):i21-i23 [FREE Full text] [doi: 10.1136/tc.2007.019984] [Medline: 18048625]
31. Macfarlane JT, Holmes WF, Macfarlane RM. Reducing reconsultations for acute lower respiratory tract illness with an information leaflet: a randomized controlled study of patients in primary care. Br J Gen Pract 1997 Nov;47(424):719-722 [FREE Full text] [Medline: 9519518]

XSL•FO
RenderX

32. Steele M, Dow L, Baxter G. Promoting public awareness of the links between lifestyle and cancer: a controlled study of the usability of health information leaflets. Int J Med Inform 2011 Dec;80(12):e214-e229. [doi: 10.1016/j.ijmedinf.2011.08.012] [Medline: 22020199]
33. Jenner EA, Jones F, Fletcher BC, Miller L, Scott GM. Hand hygiene posters: selling the message. J Hosp Infect 2005 Feb;59(2):77-82. [doi: 10.1016/j.jhin.2004.07.002] [Medline: 15702513]
34. Winterbottom A, Conner M, Mooney A, Bekker H. Evaluating the quality of patient leaflets about renal replacement therapy across UK renal units. Nephrol Dial Transplant 2007 Aug;22(8):2291-2296. [doi: 10.1093/ndt/gfm095] [Medline: 17369614]
35. McNulty CA, Nichols T, Boyle PJ, Woodhead M, Davey P. The English antibiotic awareness campaigns: did they change the public's knowledge of and attitudes to antibiotic use? J Antimicrob Chemother 2010 Jul;65(7):1526-1533. [doi: 10.1093/jac/dkq126] [Medline: 20488985]
36. Lee C. The interplay between media use and interpersonal communication in the context of healthy lifestyle behaviors: reinforcing or substituting? Mass Commun Soc 2009 Dec;13(1):48-66 [FREE Full text] [doi: 10.1080/15205430802694869] [Medline: 25598709]
37. Were MC, Kamano JH, Vedanthan R. Leveraging digital health for global chronic diseases. Glob Heart 2016 Dec;11(4):459-462. [doi: 10.1016/j.gheart.2016.10.017] [Medline: 27938842]
38. Roland D, Spurr J, Cabrera D. Preliminary evidence for the emergence of a health care online community of practice: using a netnographic framework for Twitter hashtag analytics. J Med Internet Res 2017 Jul 14;19(7):e252 [FREE Full text] [doi: 10.2196/jmir.7072] [Medline: 28710054]
39. Diddi P, Lundy LK. Organizational twitter use: content analysis of tweets during breast cancer awareness month. J Health Commun 2017 Mar;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
40. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]
41. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
42. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: new perspectives. Sport Educ Soc 2019;24(7):673-688 [FREE Full text] [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]
43. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
44. Dumas A, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55 [FREE Full text] [doi: 10.2196/jmir.9230] [Medline: 29463487]
45. Yu P, Li H, Gagnon M. Health IT acceptance factors in long-term care facilities: a cross-sectional survey. Int J Med Inform 2009 Apr;78(4):219-229. [doi: 10.1016/j.ijmedinf.2008.07.006] [Medline: 18768345]
46. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electron Mark 2012 Dec 11;23(1):49-61. [doi: 10.1007/s12525-012-0112-4]
47. Ramírez-Correa P, Grandón EE, Ramírez-Santana M, Órdenes LB. Explaining the use of social network sites as seen by older adults: the enjoyment component of a hedonic information system. Int J Environ Res Public Health 2019 May 14;16(10):1673 [FREE Full text] [doi: 10.3390/ijerph16101673] [Medline: 31091670]
48. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: toward an integrative view. Inf Manag 2006 Apr;43(3):350-363. [doi: 10.1016/j.im.2005.08.006]
49. Lin H, Ho W. Cultural effects on use of online social media for health-related information acquisition and sharing in Taiwan. Int J Hum Comput Interact 2018 Apr 10;34(11):1063-1076. [doi: 10.1080/10447318.2017.1413790]
50. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mhealth services in China: empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
51. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]
52. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009 Jan 13;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
53. Ho SY. The effects of location personalization on individuals' intention to use mobile services. Decis Support Syst 2012 Nov;53(4):802-812. [doi: 10.1016/j.dss.2012.05.012]
54. Kumar RL, Smith MA, Bannerjee S. User interface features influencing overall ease of use and personalization. Inf Manag 2004 Jan;41(3):289-302. [doi: 10.1016/s0378-7206(03)00075-2]
55. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 1995 Feb 1;38(1):24-59. [doi: 10.2307/256727]

XSL•FO
RenderX

56. Ruppel EK. Scanning health information sources: applying and extending the comprehensive model of information seeking. J Health Commun 2016;21(2):208-216. [doi: 10.1080/10810730.2015.1058438] [Medline: 26716985]
57. Macapagal K, Janssen E, Matson M, Finn PR, Heiman JR. The impact of gain- and loss-framed messages on young adults' sexual decision making: an experimental study. Arch Sex Behav 2017 Feb;46(2):385-394 [FREE Full text] [doi: 10.1007/s10508-015-0679-x] [Medline: 26696408]
58. Geven D, Karahanna E, Straub DW. Trust and TAM in online shopping: an integrated model. MIS Q 2003 May;27(1):51-59. [doi: 10.2307/30036519]
59. Wu I, Chen J. An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: an empirical study. Int J Hum Comput Stud 2005 Jun;62(6):784-808. [doi: 10.1016/j.ijhcs.2005.03.003]
60. Chinomona R. The influence of perceived ease of use and perceived usefulness on trust and intention to use mobile social software. African J Phys Heal Educ Recreat Danc 2013;19(2):258-273 [FREE Full text]
61. Lippert SK, Forman H. A supply chain study of technology trust and antecedents to technology internalization consequences. Int Jnl Phys Dist Log Manage 2006 Apr;36(4):271-288. [doi: 10.1108/09600030610672046]
62. Hashim K, Ishak S, Ahmad M. A study on social media application as a tool to share information during flood disaster. J Eng Appl Sci 2015;10(3):959-967 [FREE Full text]
63. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Bus Horiz 2011 May;54(3):241-251. [doi: 10.1016/j.bushor.2011.01.005]
64. Lu H, Yang Y. Toward an understanding of the behavioral intention to use a social networking site: an extension of task-technology fit to social-technology fit. Comput Human Behav 2014;34:323-332. [doi: 10.1016/j.chb.2013.10.020]
65. Bandyopadhyay T, Meso P, Negash S. Mobile IT in health–the case of short messaging service in an HIV awareness program. Inf Technol Dev 2017 Aug 23;24(2):359-397. [doi: 10.1080/02681102.2017.1363029]
66. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
67. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [doi: 10.1037/0033-2909.121.1.3] [Medline: 9000890]
68. Latimer AE, Brawley LR, Bassett RL. A systematic review of three approaches for constructing physical activity messages: what messages work and what improvements are needed? Int J Behav Nutr Phys Act 2010 May 11;7:36 [FREE Full text] [doi: 10.1186/1479-5868-7-36] [Medline: 20459779]
69. Shamaskin AM, Mikels JA, Reed AE. Getting the message across: age differences in the positive and negative framing of health care messages. Psychol Aging 2010 Sep;25(3):746-751. [doi: 10.1037/a0018431] [Medline: 20677886]
70. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Framing of health information messages. Cochrane Database Syst Rev 2011 Dec 7;7(12):CD006777. [doi: 10.1002/14651858.CD006777.pub2] [Medline: 22161408]
71. Meyerowitz BE, Chaiken S. The effect of message framing on breast self-examination attitudes, intentions, and behavior. J Pers Soc Psychol 1987 Mar;52(3):500-510. [doi: 10.1037//0022-3514.52.3.500] [Medline: 3572721]
72. Levin IP, Schneider SL, Gaeth GJ. All frames are not created equal: a typology and critical analysis of framing effects. Organ Behav Hum Decis Process 1998 Nov;76(2):149-188. [doi: 10.1006/obhd.1998.2804] [Medline: 9831520]
73. Vishwanath A. From belief-importance to intention: the impact of framing on technology adoption. Commun Monogr 2009 May 13;76(2):177-206. [doi: 10.1080/03637750902828438]
74. Hong S, Thong J, Moon J, Tam K. Understanding the behavior of mobile data services consumers. Inf Syst Front 2008 Jun 12;10(4):431-445 [FREE Full text] [doi: 10.1007/s10796-008-9096-1]
75. El-Wajeeh M, Galal-Edeen G, Mokhtar H. Technology acceptance model for mobile health systems. J Mob Comput Appl 2014;1(1):21-33 [FREE Full text] [doi: 10.9790/0050-0112133]
76. Zaini M, Shahibi M, Kadir W, Ibrahim Z. Identification of trust factor in digital health information from the internet among clinic attendees. Int J Comput Inf Technol 2013;2(5):880-887 [FREE Full text]
77. Field A. Discovering Statistics Using SPSS. Third Edition. Thousand Oaks, CA: Sage Publications; 2009.
78. Hair J, Black W, Babin BJ, Anderson RE. Multivariate Data Analysis: With Readings. Seventh Edition. England, UK: Pearson; 2014.
79. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/bf02291575]
80. Tabachnick B, Fidell L. Using multivariate statistics. In: Using Multivariate Statistics. Sixth Edition. England, UK: Pearson; 2013.
81. MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Q 2011;35(2):293. [doi: 10.2307/23044045]
82. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
83. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
84. Maydeu-Olivares A, Shi D. Effect sizes of model misfit in structural equation models: standardized residual covariances and residual correlations. Methodology 2017 Jun;13(Supplement 1):23-30. [doi: 10.1027/1614-2241/a000129]

XSL•FO
RenderX

85. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
86. Lee-Won R, Na K, Coduto K. The effects of social media virality metrics, message framing, and perceived susceptibility on cancer screening intention: the mediating role of fear. Telemat Informatics 2017 Dec;34(8):1387-1397. [doi: 10.1016/j.tele.2017.06.002]
87. Borah P, Xiao X. The importance of 'likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
88. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on Twitter. PLoS One 2014;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
89. Nabi RL, Walter N, Oshidary N, Endacott CG, Love-Nichols J, Lew ZJ, et al. Can emotions capture the elusive gain-loss framing effect? A meta-analysis. Communic Res 2019 Jul 13;1:1-24. [doi: 10.1177/0093650219861256]

Abbreviations

AVE: average variance extracted
CA: Cronbach alpha
CFA: confirmatory factor analysis
CFI: comparative fit index
CR: composite reliability
EFA: exploratory factor analysis
eHealth: electronic health
IFI: incremental fit index
KMO: Kaiser-Meyer-Olkin
NFI: normed fit index
RMSEA: root mean square error of approximation
SEM: structural equation modeling
TAM: technology acceptance model
TLI: Tucker-Lewis Index
TTF: task technology fit

Edited by G Eysenbach; submitted 10.09.19; peer-reviewed by J Hazzam, JR Bautista; comments to author 03.10.19; revised version received 27.11.19; accepted 06.02.20; published 23.07.20.

Please cite as:
Alsisi EA, Al-Ashaab A, Abualfaraa WA
The Development of a Smart Health Awareness Message Framework Based on the Use of Social Media: Quantitative Study
J Med Internet Res 2020;22(7):e16212
URL: https://www.jmir.org/2020/7/e16212
doi:10.2196/16212
PMID:32459627

©Elaf Ali Alsisi, Ahmed Al-Ashaab, Wadhah Ahmed Abualfaraa. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Sudden Infant Death Syndrome on Facebook: Qualitative Descriptive Content Analysis to Guide Prevention Efforts

Kelly Pretorius[1], PhD, MPH, MSc, PNP, RN; Eunju Choi[1], MSc, RN; Sookja Kang[1], BSc, RN; Michael Mackert[1], PhD

The University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Kelly Pretorius, PhD, MPH, MSc, PNP, RN
The University of Texas at Austin
1710 Red River St
Austin, TX, 78712
United States
Phone: 1 5124717913
Email: kpretorius@utexas.edu

Abstract

Background: Sudden unexpected infant death (SUID), which includes the diagnosis of sudden infant death syndrome (SIDS), is a leading cause of infant mortality in the United States. Despite prevention efforts, many parents continue to create unsafe infant sleep environments and use potentially dangerous infant sleep and monitoring devices, ultimately leading to sleep-related infant deaths. Analyzing Facebook conversations regarding SIDS may offer a unique maternal perspective to guide future research and prevention efforts.

Objective: This study aims to describe and analyze conversations among mothers engaged in discussions about SIDS on a Facebook mother's group. We were interested in understanding maternal knowledge of SIDS, identifying information sources for SIDS, describing actual infant sleep practices, exploring opinions regarding infant sleep products and monitoring devices, and discovering evidence of provider communication regarding SIDS.

Methods: We extracted and analyzed 20 posts and 912 comments from 512 mothers who participated in a specific Facebook mother's group and engaged in conversations about SIDS. There were 2 reviewers who coded the data using qualitative descriptive content analysis. Themes were induced after discussion among researchers and after the study objectives were addressed.

Results: The theme of social support emerged, specifically informational and emotional support. A variety of informational sources for SIDS and safe sleep were identified, as was a continuum of infant sleep practices (ranging from unsafe to safe sleep per the American Academy of Pediatrics standards). There was widespread discussion regarding infant sleep products and monitoring devices. Embedded within conversations were (1) confusion among commonly used medical terminology, (2) the practice of unsafe infant sleep, (3) inconsistency in provider communication about SIDS, and (4) maternal anxiety regarding SIDS.

Conclusions: We uncovered new findings in this analysis, such as the commonality of infant sleep products and monitoring devices and widespread maternal anxiety regarding SIDS. Additionally, mothers who participated in the Facebook group provided and received informational and emotional support regarding SIDS via this social media format. Such results can guide future prevention efforts by informing health communication regarding SUID and safe sleep. Future provider and public health agency communication on the topic of SUID and safe sleep should be simple and clear, address infant sleep products and monitoring devices, address maternal anxiety regarding SIDS, and address the common practice of unsafe sleep.

(J Med Internet Res 2020;22(7):e18474) doi:10.2196/18474

KEYWORDS

sudden infant death; SIDS; infant mortality; safe sleep; social media; social support; health communication; maternal health; qualitative research; health care providers

XSL•FO
RenderX

Introduction

Background

Sudden unexpected infant death (SUID) and sudden infant death syndrome (SIDS) are a leading cause of infant mortality in the United States, resulting in approximately 3500 infant deaths annually [1]. SUID includes the diagnosis of SIDS and other unintentional causes of infant death: asphyxia, accidental suffocation and strangulation in bed, or ill-defined deaths. Risk factors for SUID include prone sleeping, bed sharing, soft bedding, unsafe sleep surfaces, prematurity, and smoke exposure [1]. Although the Back to Sleep campaign [2] reduced SIDS deaths by more than 50%, SIDS remains to be the leading cause of postneonatal mortality in the United States [3]. Furthermore, there has been an increase in infant deaths because of accidental suffocation [4].

Despite recommendations for SUID prevention, as of 2015, approximately 21% of mothers reported placing infants on their stomach to sleep, 61.4% reported bed sharing, and 38.5% reported using soft bedding, such as blankets or bumper pads [5]. Another study found that 54.7% of mothers reported the use of hazardous bedding for their infants [6]. Furthermore, potentially dangerous infant sleeping devices (eg, Rock ‘n Play [7], baby swings) and sleep monitoring devices (eg, Owlet, Snuza) are increasingly common, yet little is known about their use [8].

Mothers place infants to sleep in the prone position for perceived safety, infant comfort, and concern for choking [9,10]. Colson et al [11] found that mothers who were advised consistently by their doctors to practice supine sleeping were less likely to place infants prone. Mothers were also more likely to follow social norms or model behaviors that were perceived as positive [11]. Such perspectives may be influenced by cultural norms and family members [10]. Extended social networks, such as social media, are also potentially influential. Social media is available 24/7 for people to seek information or leave messages for peers [12]. Thus, mothers are increasingly using social media for parenting and health information [13,14]. As of 2015, 75% of parents used social media, often for parenting information or support [14]. In a review of parental use of social media for parenting in the United States, Facebook was the most common social media format [15]. Facebook was popular among all races or ethnicities, including African Americans [15]—those most at risk for SUID [1].

Objectives

Thus, this study aimed to (1) understand maternal knowledge of SIDS, (2) identify information sources for SIDS, (3) describe actual infant sleep practices, (4) explore opinions regarding infant sleeping devices and sleep monitoring devices as they relate to SIDS, and (5) discover evidence of provider communication regarding SIDS among mothers engaged in discussions about SIDS on a Facebook mother’s group.

Methods

This is a qualitative study of publicly available data extracted from a Facebook mother’s group in May 2019. At that time, there were 17,545 members in the group. Once on the Facebook page, *SIDS* was entered in the search toolbar. The results were filtered to include the following: *all posts*, *most recent*, posted by *anyone*, tagged location of *anywhere*, and *any date* posted. The phrase *safe sleep* was trialed but ultimately not selected as this resulted in erroneous conversations. The term *SIDS*, on the other hand, resulted in a total of 20 posts that were all relevant to SIDS or safe sleep. Each post and related conversations (912 comments) were manually copied and pasted into a spreadsheet. To protect the identity of the participants, only the initials of the participant were copied to the spreadsheet to enable the researchers to follow the conversations. After the posts and replies were copied, the spreadsheet was manually compared with the actual Facebook page to ensure accuracy. Personal and identifying information were removed to maintain confidentiality.

The posts and related conversations were then converted to 20 documents to be uploaded to Atlas.ti [16] for further analysis. The qualitative descriptive analysis process was completed as per Miles et al [17], and inductive coding was allowed for codes and themes to emerge progressively.

After analyzing 25% of the content, a preliminary codebook [18] was developed as coding became redundant at that time. To ensure trustworthy findings and increase reliability [17], 2 investigators coded all of the data. First-cycle coding [17] was completed by 2 investigators (KP and EC or SK) to assign descriptive coding of the data using qualitative analysis software (Atlas.ti, version 8.4.2). The discrepancies in coding were discussed among the members and the codebook was adjusted as needed. Second-cycle coding was then completed to identify themes [17]. This study was submitted for review by the institutional review board at the University of Texas at Austin and deemed exempt given that this research project utilized information from publicly available datasets.

Results

General Description of the Data

A total of 526 individual mothers participated in the 20 posts and 912 related comments. The number of comments on the posts ranged from 4 to 152, with an average of 45.6 comments per post. Of the 20 posts, 6 included pictures. Table 1 includes a general description of each post and the total number of comments for each post.

The post that received the most likes (n=20) was a *Seattle Times* article about a doctor who linked hearing dysfunction to SIDS; if infants failed the screening, they would undergo a more thorough exam because they were more at risk than those who passed. The post that received the second highest number of likes (n=12) was a post from a mother who wanted to know who else “went against the recommendations” and let their babies sleep prone. Multiple posts received 0 likes and included the following topics: breathable bumpers, babies rolling prone, co-sleeping, and an infant who “hated” sleeping flat. Table 2 demonstrates the Facebook emotions and the number of comments that received such emotions. Of note, people who participated in *liking* a comment did not necessarily add a

XSL•FO
RenderX

remark to the post. Therefore, the total number of participants in the analyzed conversations may have actually exceeded 526.

Table 1. General description of the Facebook posts and number of comments (in order, from most to least number of comments).

Comments, n	General content of original post (and ultimate conclusion, if known)
155	Mother attended a baby safety class and was told she should keep her baby in the same room for a year to prevent SIDS[a]. She was not planning on doing this and wanted to know what other moms' opinions were on this topic
152	Mother is terrified of SIDS and is concerned because her husband would like to co-sleep. She would also like to co-sleep but is very anxious about it. She ultimately decided to co-sleep after receiving much input from other mothers
89	Mother has newborn who did not sleep well in bassinet and slept well in the Rock 'n Play. She is not sure if the Rock 'n Play is safe because of SIDS and asked if other mothers use it for sleep. After much input from other mothers, she chose not to use the Rock 'n Play because her pediatrician advised against it
85	Mother cannot stop thinking about SIDS; she is terrified and asked other mothers how to cope with this stress. After much input from other mothers, she mentioned that she will likely proceed with purchasing a baby monitoring device
81	Mother wanted to know how many "went against the recommendations" and allowed their babies to sleep prone. Her mother and mother-in-law reassured her that prone sleeping is acceptable
42	Mother's friend lost a baby to SIDS and she asked for suggestions on how to support her. After much input from other mothers, she thanked the group and shared her appreciation for the Facebook group
41	Mother is pregnant with her third baby and is scared of SIDS. She asked about other mothers' experiences with the Owlet. After much input from other mothers, she mentioned that she will likely co-sleep but may consider placing the baby in a crib if she has reassurance from a monitor
36	Mother is terrified of SIDS and sleeps with baby in an in-bed bassinet and with a baby monitor. She asked if her feelings were normal
36	Mother shared link to an article discussing a study that found swaddling may cause SIDS. There was much input from mothers; some comments made light of the article and scrutinized scientific studies
31	Mother asked for input regarding co-sleeping. She fell asleep at the hospital with her baby and the hospital "freaked" her out because of SIDS, but she is considering co-sleeping because she breastfeeds
27	Mother asked other "tummy sleeper" mothers for reassurance. She quoted Dr Sears, who condones prone sleeping, and mentioned that other SIDS risks were low. After much input from other mothers, she mentioned that she found a swaddle on the web and that her baby was able to sleep supine
25	Mother asked for input about crib bumpers, mentioning that SIDS is no longer an issue because her baby is older than 6 months. After much input from other mothers, she mentioned that she will try bumpers and put stuffed animals in the crib corners (a suggestion from another mother)
24	Mother asked if other mothers let their babies sleep prone as her baby sleeps better this way; however, she is worried about SIDS
23	Mother asked if she can leave her baby prone once he or she rolls on his or her own; she is nervous because her baby is at the peak age for SIDS
20	Mother asked if she can leave her baby on stomach once he or she rolls on his or her own; she is following safe sleep guidelines and has a baby monitor but is still nervous about SIDS. After much input from other mothers (including recommendations for sleeping devices or monitors), she thanked the group and said that she felt better
18	Mother asked when she can stop worrying about SIDS because she would like to use crib bumpers and a blanket
11	Mother is concerned about SIDS after hearing "horror stories" and is considering the Owlet. She asked if anyone is selling one
6	Mother asked if other mothers' babies will not sleep flat. Her baby is sleeping on a blanket on top of a nursing pillow, but she is scared of SIDS and knows flat is safer
6	Mother asked about breathable crib bumpers. She knows crib bumpers are controversial because of SIDS but wants to keep baby from hitting the crib and waking up. After much input, she thanked the group and says she will wait a bit longer but wanted to get input beforehand
4	Mother shared link to an article about a researcher who linked SIDS to children with audiologic problems

[a]SIDS: sudden infant death syndrome.

XSL•FO
RenderX

Table 2. Frequency of Facebook emotions per comments.

Facebook emotion	Comments receiving the emotion, n
1 like	241
2 likes	72
3 likes	36
4 likes	15
5 likes	12
1 love	45
2 loves	10
1 sad	10
2 sad	3

Themes

In this analysis, the theme of social support emerged. Two types of social support were evident: informational and emotional support. The types of social support have been defined and modified by Cutrona and Suhr [19]. Informational support involves advice or suggestions, and emotional support involves sharing concern. Such categories of social support are applicable to web-based environments [20,21].

Informational Support

Informational support was evident in the following discussions among mothers: (1) asking questions about SIDS, infant sleep, or baby products; (2) sharing personal experiences of provider communication regarding SIDS or safe sleep; (3) sharing personal definitions or beliefs regarding SIDS or safe sleep (including discussions on vaccines and SIDS); and (4) sharing informational sources for SIDS or safe sleep. Overwhelmingly, the information shared demonstrated misinformation and inaccurate use of terminology.

Asking for Recommendations

Mothers openly asked for recommendations or feedback regarding infant sleep practices, use of baby products or monitoring devices, or other topics related to SIDS or safe sleep. Mothers asked for opinions about "what other moms' opinions [were] on the subject" of prone sleeping, the use of the Rock 'n Play, when to transition to a crib, crib bumpers, whether they should seek help for their anxiety about SIDS, or what monitor system to purchase (Owlet vs Snuza). Many of the responses involved information sharing of provider communication, personal definitions or beliefs, and information sources.

Provider Communication

Mothers shared information provided by their health care provider regarding SIDS, safe sleep, and the use of baby products. Some mothers shared their doctor's knowledge of SIDS, such as the risk of SIDS being highest before 4 months of age or that the risk of SIDS from prone sleeping was "incredibly slim." Another mother tried to distinguish co-sleeping from bed sharing; she mentioned that her doctor helped her understand that they were different. Some communication with providers endorsed safe sleep, whereas others condoned unsafe sleep. For instance, mothers mentioned that their pediatricians recommended supine sleeping and pacifier use and encouraged mothers not to use the Rock 'n Play; however, other mothers mentioned that their pediatricians were "ok" with the Rock 'n Play. One mother said that her pediatrician was "ok" with her use of a breathable crib bumper, and another mother said that her pediatrician endorsed prone sleeping for naptime, just not bedtime.

Personal Definitions or Beliefs

Mothers also shared information by providing their own personal definitions or beliefs regarding SIDS or safe sleep. There were approximately 80 personal definitions or beliefs shared, some of which were accurate and some were not. Some mothers felt the risk of SIDS was exaggerated; one mother did her "own research" and said that SIDS is actually not common as only "2000" infants die yearly. Another mother mentioned that "they are overzealous about the SIDS thing," whereas another mother agreed that it is very rare. Similarly, another mother, who self-identified as a pediatric provider, said that most babies were not at risk for SIDS if they did not have other risk factors such as smoking, neurological issues, or vaccines. One mother even expressed anger with the Back to Sleep campaign—she felt the campaign caused flat heads and developmental delays. There was also confusion about SIDS versus suffocation death and co-sleeping versus bed sharing. One mother said that co-sleeping was not a risk for suffocation, especially if one was following safe bed-sharing guidelines. Many mothers felt that that co-sleeping was "ok" if done safely and that SIDS had "nothing to do with bed sharing." One mother mentioned that "statistically it is safe for breastfed babies to co-sleep." Although many of the shared definitions and beliefs were inaccurate, including a long discussion of how vaccines cause SIDS, some mothers shared accurate information. For instance, some mothers commented that SIDS can occur anywhere and that co-sleeping is a risk for suffocation. Other mothers explained ways to prevent SIDS: keeping the house cool, use of a pacifier, no blankets, not smoking, and placing the baby supine. Finally, some mothers disagreed that vaccines cause SIDS.

Information Sources

Mothers shared information sources with other mothers, as it related to SIDS or safe sleep. The list of information sources is shown in Table 3. Mothers also shared links to baby products or monitoring products to facilitate purchases.

XSL•FO
RenderX

Table 3. List of information sources shared or discussed in the Facebook group (in alphabetical order).

Information sources	Direct link or sources	Descriptions
AAP[a]	Not provided	Not provided
AAP press release	Link: https://www.aap.org/en-us/about-the-aap/aap-press-room/Pages/Bed-Sharing-Remains-Greatest-Risk-Factor-for-Sleep-Related-Infant-Deaths.aspx	The press release was about a study published in *Pediatrics* titled *Bed-sharing remains greatest risk factor for sleep-related infant deaths*
Babybargains.com	Link: www.Babybargains.com	From the website: "We obsess over baby gear . . . so you don't have to. Baby Bargains has one mission: help you find the best gear for your baby with unbiased reviews by experts with 20 years of experience"
Baby Safety Academy	Not provided	Mom shared information from attending this class: "Infants should sleep in the same room as the parents for the first year of life"
Book: *Babywise*	Not provided	Not provided
Book: *Bearing the Unbearable*	Not provided	Not provided
Biologically Normal Infant Sleep	Facebook group	Per the mother who shared this source, this group had information on "safe co-sleeping" based on Dr James McKenna
Compassionate Friends	Online support group for parents who have lost a child	Not provided
Dr Sears	Not provided	Per the mother who shared this source, Dr Sears says the following: "The front-sleeping risk factor for SIDS[b] doesn't mean that you should worry every time you place your baby down to sleep. Just be sure to place your baby to sleep on a safe bedding surface. After all, over 99.9 percent of tummy-sleeping infants wake up every morning"
Ezinearticles.com	Link: https://ezinearticles.com/?Preventing-Tragedies---Moving-Air-From-a-Fan-is-All-You-Need-to-Prevent-SIDS%3F&id=1650718	The article was titled *Preventing tragedies - Moving air from a fan is all you need to prevent SIDS*. The article cited a study by Kaiser Permanente, published in *Archives of Pediatrics and Adolescent Medicine*, finding that fan use can prevent SIDS
Family Sleep Institute	Link: https://familysleepinstitute.com/position-statements-2/	The position statement discussed the update in safe sleep guidelines and provided a link to Charlie's Kids (www.charlieskids.org) for further guidance on safe sleep
Gavin's Gift of Grace	Facebook group	Not provided
Google	Not provided	Not provided
Gracie Faith SIDS Prevention Program Inc	Not provided	Not provided
Hand to Hold	Support group for parents who have lost a child	Not provided
New York Times	Link: https://well.blogs.nytimes.com/2016/05/09/swaddling-may-increase-the-risk-of-sids/	The article was about a study published in *Pediatrics*, finding that swaddling may contribute to SIDS
New Zealand study	Not provided	Per the mother who shared this source, a study in New Zealand found that toxins in mattresses contribute to SIDS
NPR[c]	Link: https://www.npr.org/sections/health-shots/2017/06/05/531582634/babies-sleep-better-in-their-own-rooms-after-4-months-study-finds	The article was about a study published in *Pediatrics*, finding that babies sleep better in their own rooms after 4 months
Rhett Sullivan Foundation	Link: https://rhettsullivan.org	From the website: the mission is to help "families who experience unexpected early child loss"

XSL•FO
RenderX

Information sources	Direct link or sources	Descriptions
Safe Infant Sleep - Evidence-Based Support Group	Facebook group	Not provided
Safe Sleep and Baby Care - Evidence Based Support	Facebook group	Not provided
Seattle Times	Link: https://www.seattletimes.com/seattle-news/health/one-seattle-childrens-doctor-thinks-he-close-to-stopping-sids/	The article is about a doctor linking audiology problems to SIDS
Book: *Sweet Sleep: Nighttime and Naptime Strategies for the Breastfeeding Family*, written by La Leche League International and others	Link: https://books.apple.com/us/book/sweet-sleep/id813166216	From the website: "Research finds that most breastfeeding mothers do sleep with their babies at some point and preparing for bedsharing is safer than accidentally falling asleep together"
The Pediatric Insider	Link: https://pediatricinsider.wordpress.com/2015/05/28/swings-slings-and-car-seats-are-not-for-sleeping/	The article, written by a pediatrician, advises against using swings, slings, and car seats for sleeping. The blog referenced an article published in *Pediatrics* titled *Hazards associated with sitting and carrying devices for children two years and younger*

[a]AAP: American Academy of Pediatrics.

[b]SIDS: sudden infant death syndrome.

[c]NPR: National Public Radio.

Emotional Support

Emotional support was evident in the following discussions among mothers: (1) encouraging each other to "do what's best for you or your family," (2) telling each other that "it will get better with time," (3) stories of infant or child death, (4) expressions of maternal anxiety about SIDS, and (5) general comments that relayed support.

Do What's Best

Mothers offered emotional support by encouraging each other to do what was best for their family when talking about SIDS or safe sleep, regardless of the infant sleep environment or use of baby products. For example, it was common for a mother to comment "I think you need to do whatever you are comfortable with" or "you have to do whatever you feel comfortable with because if not, you'll never sleep" regarding infant sleep environments. Many mothers supported following a "mother's instinct" or "mother's gut." For example, one mother said "research, educate yourself, and trust your mama gut" when discussing how to help a new mother ease her worries about SIDS. Another mother commented: "whatever you choose to do will be right, because you are the mom. Your instincts will most always be correct." There was also a focus on doing what was best for each family, despite what the safe sleep recommendations might be. For instance, it was common to see mothers comment "I think every family is different and not everything works for everyone" or "I think whatever you feel is right for you and your baby, just because it's recommended doesn't mean you have to follow." One mother addressed the changing infant sleep recommendations, saying "…info changes all the time and we should all do what works for us and our baby." However, some mothers did caution about the support and advice being given. For example, one mother said, "do what works best for you but keep in mind you're asking for opinions and opinions do not change the research or statistics."

It Gets Better With Time

Similar to the emotional support for mothers to do what was best for their family, emotional support was also prevalent in discussions of how worrying about SIDS gets better with time. Mothers reassured each other that the "worry of SIDS" was normal and that it would get better over time. One mother celebrated when her infant turned 4 months old and encouraged another mother that with time, she will realize "wow. I haven't thought about SIDS in weeks." Another mother said that her concerns subsided eventually, even though she still found herself standing over the crib at times.

Stories of Death

Emotional support was also prevalent in conversations surrounding stories of infant or child death. These conversations included mothers asking how to support mothers who lost an infant or child. For example, when discussing how to support a mother who had recently lost an infant to SIDS, one mother said, "my friend lost her daughter to SUIDS and talking about her daughter makes her happy." Other mothers shared stories of infant or child death, including infants who died from co-sleeping or from "sleeping in rock and plays."

Maternal Anxiety

Emotional support was evident in conversations surrounding the concept of maternal anxiety related to SIDS. Mothers openly asked if their anxiety about SIDS was normal and how to cope with such feelings. One mother said that postpartum stress is "absolutely awful;" she explained that she cried daily for weeks and felt that this was normal. Many mothers shared stories of watching their baby sleep throughout the night, mentioning they were "terrified of SIDS". Mothers replied by offering support, by encouraging the use of baby products (Owlet or Snuza), recommending that the mother speak with a health care provider, or normalizing such feelings. For instance, it was common for a mother to say the Owlet was the "only way I [could] sleep…and helped my anxiety."

XSL•FO
RenderX

Supportive Comments

Finally, emotional support was embedded within conversations around SIDS or safe sleep when mothers frequently shared supportive or encouraging comments. There were approximately 100 comments demonstrating support, evident in conversations of various topics. It was common to read "Good luck, mama," "I was right there with you," or "Hang in there." Mothers demonstrated unity in feeling anxiety about SIDS, encouraged each other to not worry so much, and to get some sleep. Often, these words of encouragement were accompanied by emojis, such as hearts or kisses.

Information Sources

Of the 20 main posts, 4 included references to an information source. Of those, the following were mentioned: *Seattle Times*, *Time Magazine*, Dr Sears, and Baby Safety Academy. It is worth noting that the American Academy of Pediatrics (AAP) was mentioned at least nine times throughout the posts and related comments. Table 3 lists the information sources that were shared in the posts and related conversations, with a brief description or direct link, if provided by the participant.

Parental Practices of Infant Sleep

Of the 20 main posts, 12 involved infant sleeping practices. Of these, 2 were considered safe sleep environments; both of these posts were from mothers inquiring about what to do once their infant began rolling onto their stomach. Of the 12 posts that involved infant sleeping practices, 10 were considered unsafe practices. These included discussions about the following infant sleep practices: co-sleeping (2), sleeping on a nursing pillow (1), prone sleeping (3), crib bumpers (3), and the Rock 'n Play (1). In analyzing the data, it became apparent that parental practices of infant sleep are best understood on a continuum rather than as a black and white matter. Infant sleep practices are surprisingly complex, and mothers adjust to the baby's needs as well as their own family's needs.

Unsafe Sleep

Many mothers changed sleep practices based on the infant and the needs of the family. For instance, many mothers did not bed-share with their first child, but chose to bed-share with their second. One mother encouraged other mothers to "listen to [their] instincts." Another mother said that it just depends on the "family dynamic."

Many mothers were aware of safe sleep recommendations but chose to practice unsafe sleep. For example, one mother encouraged others to join an evidenced-based group on safe sleep but mentioned that she bed shared with her first child and might with her second. There was also confusion around safe sleep recommendations and infant sleep practices. For example, one mother bed shared with her children, but only for the first few months before transitioning them to a crib, because "SIDS is less likely if they're in bed." Many mothers also allowed infants to sleep prone for nap time, but not at nighttime. One mother had 2 of her children sleep in the Rock 'n Play and her third in a DockATot; she wanted them next to her, but in their own safe space, because she was "SIDS traumatized."

Within conversations surrounding such practices, there was also evidence of motivation for unsafe sleep practices. One mother chose to bed-share because she had a cesarean section and it was a "chore" to walk to a crib. Many mothers chose to bed-share because their babies had reflux and they wanted to "be right next to [the baby] if anything happened." Some mothers simply believed that their baby was safer next to them in bed. Many mothers "didn't plan to co-sleep" but had a baby that would only sleep in bed with the mother or had choking because of reflux. Another mother allowed her infant to sleep in bed on a pillow because the mother could not rest "without [the baby] being near," despite knowledge that such practices were not recommended. Mothers also explained that it was not feasible for infants to sleep in their bedrooms for the recommended 6 to 12 months because of personal difficulty sleeping and returning to work.

Among discussions of unsafe sleep practices were elements of confidence in such practices. Mothers felt confident they would "wake up the moment anything [happened]" or mentioned that they would never roll onto their infant when bed sharing. Another mother who was bed sharing explained that she was not worried about suffocation as the blanket did not reach the infant's head and that she and her husband do not move throughout the night.

Safe Sleep

Although many comments described unsafe sleep environments, some mothers encouraged safe infant sleep environments. For instance, in responding to a mother who asked for advice about her infant who preferred to sleep prone, whereas many mothers endorsed this practice and recommended the Rock 'n Play, one mother commented "SIDS is most common at 3 months old. Back is best." Other mothers commented that the mother was taking a significant risk, and some explained why the mother should follow the "AAP safe sleep guidelines."

Sleeping Devices or Sleep Monitoring Devices

Discussions surrounding infant sleeping devices or monitoring devices permeated many conversations regarding SIDS or safe sleep. Mothers asked specific questions about the use of infant monitoring devices or sleep devices and encouraged other mothers to use devices to cope with their fear of SIDS. Of the 20 main posts, 10 involved the topic of sleeping devices or monitors. The breakdown of the sleeping device or sleep monitoring devices discussed in the original 20 posts are displayed in Table 4. Throughout all of the data analyzed, the device most commonly mentioned was the Owlet monitor, which was mentioned 112 times. The Snuza monitor was mentioned 53 times and the Angelcare monitor, 5 times. The Rock 'n Play was mentioned 26 times and the DockATot, 4 times.

XSL•FO
RenderX

Table 4. Sleeping device or sleep monitoring device mentioned in main posts.

Sleeping devices or sleep monitoring devices	Posts, n
Crib bumpers	4
Owlet monitor	2
In-bed bassinet	1
Nursing pillow	1
Rock ‘n Play	1
Mimo Monitors	1

Mothers frequently mentioned that sleep monitoring devices provided “peace of mind” and used Facebook as a marketplace to buy and sell such devices. It was common for mothers to ask other mothers about their experience with the Owlet; one mother was considering purchasing one to “ease my mind… [and] help me to sleep better at night instead of always getting up to check that they’re still breathing.” Many responses encouraged the use of sleep monitoring devices. For example, many mothers said, “Owlet is my life” or “Owlet helped my SIDS anxiety” and expressed that the Owlet was one of the necessities of being a mom. For instance, one mother said that she could not live without the Owlet, diapers, and clothing. In comparison to many mothers recommending sleep monitoring devices, there were few mothers who advised against the use of sleep monitoring devices. One mother explained that the false alarms caused her more distress and advised against using the Owlet. Another mother said that it would be “one more thing to obsess over.”

There were also frequent conversations about other baby products. The Rock ‘n Play was by far the most frequently discussed device; however, mothers also discussed the HALO SleepSack, Merlin’s Magic Sleepsuit, DockATot, Love To Dream SWADDLE UP, Woombie, Snuggle Nest, etc. These discussions suggested strategies for improving infant sleep and either encouragement for use for a safer option (vs the Rock ‘n Play) or sometimes encouragement of use despite the known risk of SIDS.

Provider Communication

In the dataset, there were approximately 30 mentions of health care provider communication regarding SIDS. Of the 20 original posts, 2 included health care providers and involved provider communication. These 2 examples include a mother who described how the hospital “freaked” her out when she fell asleep with the baby in the hospital bed and another mother who asked about prone sleeping—as doctors used to recommend this practice. Other conversations included pediatricians recommending safe sleep practices and pediatricians condoning unsafe sleep practices. For example, one pediatrician “had no problem” with a baby sleeping in the Rock ‘n Play, and a self-identified pediatric nurse claimed that she had never heard of some of the safe sleep recommendations. However, when one mother told her pediatrician about her use of the Rock ‘n Play, she was informed of the risks of suffocation.

Discussion

Principal Findings

As suggested by Huo and Turner [12], studies on social media user perspectives are needed to guide the development of social media interventions. However, we believe this analysis guides future SUID prevention interventions beyond social media, providing guidance for health care providers, public health agencies, and health campaigns. Similar to prior social media analyses [22,23], this analysis has revealed new findings regarding maternal perspectives of SIDS and safe sleep that are otherwise not discussed in the literature. This analysis also supports prior findings regarding parental practices of infant sleep and provider communication of SUID prevention.

Themes

Social media has been identified as a supportive environment in prior analyses and studies [15,24-26], which was consistent with the identified theme of support. Understanding the types of support, informational and emotional, provided among mothers and in this format is a new finding. This Facebook mother’s group promoted participant engagement and honesty about very personal health-related concerns, resulting in discussions about maternal anxiety, personal practices of infant sleep, and discussions about infant sleeping devices and sleep monitoring devices. This frankness among social media users has previously been discovered in discussions about health-related topics [23]. One implication of this finding is the potential of emulating this type of supportive environment in other settings. For instance, health care organizations could consider creating more supportive environments in our current health care system to promote open discussions with parents. Additionally, many of the comments and conversations in this analysis provided opportunities for correction of misinformation or discussions regarding options for safer infant sleep. Although mothers receive information via many different routes [13], social media remains incredibly popular [27] and is a platform where patients can receive and communicate health information. Since Facebook mother groups are available 24/7, and most health care providers are not, health care organizations should also rethink how services are structured. For instance, if a nurse or provider were truly accessible 24/7 for advice via a social media format, this could potentially influence and impact parental decisions and decrease the spread of misinformation that is otherwise widespread among social media sites [28].

XSL•FO
RenderX

Information Sources

The information sources shared among the mothers regarding SIDS or safe sleep demonstrated a variety of sources. We know that mothers obtain information from many different sources [13], and this study supports this. However, this study identified sources specific to SIDS and safe sleep. Consistent again with a Twitter analysis on SIDS and safe sleep [24], there was evidence of news media organizations, such as *National Public Radio*, being shared. However, a concerning finding was the sharing of informal and potentially inaccurate information sources, such as Dr Sears, other Facebook groups, and controversial books such as *Babywise*. Thus, shared sources were often inconsistent with the AAP recommendations. This is similar to the findings from a prior study of Google searches that identified information largely contradicting AAP recommendations of infant safe sleep [29]. The commonality of shared links for purchase of baby products was also worrisome given that many of the shared products are not considered safe for infant sleep nor recommended by the AAP [3]. Thus, health care providers may consider asking parents where they obtain health information to openly discuss and potentially correct any misinformation that is shared in that format.

Parental Practices of Infant Sleep

Another finding is that infant sleep practices are not straightforward; safe sleep and unsafe sleep are best described on a continuum. Mothers alter infant sleep practices based on the infant and family's needs. Many mothers in the group were aware of safe sleep recommendations but chose not to follow them because they were not feasible. This finding supports a prior study's conclusion that parental motivation to bed-share trumped known risks of unsafe sleep [30]. The discovery of mothers who are practicing unsafe and ever-changing infant sleep practices is also consistent with studies demonstrating parental practice of unsafe sleep [5,6,31,32] and the changing of infant sleep environments throughout the night [33]. In this Facebook group, some mothers endorsed safe sleep and recommended following the AAP guidelines; however, other mothers believed they were following the recommendations and were actually creating dangerous infant sleep environments. Safe sleep is seemingly complex and further complicated by the confusion surrounding definitions of SIDS and safe sleep. For instance, this analysis demonstrated wide confusion and inaccurate use of the terms *co-sleeping* and *bed sharing*. These findings imply that simpler terms should be used in public health campaigns and when educating families about SUID prevention. This finding also informs health communication; health communication about safe sleep and SUID prevention should use simple language, be direct, and be clear.

Sleeping Devices or Sleep Monitoring Devices

This study demonstrates the popularity and commonality of infant sleep monitoring devices and baby products; thus, this topic can no longer be ignored. This is similar to the findings from a Twitter analysis on SIDS and safe sleep, where conversations and advertising about such products were widespread [24]. It is worth noting that the literature is sparse regarding infant sleep monitoring products and baby products. Health care providers need to educate themselves on popular products, so they can effectively discuss this topic with parents. For example, it may be helpful for health care providers to directly ask parents "What devices are you using to help your baby sleep?" Public health agencies should also consider addressing such devices in campaigns and in health messaging.

Provider Communication

Furthermore, when discussing the topic of SUID and safe sleep, health care providers and public health agencies need to continue to provide accurate information on a consistent basis. Health care provider advice on safe sleep impacts parental decisions [11], and inconsistency has been demonstrated among safe sleep messaging provided by health care professionals [34,35]. As this study also supports inconsistency in health care provider communication about safe sleep, there is room for improvement. Instead of assuming parental knowledge of safe sleep guidelines and that families are following such guidelines, health care providers should encourage open and honest conversations about infant sleep practices that occur throughout the night and at naptime. Health care providers should also recognize the need to consistently and correctly share safe sleep recommendations with families and caregivers.

Maternal Anxiety

Finally, the prevalence of discussions surrounding the topic of maternal anxiety related to SIDS and how mothers reassured each other that these feelings were normal is concerning. Many mothers coped with these feelings by using baby monitoring devices or baby sleeping devices rather than taking precautionary measures to prevent SUID and sleep-related infant death. Persons with mental illness who openly share their stories and feelings offer insight into various illnesses and symptomology [36]; thus, this analysis may have actually identified mothers who have postpartum disorders. According to O'Hara and Wisner [37], the prevalence of major and minor depression in pregnancy and the postpartum period is 20%; early symptoms can be detected through screening, and early treatment is essential for the well-being of the mothers and children. Although some mothers shared that they had sought medical treatment for postpartum disorders, many did not. This topic raises the question of whether our health care system is successfully screening and identifying mothers who may need additional support and treatment for postpartum anxiety and depression.

Limitations

Although this study provides new insights into maternal perspectives regarding SIDS and safe sleep, it is not without limitations. This particular Facebook mother's group was selected because of the large number of participants; however, the views expressed may not represent potential opinions, concerns, and views of other Facebook mother group participants (or other mothers for that matter). Furthermore, demographic information of the participants was not obtained, limiting the generalizability of the findings. This analysis only included mothers; therefore, other perspectives (such as that of the father or other caregivers) were missed. Although many of the findings were consistent with a different social media

XSL•FO
RenderX

analysis on SIDS and safe sleep [24], an analysis of another social media format may have resulted in contradictory findings. Additionally, social media language can be difficult to analyze; there are often typos, informal writing, or abbreviations that can make the analysis challenging [38]. Although measures were taken to prevent misinterpretation and the research team refrained from interpretation as much as possible, it is possible that some of the original content may have been misunderstood or misinterpreted. Finally, the participants often used safe sleep definitions inaccurately, potentially impacting the analysis. For instance, *co-sleeping*, *bed sharing*, and *room sharing* were often used interchangeably. The research team interpreted the meaning of the terms in the context used; however, it is possible that this was interpreted inaccurately.

Conclusions

Despite such limitations, this analysis provides new information regarding maternal perspectives on SIDS and safe sleep: (1) a wide variety of information sources, (2) widespread utilization of infant sleep products and monitoring devices, and (3) maternal anxiety regarding SIDS. This study demonstrated confusion among the terminology commonly used in the medical community when speaking of SUID and safe sleep, implying that future communication should aim for simpler and clearer terms. Widespread practices of unsafe sleep and inconsistency in provider communication regarding SUID prevention are not new findings but emphasize the need for continued efforts in SUID education and prevention. Health care providers, health care organizations, and public health agencies should incorporate these findings in future research and health campaigns and when directly communicating with families about SIDS prevention and safe sleep. Furthermore, such organizations should consider using social media in their marketing efforts and actively engage in such formats to correct misinformation.

Acknowledgments

The authors would like to thank Dr Lynn Rew for her invaluable input throughout the study. The authors also acknowledge financial support from Sigma Theta Tau and the National Association of Pediatric Nurse Practitioners.

Authors' Contributions

KP conceived the idea for the study and extracted the data. KP, EC, and SK analyzed the data. MM provided feedback during the design and analysis phases. KP wrote the manuscript, and all authors discussed the results and reviewed the manuscript before submission.

Conflicts of Interest

None declared.

References

1. Carlin RF, Moon RY. Risk factors, protective factors, and current recommendations to reduce sudden infant death syndrome: a review. JAMA Pediatr 2017 Feb 1;171(2):175-180. [doi: 10.1001/jamapediatrics.2016.3345] [Medline: 27918760]
2. Havens DH, Zink R. The 'back to sleep' campaign. J Pediatr Health Care 1994 Sep;8(5):240-242 [FREE Full text] [doi: 10.1016/0891-5245(94)90070-1]
3. Moon RY, Task Force on Sudden Infant Death Syndrome. SIDS and other sleep-related infant deaths: evidence base for 2016 updated recommendations for a safe infant sleeping environment. Pediatrics 2016 Nov;138(5):- [FREE Full text] [doi: 10.1542/peds.2016-2940] [Medline: 27940805]
4. Gao Y, Schwebel DC, Hu G. Infant mortality due to unintentional suffocation among infants younger than 1 year in the United States, 1999-2015. JAMA Pediatr 2018 Apr 1;172(4):388-390 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.4887] [Medline: 29459979]
5. Bombard JM, Kortsmit K, Warner L, Shapiro-Mendoza CK, Cox S, Kroelinger CD, et al. Vital signs: trends and disparities in infant safe sleep practices - United States, 2009-2015. MMWR Morb Mortal Wkly Rep 2018 Jan 12;67(1):39-46 [FREE Full text] [doi: 10.15585/mmwr.mm6701e1] [Medline: 29324729]
6. Shapiro-Mendoza CK, Colson ER, Willinger M, Rybin DV, Camperlengo L, Corwin MJ. Trends in infant bedding use: national infant sleep position study, 1993-2010. Pediatrics 2015 Jan;135(1):10-17 [FREE Full text] [doi: 10.1542/peds.2014-1793] [Medline: 25452654]
7. Fisher-Price Recalls Rock ‘n Play Sleepers Due to Reports of Deaths. United States Comsumer Product Safety Commission. 2019. URL: https://www.cpsc.gov/Recalls/2019/fisher-price-recalls-rock-n-play-sleepers-due-to-reports-of-deaths [accessed 2019-09-18]
8. King D. Marketing wearable home baby monitors: real peace of mind? Br Med J 2014 Nov 18;349:g6639. [doi: 10.1136/bmj.g6639] [Medline: 25406191]
9. Colson ER, McCabe LK, Fox K, Levenson S, Colton T, Lister G, et al. Barriers to following the back-to-sleep recommendations: insights from focus groups with inner-city caregivers. Ambul Pediatr 2005;5(6):349-354. [doi: 10.1367/A04-220R1.1] [Medline: 16302836]

XSL•FO
RenderX

10. Gaydos LM, Blake SC, Gazmararian JA, Woodruff W, Thompson WW, Dalmida SG. Revisiting safe sleep recommendations for African-American infants: why current counseling is insufficient. Matern Child Health J 2015 Mar;19(3):496-503. [doi: 10.1007/s10995-014-1530-z] [Medline: 24889117]
11. Colson ER, Geller NL, Heeren T, Corwin MJ. Factors associated with choice of infant sleep position. Pediatrics 2017 Sep;140(3):-. [doi: 10.1542/peds.2017-0596] [Medline: 28827382]
12. Huo J. Social media in health communication. In: Bian J, Guo Y, He Z, Hu X, editors. Social Web and Health Research: Benefits, Limitations, and Best Practices. New York, USA: Springer; 2019:53-82.
13. Moon RY, Mathews A, Oden R, Carlin R. Mothers' perceptions of the internet and social media as sources of parenting and health information: qualitative study. J Med Internet Res 2019 Jul 9;21(7):e14289 [FREE Full text] [doi: 10.2196/14289] [Medline: 31290403]
14. Duggan MA, Lampe C, Ellison N. Parents and Social Media: Mothers are Especially Likely to Give and Receive Support on Social Media. Pew Research Center. 2015. URL: http://www.pewresearch.org/wp-content/uploads/sites/9/2015/07/Parents-and-Social-Media-FIN-DRAFT-071515.pdf [accessed 2018-08-18]
15. Pretorius K, Johnson KE, Rew L. An integrative review: understanding parental use of social media to influence infant and child health. Matern Child Health J 2019 Oct;23(10):1360-1370. [doi: 10.1007/s10995-019-02781-w] [Medline: 31222601]
16. Update History: Version 8.4.2. ATLAS.ti: The Qualitative Data Analysis & Research Software. 2018. URL: https://atlasti.com/update-history/ [accessed 2020-07-02]
17. Miles MA, Michael SJ. Qualitative Data Analysis: A Methods Sourcebook. Third Edition. Thousand Oaks, CA: Sage Publications; 2014.
18. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and using a codebook for the analysis of interview data: an example from a professional development research project. Field Methods 2010 Dec 27;23(2):136-155. [doi: 10.1177/1525822X10388468]
19. Cutrone CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 2016 Jun 30;19(2):154-174. [doi: 10.1177/009365092019002002]
20. Braithwaite DO, Waldron VR, Finn J. Communication of social support in computer-mediated groups for people with disabilities. Health Commun 1999;11(2):123-151. [doi: 10.1207/s15327027hc1102_2] [Medline: 16370973]
21. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
22. Miller WR, Gesselman AN, Garcia JR, Groves D, Buelow JM. Epilepsy-related romantic and sexual relationship problems and concerns: indications from internet message boards. Epilepsy Behav 2017 Sep;74:149-153. [doi: 10.1016/j.yebeh.2017.06.023] [Medline: 28756337]
23. Priest C, Knopf A, Groves D, Carpenter JS, Furrey C, Krishnan A, et al. Finding the patient's voice using big data: analysis of users' health-related concerns in the chacha question-and-answer service (2009-2012). J Med Internet Res 2016 Mar 9;18(3):e44 [FREE Full text] [doi: 10.2196/jmir.5033] [Medline: 26960745]
24. Pretorius KA, Mackert M, Wilcox GB. Sudden infant death syndrome and safe sleep on Twitter: analysis of influences and themes to guide health promotion efforts. JMIR Pediatr Parent 2018 Sep 7;1(2):e10435 [FREE Full text] [doi: 10.2196/10435] [Medline: 31518314]
25. Fiks AG, Gruver RS, Bishop-Gilyard CT, Shults J, Virudachalam S, Suh AW, et al. A social media peer group for mothers to prevent obesity from infancy: the Grow2Gether randomized trial. Child Obes 2017 Oct;13(5):356-368 [FREE Full text] [doi: 10.1089/chi.2017.0042] [Medline: 28557558]
26. Asiodu IV, Waters CM, Dailey DE, Lyndon A. Infant feeding decision-making and the influences of social support persons among first-time African American mothers. Matern Child Health J 2017 Apr;21(4):863-872 [FREE Full text] [doi: 10.1007/s10995-016-2167-x] [Medline: 27565664]
27. Perrin AM. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-10-03]
28. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]
29. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the internet: let's Google it. J Pediatr 2012 Dec;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]
30. Chianese J, Ploof D, Trovato C, Chang JC. Inner-city caregivers' perspectives on bed sharing with their infants. Acad Pediatr 2009;9(1):26-32. [doi: 10.1016/j.acap.2008.11.005] [Medline: 19329088]
31. Heere M, Moughan B, Alfonsi J, Rodriguez J, Aronoff S. Factors associated with infant bed-sharing. Glob Pediatr Health 2017;4:2333794X17690313 [FREE Full text] [doi: 10.1177/2333794X17690313] [Medline: 28229101]
32. Hwang SS, Smith RA, Barfield WD, Smith VC, McCormick MC, Williams MA. Supine sleep positioning in preterm and term infants after hospital discharge from 2000 to 2011. J Perinatol 2016 Sep;36(9):787-793. [doi: 10.1038/jp.2016.80] [Medline: 27171759]

XSL•FO
RenderX

33. Batra EK, Teti DM, Schaefer EW, Neumann BA, Meek EA, Paul IM. Nocturnal video assessment of infant sleep environments. Pediatrics 2016 Sep;138(3):- [FREE Full text] [doi: 10.1542/peds.2016-1533] [Medline: 27527797]
34. Burrell TD, McDonald EM, Mahoney P, Musci RJ, Shields W, Gielen A, et al. Content of infant safe sleep counseling and maternal reported practices in an urban clinic. Acad Pediatr 2019;19(7):801-807. [doi: 10.1016/j.acap.2019.06.014] [Medline: 31254631]
35. Schaeffer P, Asnes AG. What do pediatricians tell parents about bed-sharing? Matern Child Health J 2018 Jan;22(1):51-58. [doi: 10.1007/s10995-017-2353-5] [Medline: 28766090]
36. Ismail N. Social mediapsychological disorder. In: Bian J, Guo Y, He Z, Hu X, editors. Social Web and Health Research: Benefits, Limitations, and Best Practices. New York, USA: Springer; 2019:171-192.
37. O'Hara MW, Wisner KL. Perinatal mental illness: definition, description and aetiology. Best Pract Res Clin Obstet Gynaecol 2014 Jan;28(1):3-12 [FREE Full text] [doi: 10.1016/j.bpobgyn.2013.09.002] [Medline: 24140480]
38. Ru B. A literature review of social media-based data mining for health outcomes research. In: Bian J, Guo Y, He Z, Hu X, editors. Social Web and Health Research: Benefits, Limitations, and Best Practices. New York, USA: Springer; 2019.

Abbreviations

AAP: American Academy of Pediatrics
SIDS: sudden infant death syndrome
SUID: sudden unexpected infant death

Edited by G Eysenbach; submitted 27.02.20; peer-reviewed by D Carvalho, E Da Silva; comments to author 20.03.20; revised version received 19.05.20; accepted 04.06.20; published 30.07.20.

Please cite as:
Pretorius K, Choi E, Kang S, Mackert M
Sudden Infant Death Syndrome on Facebook: Qualitative Descriptive Content Analysis to Guide Prevention Efforts
J Med Internet Res 2020;22(7):e18474
URL: http://www.jmir.org/2020/7/e18474/
doi:10.2196/18474
PMID:32729842

©Kelly Pretorius, Eunju Choi, Sookja Kang, Michael Mackert. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Assessing the Credibility and Authenticity of Social Media Content for Applications in Health Communication: Scoping Review

Eva L Jenkins[1], BNutSci (Hons); Jasmina Ilicic[2], PhD; Amy M Barklamb[1], BNutSci; Tracy A McCaffrey[1], PhD

[1]Department of Nutrition, Dietetics and Food, Monash University, Notting Hill, Australia

[2]Monash Business School, Monash University, Caulfield East, Australia

Corresponding Author:
Tracy A McCaffrey, PhD
Department of Nutrition, Dietetics and Food
Monash University
Level 1
264 Ferntree Gully Road
Notting Hill, 3168
Australia
Phone: 61 400421391
Email: tracy.mccaffrey@monash.edu

Abstract

Background: Nutrition science is currently facing issues regarding the public's perception of its credibility, with social media (SM) influencers increasingly becoming a key source for nutrition-related information with high engagement rates. Source credibility and, to an extent, authenticity have been widely studied in marketing and communications but have not yet been considered in the context of nutrition or health communication. Thus, an investigation into the factors that impact perceived source and message credibility and authenticity is of interest to inform health communication on SM.

Objective: This study aims to explore the factors that impact message and source credibility (which includes trustworthiness and expertise) or authenticity judgments on SM platforms to better inform nutrition science SM communication best practices.

Methods: A total of 6 databases across a variety of disciplines were searched in March 2019. The inclusion criteria were experimental studies, studies focusing on microblogs, studies focusing on healthy adult populations, and studies focusing on either source credibility or authenticity. Exclusion criteria were studies involving participants aged under 18 years and clinical populations, gray literature, blogs, WeChat conversations, web-based reviews, non-English papers, and studies not involving participants' perceptions.

Results: Overall, 22 eligible papers were included, giving a total of 25 research studies. Among these studies, Facebook and Twitter were the most common SM platforms investigated. The most effective communication style differed depending on the SM platform. Factors reported to impact credibility included language used online, expertise heuristics, and bandwagon heuristics. No papers were found that assessed authenticity.

Conclusions: Credibility and authenticity are important concepts studied extensively in the marketing and communications disciplines; however, further research is required in a health context. Instagram is a less-researched platform in comparison with Facebook and Twitter.

(J Med Internet Res 2020;22(7):e17296) doi:10.2196/17296

KEYWORDS

review; trust; social media; nutrition science; health; communication; health communication

Introduction

Background

Science, particularly the discipline of nutrition science, is currently facing credibility issues in the eyes of the general public [1,2]. Although nutrition science has contributed to countless discoveries and progressions in science, it is more complicated than other scientific disciplines in multiple ways. First, food is an essential part of every human's life; thus, many people have a vested interest in nutrition and care greatly about their health [3]. Second, the significant research funding provided by the food industry creates conflicts of interest and

XSL•FO
RenderX

often high levels of skepticism from the public [1,4,5]. Researcher Marion Nestle identified 76 industry-funded studies between March and October 2015, of which 70 reported results that were favorable to the sponsor's interest, highlighting the potential bias in the industry [6]. Finally, there is an ongoing challenge to produce evidence-based science to facilitate recommendations that promote health on a population level, such as dietary guidelines [1].

The advent of the internet and, in particular, social media (SM; see Multimedia Appendix 1 for a glossary of terms) has permanently changed communication worldwide. Similar to traditional media (eg, newspapers, television, and radio), websites initially existed for one-way information dissemination through static web pages, known as *Web 1.0* [7]. However, progression to the second generation of the internet, *Web 2.0* (which includes SM), has facilitated a two-way interaction between users, creating a platform for collaboration, sharing, and socialization [7].

Currently, the use of SM in health interventions has had limited effectiveness, with participant engagement rates (Multimedia Appendix 1) being highly variable, ranging from 3% to 69% [8,9]. Health promotion organizations have maintained a one-sided communication approach, sharing serious and factual messages across their platforms, often reaching only a limited audience [10,11]. In contrast, product and corporate brands (Multimedia Appendix 1) have effectively adapted their marketing strategies to utilize the features of SM by being more likely to use hashtags to increase the reach of their posts, interact with their followers to create a two-way communication channel, and run promotions such as competitions to encourage user engagement [10]. Many large and well-known corporations such as soft drink companies are faceless and do not have a permanent ambassador and instead utilize celebrity endorsers (Multimedia Appendix 1) periodically to promote their products and improve their image and reputation [12].

Similar to celebrity endorsers, SM influencers (SMIs) or "individuals or groups of individuals who can shape attitudes and behaviours through online channels," are arguably human brands (Multimedia Appendix 1), which often enable consumers to see them as regular individuals with whom they share common values [11,13]. SMIs are often referred to as microcelebrities (Multimedia Appendix 1) as they provide insight into their private lives and create the perception that they are constantly accessible and intimately invested in their audience [14]. The number of people that connect and engage with posts by SMIs is high, as their audience consists of like-minded people, allowing brands that partner with them to target specific demographic and lifestyle groups [15]. In contrast, nutrition professionals (Multimedia Appendix 1), who typically promote evidence-based science on SM, must maintain a sense of professionalism online and, therefore, cannot create the same type of content without risking their career prospects [16].

In the existing posttruth era (Multimedia Appendix 1), science experts (including nutrition professionals) are often less highly regarded, and emotional appeals are often the most effective methods of communication [2,17]. Consumers are frequently turning to celebrities and SMIs (who hold no formal qualification) for health and lifestyle advice, creating many implications for their health and well-being [18,19]. In particular, health and wellness advice spread on online platforms tends to be misinformation rather than evidence-based information [20]. For example, the *A-list* celebrity Gwyneth Paltrow has consistently been in the public eye after controversial and dangerous health claims were made by her health and wellness brand, *Goop* [21,22]. Numerous articles on *Goop's* website and Instagram promote detoxifying the body (a process that is naturally performed by the liver) and cutting out essential food groups from the diet (eg, carbohydrates) [23]. These recommendations are based on anecdotal evidence and pseudoscience, perpetuating disordered eating habits and nutritional imbalances [24].

In the marketing literature, underlying factors such as the endorser's expertise, trustworthiness, attractiveness, and authenticity have been shown to influence people's behavior in traditional forms of advertising (eg, television commercials, celebrity partnerships) [25,26]. This review was focused on various theories and models from psychology literature that draw on the concepts of expertise, trustworthiness, and authenticity: self-determination theory (SDT; Multimedia Appendix 1), source credibility model (Multimedia Appendix 1), and the elaboration likelihood model (ELM; Multimedia Appendix 1). The following section (*Theoretical Framework*) summarizes these concepts and their nexus.

Theoretical Framework

Authenticity is the concept of "being true to the self in terms of an individual's thoughts, feelings, and behaviours reflecting their true identity" [27]. SDT encompasses the concept of authenticity with 3 primary components: autonomy, competence, and relatedness [28]. SDT posits that authenticity involves an individual's engagement in intrinsically motivated behaviors, *behaviors that come from a person's innate desires and passions* [28]. In marketing literature, individuals tend to perceive another person (eg, a celebrity) as authentic when the other person's actions reflect his or her autonomous, self-determining, true self [26]. Celebrities who are perceived as authentic have a higher level of influence over others, both online and offline [29]. Many young people do not verify information found online, leaving them particularly susceptible to celebrity influence [30]. Therefore, exploring the factors that impact authenticity on SM could be useful to inform health communication and behavior change campaigns.

The source credibility model suggests that a credibility judgment is determined based on the source's (eg, a celebrity's) expertise, trustworthiness, and attractiveness [31]. The 3 dimensions of source credibility differ: (1) expertise refers to the perceived knowledge and education level of the source, (2) trust refers to the listener's confidence in and level of acceptance of the speaker or message, and (3) attractiveness refers to the perceived physical attractiveness of the source. As a credible message is influential, many individuals and brands place a high level of importance on creating and maintaining credibility [32]. Typically, SMIs are perceived as credible as they are physically attractive and share aesthetically pleasing photos relevant to

XSL•FO
RenderX

their field of perceived expertise (eg, health) to showcase their desirable lifestyle [33]. Similarly, large corporations utilize attractive and trustworthy celebrity endorsers to be the face of their marketing campaigns to increase the brand's credibility [25]. More recently, the use of SM platforms for health advice has made it difficult for laypeople to differentiate a credible, evidence-based message from a noncredible message in an environment where everyone appears to have expertise [34].

Message content is assessed through cognitive processing (Multimedia Appendix 1), explained by the ELM (Multimedia Appendix 1). The ELM describes how people manage the information they encounter and the way in which it influences attitude change [35,36]. There are 2 processing routes: central and peripheral [35,36]. This review focuses on the peripheral route of processing, where messages are evaluated using heuristics as there is low motivation to critically evaluate the content of a message [37]. On SM, heuristics are commonly used to assess the credibility of a message, for example, using celebrity endorsers to promote a product triggers familiarity and infers the product's credibility by association. Before Facebook (Australia only) and Instagram trialed the removal of the number of publicly visible likes on a post, consumers assessed credibility through bandwagon heuristics (Multimedia Appendix 1), triggered by a mass of user opinion (eg, seeing a high number of comments on a SM post) [38]. The herd mentality (Multimedia Appendix 1) of liking what others like arises through the process of status-seeking and the need to be associated with others [39]. Another common way for consumers to make prompt judgments is via the expertise heuristic, cued when a consumer sees an official authority (such as an organization) as the source of information, whether it is an SM post, a news article, or a website [40]. By quickly associating an organization or individual as an expert source, less motivation is required to assess source credibility.

Purpose and Objective

Previous reviews have assessed the credibility of health information online, finding that factors such as clear website layout and professional design increased credibility [41-43]. However, much of this research is not applicable to SM platforms, which are often less curated and focus on fast-paced status updates.

As SM is a relatively new area of research, a scoping review was considered the most appropriate method to explore the topic area. Our research in applying social marketing techniques to the field of nutrition has led us to recognize the importance of using marketing and communication techniques, particularly on SM [9,11,44]. The aim of this scoping review was to understand and explore the factors that affect consumers' perceptions of message and source credibility (ie, expertise and trustworthiness) and authenticity on SM platforms to better inform nutrition science SM communication best practices. A secondary objective was to examine the fields that are currently undertaking this type of research and the theories used to inform the research.

Methods

Search Strategy and Databases

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Scoping Review Checklist and the Joanna Briggs Institute Reviewer's Manual were used throughout the review process [45,46].

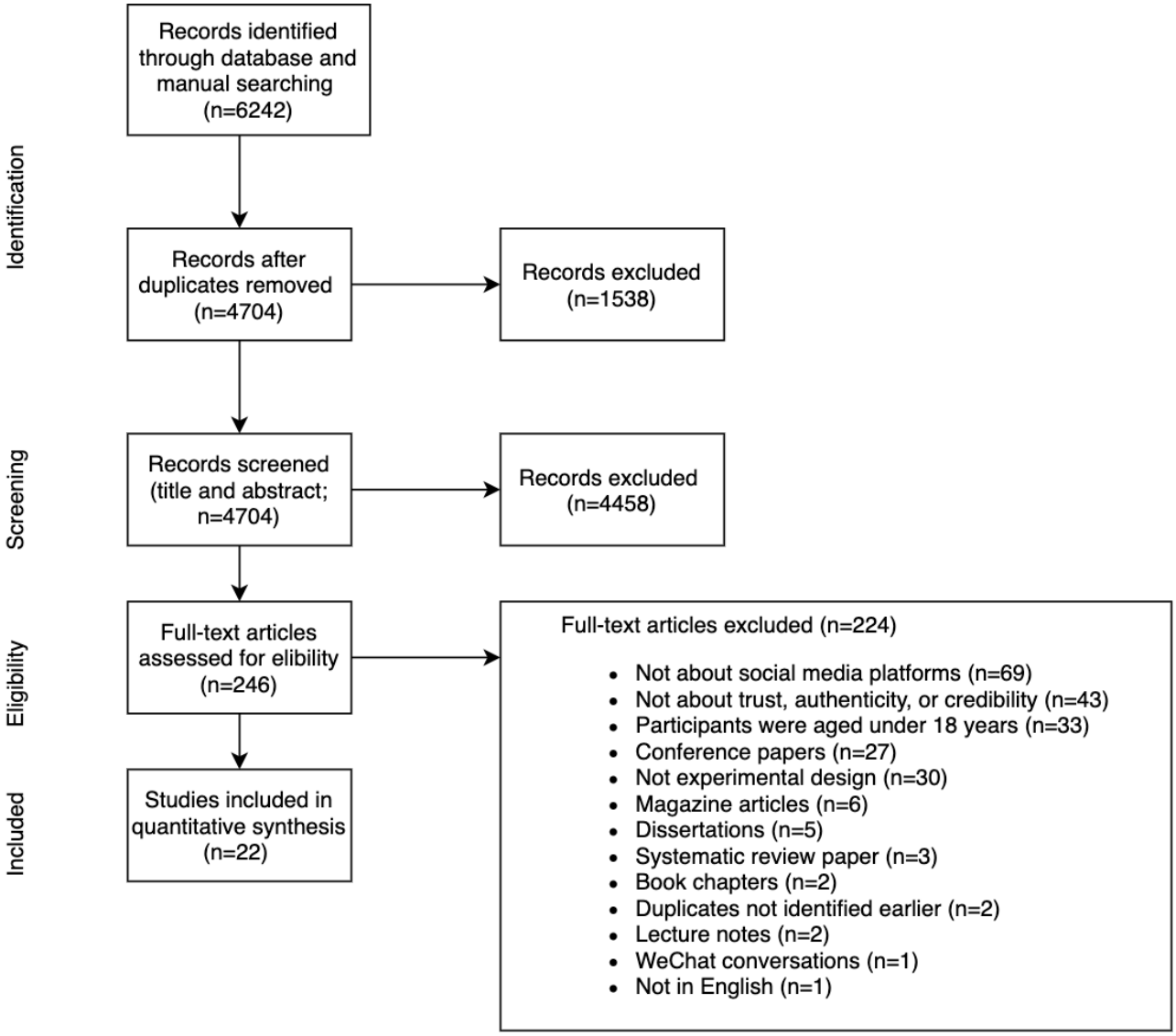
Key databases from health, psychology, and business disciplines were used to conduct the final search in conjunction with a university librarian on March 27, 2019. Cumulative Index of Nursing and Allied Health Literature (CINAHL) Plus (422 results), Scopus (2199 results), Excerpta Medica dataBASE (EMBASE; 697 results), Ovid Medical Literature Analysis and Retrieval System Online (MEDLINE; 326 results), PsycINFO (1223 results), and Business Source Complete (1375 results) were searched for title, abstract, and keywords to identify the initial 6242 articles (example search strategy provided in Multimedia Appendix 2). All articles were imported into Covidence online software (Veritas Health Innovation) to manage the reviewing process. Manual searches were conducted by checking the reference lists of included studies to identify additional papers that the search may have missed; however, no papers were found.

Inclusion and Exclusion Criteria

The inclusion criteria were experimental studies (ie, stimuli provided), studies focusing on microblogs (eg, Facebook, Twitter; Multimedia Appendix 1), studies focusing on healthy adult populations (as clinical populations often use SM to search for very specific health information), and studies focusing on source credibility or authenticity. No date restrictions were used. Exclusion criteria were studies involving participants aged under 18 years and clinical populations, gray literature, blogs, WeChat conversations, web-based reviews, studies that did not involve participants' perceptions, and non-English papers (Figure 1).

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews flow diagram of the search and study selection process.



Screening

Overall, 2 investigators (ELJ and AMB) independently screened the title and abstract of the included papers against the inclusion and exclusion criteria. This process was repeated for full-text screening, with any conflicts being discussed until a joint consensus was reached. There were 22 final papers (Figure 1).

Data Extraction

Data extraction was conducted independently by a researcher (ELJ) using Microsoft Excel 2019 (Microsoft Corporation) and was then cross-checked by a fellow researcher (AMB). Following data extraction, the results were collated based on various parameters such as SM platform focused on manipulation stimuli, outcome, and scales used for data collection. All papers were given a category to summarize their topic area, such as *celebrity*, *political*, *marketing*, *health*, or *general* (for studies that assessed source credibility without an overarching theme).

Results

Key Characteristics

Of the 22 papers that met the inclusion criteria, 3 papers reported on 2 separate studies, giving a total of 25 research studies [47-49]. The number of participants in the included studies ranged from 85 to 3476 [49,50]. Most studies were conducted in the United States (n=16) and had between-subject experimental designs (n=19). All studies used convenience sampling, excluding one that used random sampling [51]. Many studies (n=19) had no inclusion or exclusion criteria for recruitment, whereas only 2 studies specified exclusion criteria that required the participants to have an active SM account [52,53]. Student populations were predominant (n=17), where college students participated in exchange for course credit. Paid online participants were recruited in 5 of the studies via either Mechanical Turk (an Amazon platform) or professional market research companies [50-52,54,55]. Generally, the student population had a lower average age than the paid participants. Demographic data were not reported in 3 studies [56-58].

XSL•FO
RenderX

A total of 4 microblogging platforms were used: Twitter (n=10), Facebook (n=8), Instagram (n=3), and YouTube (n=3; Multimedia Appendix 1). Furthermore, 2 papers focused on both Facebook and Twitter [50,59]. Credibility was the most reported outcome (n=18), followed by trust (n=4); authenticity was not reported in the included studies. The predominant fields of research included communications (n=11), psychology (n=4), and marketing (n=2). The source credibility model (n=13) and the modality, agency, interactivity, and navigability (MAIN) model (n=7) were frequently used to inform research across disciplines. Other theories included social capital theory (n=2) and self-disclosure theory (n=2). The most commonly manipulated variables were the number of likes (n=4), source of the post (n=4), number of followers (n=3), and number of retweets (n=3).

Scales

All papers used a scale to assess credibility, with the most common being McCroskey and Teven's Source Credibility Scale (n=6; α=.94) and Ohanian's Source Credibility Scale (n=4; α>.8) [31,60]. McCroskey and Teven's validated scale, used to assess the credibility of political and public figures, includes 3 constructs: goodwill, trustworthiness, and competence [60]. Ohanian's validated scale, created to assess the credibility of celebrity endorsers, also includes 3 constructs: attractiveness, trustworthiness, and expertise [31]. Other studies created their own scales for data collection, sourcing items from the literature, and were not validated before use [47,51,61,62].

Main Observations

Most papers (n=15) involved the manipulation of text (eg, in a SM post) via Facebook or Twitter feeds to assess credibility (n=11), trust (n=4), or both (n=2). Studies primarily explored the tone of voice, bandwagon cues, and expertise [47-49,51,56,58,59,63-66]. Full details of the studies are reported in Multimedia Appendices 3-7.

Language Use

Overall, 3 papers included in the review assessed message credibility: how message characteristics impact credibility perceptions (Multimedia Appendix 1). In a study conducted on Facebook with a student sample (mean age 19 years), language usage was found to impact credibility judgments, with a gain-framed post (focusing on the benefits of exercise) eliciting positive emotions and increasing credibility when compared with a loss-framed post (focusing on the risks of not exercising; Tables 1 and 2; Multimedia Appendix 3) [47].

On Twitter, Houston et al [51] found that the tone of voice impacted credibility; nonopinionated tweets (written as a headline), which conveyed no personal opinion, were more credible than opinionated tweets that used humor or sarcasm and conveyed strong personal opinion among consumers aged over 18 years, participating for monetary rewards (Tables 1 and 2; Multimedia Appendix 4).

Similarly, Yilmaz and Johnston [59] compared language framed with personal experiences to depersonalized language (ie, factual and data-based) on Facebook and Twitter in a sample of 257 students (Tables 1 and 2; Multimedia Appendix 5). Differing results were found among the SM platforms; personalized Facebook posts were more competent and trustworthy than personalized tweets [59]. However, depersonalized tweets were more competent and trustworthy than depersonalized Facebook posts [59]. Thus, personalized language was an effective way to increase credibility on Facebook but not on Twitter [59].

XSL•FO
RenderX

Table 1. Main findings of included papers and their effect on credibility or trust, separated by manipulated variable: number of likes, number of followers, number of retweets, source, and language.

Outcomes/author (year) [citation]	Population group	Key significant results
Number of likes		
Borah and Xiao (2018) [47]	Students	In the 2 studies conducted, the number of likes did not affect source credibility overall when looking at Facebook posts (study 1: *P*=.93; study 2: *P*=.09)
Phua and Ahn (2016) [66]	Students	Brand trust was higher when likes were high on Facebook post (*P*<.005) or when friends' likes were high (*P*<.001). Friends' likes were more important in trust than overall total likes (*P*<.005). The number of likes had no direct effect on brand trust when the intensity of Facebook use was controlled for (*P*=.89)
Shen et al (2019) [50]	Paid online workers	Bandwagon cues did not impact credibility when looking at images on Twitter and Facebook (*P*=.85)
Number of followers		
Jin and Phua (2014) [48]	Students	A higher number of Twitter followers on the celebrity's account increased source credibility and intention to build an online friendship with the celebrity endorser for all dimensions of source credibility: physical attraction (*P*<.05), trustworthiness (*P*<.05), and competence (*P*<.01)
Lee (2018) [49]	Students	The number of followers on Facebook made a statistically significant difference on the believability of the answer (*P*<.05), with a high number of followers increasing believability. There were no significant results for trustworthiness or accuracy
Westerman et al (2011) [57]	Students	Trustworthiness indicated an inverted U-shaped relationship with the number of followers on Twitter (*P*=.02)
Number of retweets		
Lin and Spence (2018) [63]	Students	The highest level of trust (on Twitter) was when participants viewed the post with 400 retweets, followed by 40 retweets, whereas 4000 retweets had the lowest level of trust (*P*=.01). Participants perceived the highest levels of source competence when viewing the post with 40 retweets, followed by 400 retweets. The post with 4000 retweets had the lowest perceived competence (*P*=.01)
Lin and Spence (2019) [64]	Students	There were significant differences in trust perceptions across varying retweet conditions (*P*=.046). People who viewed the FDA's[a] Twitter page containing 4000 retweets were more likely to perceive lower organizational trust than the condition of 40 retweets (*P*<.05)
Lin and Spence (2016) [65]	Students	Participants perceived lowest competence when viewing a peer's Twitter page with no retweets (*P*<.001). The highest level of perceived source goodwill, trustworthiness, and competence was when participants viewed the CDC[b] page with no retweets (*P*<.001)
Source (expert, peer, or stranger)		
Borah and Xiao (2018) [47]	Students	In the 2 studies conducted on Facebook, the CDC and WebMD authors were seen as more credible than unknown authors (study 1: *P*<.01; study 2: *P*<.01)
Lin and Spence (2018) [63]	Students	Participants viewing an FDA expert's Twitter account were more likely to perceive higher trust (*P*=.01), competence (*P*<.001), and goodwill (*P*<.001) than those viewing a peer or stranger's account
Lin and Spence (2016) [65]	Students	Higher credibility was assigned to risk information from an expert compared with a peer and a stranger on Twitter (*P*<.001)
Language (message credibility)		
Borah and Xiao (2018) [47]	Students	In the 2 studies conducted, a gain-framed message was more credible than a loss-framed message on Facebook (study 1: *P*<.001; study 2: *P*<.001)
Houston et al (2018) [51]	Paid workers	Nonopinionated tweets were perceived as more credible than opinionated tweets (*P*<.001)
Yilmaz and Johnson (2016) [59]	Students	Personalized status updates on Facebook were seen as more competent and trustworthy than personalized tweets (*P*=.007, *P*=.001 respectively). Depersonalized tweets were more trustworthy than the source of depersonalized status updates on Facebook

[a]FDA: Food and Drug Administration.

[b]CDC: Centers for Disease Control and Prevention.

XSL•FO
RenderX

Table 2. Included papers, main outcomes, and the effect on either brand trust, message credibility, or source credibility (including trustworthiness, believability, and competence) as specified in their results.

Factors	Platform	Population	Outcome	Result[a]	Relevant papers
Language use[b]					
Gain-framed language					
	Facebook	Student	Message credibility[b,c]	Increase	Borah and Xiao [47]
Personalized language					
	Twitter	Student	Competence and trustworthiness	Decrease	Yilmaz and Johnson [59]
	Facebook	Student	Competence and trustworthiness	Increase	Yilmaz and Johnson [59]
Exposure to civil discussion					
	Facebook	Student	Trustworthiness	Increase	Antoci et al [67]
Nonopinionated language					
	Twitter	Paid worker	Message credibility	Increase	Houston et al [51]
	YouTube	Student	Message credibility	No effect	Zimmermann and Jucks [68]
Bandwagon heuristics[b]					
High number of likes					
	Facebook	Student	Source credibility and trustworthiness	No effect	Borah and Xiao [47]
		Student	Brand trust	Increase	Phua and Ahn [66]
High number of followers					
	Facebook	Student	Believability	Increase	Lee [49]
	Twitter	Student	Source credibility	Westerman and Spence: unclear; Phua and Ahn: increase	Westerman and Spence [57]; Phua and Ahn [66]
Narrow ratio of the number of followers to the number of follows					
	Twitter	Student	Competency	Increase	Westerman and Spence [57]
High number of retweets					
	Twitter	Student	Trustworthiness	Decrease	Lin and Spence [63-65]
High number of friends					
	Facebook	Student	Believability and trustworthiness	Increase	Lee [49]
Expertise heuristic[b]					
Post from expert source	Facebook and Twitter	Student	Source credibility	Increase	Borah and Xiao [47]; Lin and Spence [63,65]
Other[b]					
Interaction with followers	Twitter	Student	Source credibility	Increase	Jahng and Littau [62]
High perceived privacy control	Facebook	Student	Trust	Increase	Antoci et al [67]
Positive brand attitude	Instagram	Paid worker	Brand credibility	Increase	De Veirman and Hudders [52]; Jin and Muqaddam [55]
Prosocial attitude online	Twitter	Student	Source credibility	Increase	Jin and Phua [48]
Recency of updates (frequent)	Twitter	Student	Source credibility	Increase	Westerman and Spence [56]

XSL•FO
RenderX

Factors	Platform	Population	Outcome	Result[a]	Relevant papers
Snapshot aesthetic (vs studio aesthetic)	Instagram	Paid worker	Brand credibility	Increase	Colliander and Marder [54]
Preexisting photoshop/internet skills (when looking at photo-shopped images)	Twitter and Facebook	Paid worker	Source credibility	Decrease	Shen et al [50]
Ethos message appeal (compared with logos and pathos)	YouTube	Student	Source credibility	Increase	English et al [61]
Consumer-generated advertising (compared with firm-generated advertising)	YouTube	Student	Source credibility	Increase	Lee et al [69]
Caucasian ethnicity (compared with African American)	Facebook	Student	Source credibility	Increase	Spence et al [58]

[a]On the basis of reported results from studies summarized in Multimedia Appendices 3-7.

[b]For further context, explanation, and examples of these factors, refer to Multimedia Appendices 3-7.

[c]Credibility comprises trustworthiness, expertise, and sometimes attractiveness, depending on the individual paper.

Bandwagon Heuristics

Bandwagon cues such as the number of followers, number of retweets, and number of likes were a way in which student participants (mean age range 19-22.9 years) assessed source credibility across 30% (8/22) of the included papers; however, the findings were inconsistent among studies [47-49,56,63-66].

Borah and Xiao [47] assessed the number of likes on a Facebook post (150 likes or 2 likes) and found that the manipulation had no significant effect on credibility or trust levels (Tables 1 and 2; Multimedia Appendix 3). This differed from the study by Phua and Ahn [66], which found that brand trust increased when overall likes or friends' likes on the post were higher (Tables 1 and 2; Multimedia Appendix 3). Similarly, Lee's experiment [49], which involved a question-and-answer format on a Facebook post, found that when the source had a higher number of followers, the answer posted was rated as more believable than an answer posted by a source with a lower number of followers (Multimedia Appendix 3). However, this relationship was not observed for the accuracy or trustworthiness dimensions of source credibility. Lee's second experiment assessed the credibility of a source with a high number of friends compared with a low number, finding that the source with more friends was seen as more believable and trustworthy (Multimedia Appendix 3) [49].

On Twitter, Westerman et al [57] found no linear relationship between the number of followers and credibility. In fact, too many followers (n=70,000) or too few followers (n=70) reduced the level of trust compared to those with 7000 followers, who had the greatest level of trust (Multimedia Appendix 4) [57]. However, Jin and Phua [48] reported that a higher number of followers (n=14,677,050) increased source credibility (Tables 1 and 2; Multimedia Appendix 4). The ratio of the number of followers to the number of follows on Twitter provided a cue for participants to assess credibility, with a narrow gap (ie, similar ratio of Twitter followers to follows) being perceived as more competent than a wide gap (ie, more Twitter followers compared with follows; Multimedia Appendix 4) [57]. High numbers of retweets (from other Twitter profiles) reduced credibility in 3 studies [63-65]. When using an organization's Twitter page (*Centers for Disease Control and Prevention* [CDC] or *Food and Drug Administration* [FDA]), a post with 4000 retweets was found to be less trustworthy and competent than posts with 40 or 400 retweets (Table 1; Multimedia Appendix 4) [63-65].

Expertise Heuristics

Manipulating the source of the Facebook or Twitter posts was found to impact source credibility in 3 studies with student participants [47,63,65]. Expert sources, such as the CDC, were perceived as more credible than strangers when disseminating health information in Facebook status updates (Multimedia Appendix 3) [47].

Similarly, expert sources (eg, FDA) were considered more trustworthy and competent and, thus, overall more credible than strangers or peers on Twitter (Tables 1 and 2; Multimedia Appendix 4) [63,65]. Jin and Phua [48] assessed celebrity source credibility by manipulating a news story of the celebrity (shared on Twitter) to be either prosocial (donating to a charity) or antisocial (drug abuse; Multimedia Appendix 4). Consumers were more likely to identify with prosocial celebrity and perceive them as more credible, suggesting that reputation can affect credibility perceptions [48].

Discussion

Principal Findings

Surprisingly, there were no studies in the field of health and nutrition research included in this scoping review; however, there are some important learnings that could be utilized for nutrition and health communication. There were many different factors that affected the perceived credibility of a message and source on SM, such as language usage, expertise heuristics, and bandwagon heuristics. However, no information was found on the factors affecting perceived authenticity in this context. The scales used as well as the different models and theories to inform various fields of research are reported.

XSL•FO
RenderX

Language Usage

The results of the included studies indicated that the language used in tweets affected message credibility. Personalization was more effective on Facebook, whereas depersonalization was more effective on Twitter [59]. This can be explained by the differing functions of these platforms. The audience of a tweet is largely unknown as the public nature of Twitter allows anyone with or without an account to view a profile (provided the profile is not locked) [70]. Having a large and primarily unknown audience creates a challenge to balance self-expression and impression management; users want to share information but do not want to be negatively judged [70]. The fear of other people's negative opinions and judgment makes it easier to post factual, depersonalized tweets with no personal opinion involved [70]. Furthermore, the linguistic style of Twitter differs from other SM platforms because of the 280-character (previously 140) limit for tweets [59]. When users share long personal stories, it differs from the concise updates that are typically shared on Twitter, reducing the contextual appropriateness of the information [59]. This differs from Facebook, which has an extremely high character limit for status updates (63,206 characters) and requires both parties to accept a friend request before viewing each other's content (provided the privacy of the page is not set to public) [59]. Thus, when a personal status update is made on Facebook, the recipients are generally people who have an existing relationship with the user and an established level of trust, instantly increasing the credibility of the post [59]. Therefore, message characteristics such as language style and the platform being used should be considered by health professionals when creating SM posts.

Expertise Heuristics

Each study in the review that compared an expert to a stranger or peer found that the expert was more credible [47,63,65]. This is explained by the expertise heuristic, which is triggered when a consumer sees an official authority as a source [71]. For example, when tweets were posted from the CDC's Twitter account, they were seen as more credible than when an unknown person tweeted [65]. These findings are consistent with the existing literature on website credibility, whereby listing an author's affiliations, credentials, or qualifications triggers the expertise heuristic and increases the credibility of the information or message presented [72,73]. In addition, a recent review found that websites run by health institutions (such as the CDC) were considered more trustworthy than private websites [42]. Health practitioners should include their credentials (relevant to their field of research) on their SM page, so that their expertise is clear to the audience and credibility can be established.

Bandwagon Heuristics

A key focus of the included papers was bandwagon heuristics, which relate to the number of likes, followers, or retweets assigned to information on SM [71]. It is known that people base their own decisions on other people's endorsements and opinions, particularly when purchasing products [74]. Typically, a bigger bandwagon (ie, a high number of likes, comments, or shares) will result in a greater perception of source credibility [40,75]. However, the reviewed papers that manipulated bandwagon cues provided differing results than expected; a high number of retweets (n=4000) had the lowest perceived trustworthiness of the conditions, and two papers found no difference in credibility between the low and high like manipulations (Multimedia Appendices 3 and 4) [47,50,63,64]. Some consumers perceived popular content as having lower credibility, described as a *reverse bandwagon heuristic* (Multimedia Appendix 1) or the *snob effect* (Multimedia Appendix 1) [39,76]. This arises from the consumer's need to identify as an individual in society, without conforming to social norms [39,76]. Some people do not follow trends and are not willing to follow others without independently thinking and making their own judgments. The *snob effect* causes a deviation from the norm (ie, liking the post) and, thus, when a consumer sees a high amount of engagement on SM, it can result in negative opinions of the source and information presented, reducing the perceived credibility [63]. However, in 2019, attempting to reduce the pressure people feel when they post online, Instagram (globally) and Facebook (in Australia) trialed removing the number of likes on posts so that likes are no longer publicly visible to others [77]. Rather than displaying the number of people who have liked the post, the display now shows one user who has liked the post followed by the phrase "and others" [77,78]. Therefore, in the future, the use of bandwagon cues to judge credibility may be less prevalent as the engagement of a post can no longer be viewed publicly in a numerical format [77].

Scales

The most commonly used scale in the research papers was McCroskey and Teven's Source Credibility Scale (n=6), used to assess the credibility of a source of information (ie, a person). In contrast, message credibility was assessed by examining the characteristics of a message such as the structure, perceived accuracy, and language used [79]. Flanagin and Metzger's Scale of Message Credibility Online (2013) assesses 5 dimensions of credibility—believability, accuracy, trustworthiness, bias, and completeness—and was utilized in 2 papers [49,50]. Other papers that assessed message credibility adapted their own scale from multiple sources in the literature to evaluate the language used in Facebook posts, tweets, or message appeals in YouTube videos [47,51,61]. Internal reliability was assessed with Cronbach alpha values at an acceptable level (α>.7) in all three papers that adapted their own scale. However, there was no exploratory or confirmatory factor analysis conducted, creating implications for the validity and generalizability of the research.

Fields of Research

Communications and psychology were the predominant fields of research within the included papers. The theories and models used to underpin research overlapped among disciplines, with the source credibility model being the most common within communications, information research, psychology, and business. The MAIN model (a digital extension of source credibility model) was also used frequently in communications, information research, computer science, and psychology. Previously, source credibility and its impact have been investigated more broadly in the health discipline to assess consumers' perceptions of health information online, but this

XSL•FO
RenderX

was limited to websites and did not include SM [80]. To our knowledge, health communication research has not utilized the MAIN model or source credibility model. Using these models to inform research in different fields resulted in the perception of credible spokespeople on SM. Thus, these models should be utilized to increase the effectiveness of health communication within nutrition research.

Gaps in Knowledge

On the basis of the search strategy and the inclusion and exclusion criteria, the concept of authenticity has not been explored in this specific context. In addition, there was a lack of research regarding SMIs or nutrition professionals as the source of SM posts. Most studies focused on organizations or an unknown fictional author as the source. As SMIs are key to digital marketing and health communication, research into consumer perceptions of the source credibility and authenticity of SMIs would be beneficial to further understand how they communicate effectively with their followers. Results from these future studies would be beneficial for informing the delivery of health communication in various digital formats.

Strengths

The strength of the papers included in the scoping review was the large sample size that was achieved (range of the number of participants from included studies was 85-3476). As most papers had a student cohort participating for university credit, large numbers of participants were recruited. In addition, many of the papers completed a manipulation check during the pilot of the survey to ensure that the mock scenario they were creating was robust, for example, testing if the *high retweet* condition of 4000 retweets was actually considered *high* by the participants, limiting the confounding factors that could arise if the manipulations were perceived differently than intended [63].

Limitations

As most papers (n=21) used convenience sampling, selection biases were inherent. Furthermore, generalizability was limited to the geographical area in which the research was conducted, making it difficult to draw conclusions without conducting further research. Student samples were predominant (n=17), further limiting the variability and generalizability of results as student samples (referred to as western, educated, industrialized, rich, and democratic [WEIRD]) [81] are seen as more homogenous in terms of education level and socioeconomic status than the general public [81,82]. Furthermore, cross-sectional questionnaires were used as the method of data collection, which is self-reported, adding a level of bias to the results as participants can be deceiving intentionally or unintentionally [83]. A methodological limitation of undertaking a scoping review is the omission of quality appraisal of studies, usually conducted during data extraction in a systematic literature review.

Conclusions and Recommendations

Fostering credibility online should be considered by corporate and human health brands to create a stronger relationship with their audience. This scoping review highlighted that message and source credibility can be affected by language usage, expertise heuristics, and bandwagon cues. Gaps in the literature were identified, highlighting the need for further research on SM platforms, as Instagram and YouTube were studied less than Facebook and Twitter. The main field of research identified from the included papers was communications, with no papers from health or nutrition science. Currently, there is a limited understanding of the use of SMIs and science experts to relay health messages. Further research needs to be undertaken to apply information from communications (on source and message credibility and authenticity) in a health context and in populations other than students.

Acknowledgments

The authors thank Monash University librarian, Anne Young, for assistance when developing the search strategy.

Communicating Health is funded by the National Health and Medical Research Council (grant number GNT1115496).

Authors' Contributions

ELJ and AMB completed the screening and data extraction of the included papers. The manuscript was drafted by ELJ, with critical revisions suggested by TAM and JI. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Glossary of terms.
[DOCX File , 25 KB - jmir_v22i7e17296_app1.docx]

Multimedia Appendix 2
Search strategy (Scopus database).
[DOCX File , 13 KB - jmir_v22i7e17296_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Research studies assessing trust and credibility on Facebook.
[DOCX File , 20 KB - jmir_v22i7e17296_app3.docx]

Multimedia Appendix 4
Research studies assessing trust and credibility on Twitter.
[DOCX File , 19 KB - jmir_v22i7e17296_app4.docx]

Multimedia Appendix 5
Research studies assessing credibility on both Facebook and Twitter.
[DOCX File , 15 KB - jmir_v22i7e17296_app5.docx]

Multimedia Appendix 6
Research studies assessing credibility on Instagram.
[DOCX File , 16 KB - jmir_v22i7e17296_app6.docx]

Multimedia Appendix 7
Research studies assessing credibility on YouTube.
[DOCX File , 16 KB - jmir_v22i7e17296_app7.docx]

References

1. Penders B, Wolters A, Feskens EF, Brouns F, Huber M, Maeckelberghe EL, et al. Capable and credible? Challenging nutrition science. Eur J Nutr 2017 Sep;56(6):2009-2012 [FREE Full text] [doi: 10.1007/s00394-017-1507-y] [Medline: 28718015]
2. Penders B. Why public dismissal of nutrition science makes sense: post-truth, public accountability and dietary credibility. Br Food J 2018;120(9):1953-1964 [FREE Full text] [doi: 10.1108/BFJ-10-2017-0558] [Medline: 30581197]
3. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008 Oct(8):1-9. [Medline: 18946947]
4. Nestle M. Corporate funding of food and nutrition research: science or marketing? JAMA Intern Med 2016 Jan;176(1):13-14. [doi: 10.1001/jamainternmed.2015.6667] [Medline: 26595855]
5. Anahad O. Coca-Cola Funds Scientists Who Shift Blame for Obesity Away From Bad Diets. The New York Times. 2015. URL: https://well.blogs.nytimes.com/2015/08/09/coca-cola-funds-scientists-who-shift-blame-for-obesity-away-from-bad-diets/?_r=0 [accessed 2019-10-05]
6. Nestle M. Viewpoint: Food-industry Funding of Food and Nutrition Research. Food Politics by Marion Nestle. 2016. URL: https://www.foodpolitics.com/2016/01/viewpoint-food-industry-funding-of-food-and-nutrition-research/ [accessed 2019-09-10]
7. Dooley JA, Jones SC, Iverson D. Using web 2.0 for health promotion and social marketing efforts: lessons learned from web 2.0 experts. Health Mark Q 2014;31(2):178-196. [doi: 10.1080/07359683.2014.907204] [Medline: 24878406]
8. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
9. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MS. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
10. Lim MS, Hare JD, Carrotte ER, Dietze PM. An investigation of strategies used in alcohol brand marketing and alcohol-related health promotion on Facebook. Digit Health 2016;2:2055207616647305 [FREE Full text] [doi: 10.1177/2055207616647305] [Medline: 29942554]
11. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What people 'like': analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
12. Davies F, Slater S. Exploring the power of sporting celebrity endorsements: a comparison of contrasting sports. New York, USA: Springer; 2017. [doi: 10.1007/978-3-319-50008-9_83]
13. Sara. The Rise of Social Media Influencers (Yes, They Matter!). Pitstop Marketing. 2018. URL: https://pitstopmarketing.com.au/the-rise-of-social-media-influencers/ [accessed 2019-06-18]
14. Khamis S, Ang L, Welling R. Self-branding, 'micro-celebrity' and the rise of social media influencers. Celebrity Stud 2016 Aug 25;8(2):191-208. [doi: 10.1080/19392397.2016.1218292]
15. The Definitive Guide to Influencer Marketing – Everything You Need to Know. Influencer Marketing Hub. 2018. URL: https://influencermarketinghub.com/the-definitive-guide-to-influencer-marketing/ [accessed 2019-05-03]

XSL•FO
RenderX

16. Probst YC, Peng Q. Social media in dietetics: insights into use and user networks. Nutr Diet 2019 Sep;76(4):414-420. [doi: 10.1111/1747-0080.12488] [Medline: 30370651]
17. Laybats C, Tredinnick L. Post truth, information, and emotion. Bus Inf Rev 2016 Dec 22;33(4):204-206. [doi: 10.1177/0266382116680741]
18. van Reijmersdal E, Mariea ER, Hoy G. The Role of Social Media Influencers in the Lives of Children and Adolescents. Frontiers: Peer Reviewed Articles - Open Access Journals. 2019. URL: https://www.frontiersin.org/research-topics/9295/the-role-of-social-media-influencers-in-the-lives-of-children-and-adolescents [accessed 2019-09-01]
19. Hoffman SJ, Tan C. Biological, psychological and social processes that explain celebrities' influence on patients' health-related behaviors. Arch Public Health 2015;73(1):3 [FREE Full text] [doi: 10.1186/2049-3258-73-3] [Medline: 25973193]
20. Tonsaker T, Bartlett G, Trpkov C. Health information on the internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
21. Young S. Gwyneth Paltrow's Goop Reported to British Regulators. The Independent. 2018. URL: https://www.independent.co.uk/life-style/goop-gwyneth-paltrow-mother-load-uk-dangerous-claims-trading-standards-a8606151.html [accessed 2018-10-29]
22. d'Entremont Y. The Unbearable Wrongness of Gwyneth Paltrow. The Outline. 2017 Aug 13. URL: https://theoutline.com/post/1394/the-unbearable-wrongness-of-gwyneth-paltrow [accessed 2019-08-13]
23. Hosie R. A Nutritionist Says Goop's Advice for Losing Weight Fast is 'Extremely Damaging'. The Independent. 2017. URL: http://www.independent.co.uk/life-style/health-and-families/goop-fast-weight-loss-advice-diet-tracy-anderson-damaging-nutritionist-gwyneth-paltrow-a8091241.html [accessed 2019-08-13]
24. Turner PG, Lefevre CE. Instagram use is linked to increased symptoms of orthorexia nervosa. Eat Weight Disord 2017 Jun;22(2):277-284 [FREE Full text] [doi: 10.1007/s40519-017-0364-2] [Medline: 28251592]
25. Amos C, Holmes G, Strutton D. Exploring the relationship between celebrity endorser effects and advertising effectiveness. Int J Advert 2015 Jan 6;27(2):209-234. [doi: 10.1080/02650487.2008.11073052]
26. Ilicic J, Webster CM. Being true to oneself: investigating celebrity brand authenticity. Psychol Mark 2016 May 10;33(6):410-420. [doi: 10.1002/mar.20887]
27. van Leeuwen T. What is authenticity? Discourse Stud 2001 Nov;3(4):392-397. [doi: 10.1177/1461445601003004003]
28. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
29. Ghani S. Impact of celebrity credibility on advertising effectiveness. Pak J Commer Soc Sci 2013;7(1):107-127 [FREE Full text]
30. Krishan R. Social Media Authenticity Issues: Information, Verification and Dissemination. In: Conference on Media Literacy: Issues and Challenges. 2014 Presented at: EMEDUS'14; November 23, 2014; Kurukshetra University, Kurukshetra URL: https://www.researchgate.net/publication/305681408_Social_Media_Authenticity_Issues_Information_Verification_and_Dissemination
31. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990 Oct;19(3):39-52. [doi: 10.1080/00913367.1990.10673191]
32. Pornpitakpan C. The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Social Pyschol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
33. Lou C, Yuan S. Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. J Int Adv 2019 Feb 12;19(1):58-73. [doi: 10.1080/15252019.2018.1533501]
34. Hocevar K, Metzger M, Flanagin A. Source Credibility, Expertise, and Trust in Health and Risk Messaging. Oxford Research Encyclopedias. 2017. URL: https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-287 [accessed 2020-06-25]
35. Petty R, Cacioppo J. The Elaboration Likelihood Model of Persuasion. New York, USA: Springer; 1986.
36. Brennan L, Binney W, Parker L, Aleti T, Nguyen D. Social Marketing and Behaviour Change: Models, Theory and Applications. Massachusetts, USA: Edward Elgar Publishing; 2014.
37. White H. Elaboration Likelihood Model. Interaction Design Foundation Reviews. 2011. URL: https://www.interaction-design.org/literature/article/elaboration-likelihood-model-theory-using-elm-to-get-inside-the-user-s-mind [accessed 2020-06-25]
38. Kim J, Gambino A. Do we trust the crowd or information system? Effects of personalization and bandwagon cues on users' attitudes and behavioral intentions toward a restaurant recommendation website. Comput Hum Behav 2016 Dec;65:369-379. [doi: 10.1016/j.chb.2016.08.038]
39. Granovetter M, Soong R. Threshold models of interpersonal effects in consumer demand. J Econ Behav Organ 1986 Mar;7(1):83-99. [doi: 10.1016/0167-2681(86)90023-5]
40. Hu Y, Sundar SS. Effects of online health sources on credibility and behavioral intentions. Commun Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
41. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]

XSL•FO
RenderX

42. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 2;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
43. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369. [doi: 10.1177/1460458214559432] [Medline: 25518944]
44. Lombard C, Brennan L, Reid M, Klassen KM, Palermo C, Walker T, et al. Communicating health-optimising young adults' engagement with health messages using social media: study protocol. Nutr Diet 2018 Nov;75(5):509-519. [doi: 10.1111/1747-0080.12448] [Medline: 30009396]
45. Peters MG, McInerney P, Baldini SC, Khalil H, Parker D. Scoping Reviews. Joanna Briggs Institute. 2017. URL: https://wiki.joannabriggs.org/display/MANUAL/Chapter+11%3A+Scoping+reviews [accessed 2019-03-01]
46. PRISMA for Scoping Reviews. PRISMA. 2018. URL: http://www.prisma-statement.org/Extensions/ScopingReviews [accessed 2019-03-01]
47. Borah P, Xiao X. The importance of 'likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
48. Jin SA, Phua J. Following celebrities' tweets about brands: the impact of Twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. J Advert 2014 Apr 24;43(2):181-195. [doi: 10.1080/00913367.2013.827606]
49. Lee SY. Effects of relational characteristics of an answerer on perceived credibility of informational posts on social networking sites: The case of Facebook. Inf Res 2018;2(3):- [FREE Full text]
50. Shen C, Kasra M, Pan W, Benefield G, Malloch Y, O'Brien J. Fake images: the effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. SSRN J 2019:-. [doi: 10.2139/ssrn.3234129]
51. Houston JB, McKinney MS, Thorson E, Hawthorne J, Wolfgang JD, Swasy A. The twitterization of journalism: user perceptions of news tweets. Journalism 2018 Mar 23;21(5):614-632. [doi: 10.1177/1464884918764454]
52. de Veirman M, Hudders L. Disclosing sponsored Instagram posts: the role of material connection with the brand and message-sidedness when disclosing covert advertising. Int J Advert 2019 Feb 11;39(1):94-130. [doi: 10.1080/02650487.2019.1575108]
53. Ardiansyah Y, Harrigan P, Soutar GN, Daly TM. Antecedents to consumer peer communication through social advertising: a self-disclosure theory perspective. J Int Adv 2018 Mar 14;18(1):55-71. [doi: 10.1080/15252019.2018.1437854]
54. Colliander J, Marder B. 'Snap happy' brands: increasing publicity effectiveness through a snapshot aesthetic when marketing a brand on Instagram. Comput Hum Behav 2018 Jan;78:34-43. [doi: 10.1016/j.chb.2017.09.015]
55. Jin SV, Muqaddam A. Product placement 2.0: 'do brands need influencers, or do influencers need brands?'. J Brand Manag 2019 Feb 11;26(5):522-537. [doi: 10.1057/s41262-019-00151-z]
56. Westerman D, Spence PR, van der Heide B. Social media as information source: recency of updates and credibility of information. J Comput-Mediat Comm 2013 Nov 8;19(2):171-183. [doi: 10.1111/jcc4.12041]
57. Westerman D, Spence PR, van der Heide B. A social network as information: the effect of system generated reports of connectedness on credibility on Twitter. Comput Hum Behav 2012 Jan;28(1):199-206. [doi: 10.1016/j.chb.2011.09.001]
58. Spence PR, Lachlan KA, Westerman D, Spates SA. Where the gates matter less: ethnicity and perceived source credibility in social media health messages. Howard J Commun 2013 Jan;24(1):1-16. [doi: 10.1080/10646175.2013.748593]
59. Yilmaz G, Johnson JM. Tweeting facts, Facebooking lives: the influence of language use and modality on online source credibility. Commun Res Rep 2016 Apr 13;33(2):137-144. [doi: 10.1080/08824096.2016.1155047]
60. McCroskey JC, Teven JJ. Goodwill: a reexamination of the construct and its measurement. Commun Monogr 2009 Jun 2;66(1):90-103. [doi: 10.1080/03637759909376464]
61. English K, Sweetser KD, Ancu M. YouTube-ification of political talk: an examination of persuasion appeals in viral video. Am Behav Sci 2011 Mar 21;55(6):733-748. [doi: 10.1177/0002764211398090]
62. Jahng MR, Littau J. Interacting is believing. Journalism Mass Commun Q 2015 Oct 9;93(1):38-58. [doi: 10.1177/1077699015606680]
63. Lin X, Spence PR. Identity on social networks as a cue: identity, retweets, and credibility. Commun Stud 2018 Jul 6;69(5):461-482. [doi: 10.1080/10510974.2018.1489295]
64. Lin X, Spence PR. Others share this message, so we can trust it? An examination of bandwagon cues on organizational trust in risk. Inf Process Manag 2019 Jul;56(4):1559-1564. [doi: 10.1016/j.ipm.2018.10.006]
65. Lin X, Spence PR, Lachlan KA. Social media and credibility indicators: the effect of influence cues. Comput Hum Behav 2016 Oct;63:264-271. [doi: 10.1016/j.chb.2016.05.002]
66. Phua J, Ahn SJ. Explicating the 'like' on Facebook brand pages: the effect of intensity of Facebook use, number of overall 'likes', and number of friends' 'likes' on consumers' brand outcomes. J Mark Commun 2014 Jul 30;22(5):544-559. [doi: 10.1080/13527266.2014.941000]
67. Antoci A, Bonelli L, Paglieri F, Reggiani T, Sabatini F. Civility and trust in social media. J Econ Behav Organ 2019 Apr;160:83-99. [doi: 10.1016/j.jebo.2019.02.026]

XSL•FO
RenderX

68. Zimmermann M, Jucks R. With a view to the side: YouTube's sidebar and YouTuber's linguistic style as hints for trust-related evaluations. Int J Hum Comput Interact 2018 Sep 14;35(14):1279-1291. [doi: 10.1080/10447318.2018.1519165]
69. Lee JK, Lee S, Hansen SS. Source credibility in consumer-generated advertising in YouTube: the moderating role of personality. Curr Psychol 2016 Jun 27;36(4):849-860. [doi: 10.1007/s12144-016-9474-7]
70. Marwick AE, boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc 2010 Jul 7;13(1):114-133. [doi: 10.1177/1461444810365313]
71. Sundar SS. The MAIN Model : A Heuristic Approach to Understanding Technology Effects on Credibility. Semantic Scholar. 2007. URL: https://pdfs.semanticscholar.org/de80/aa094f380342a632eadb0ee8d4221e8920ba.pdf [accessed 2020-06-18]
72. Metzger MJ, Flanagin AJ, Eyal K, Lemus DR, Mccann RM. Credibility for the 21st century: integrating perspectives on source, message, and media credibility in the contemporary media environment. Ann Int Commun Assoc 2016 May 18;27(1):293-335. [doi: 10.1080/23808985.2003.11679029]
73. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Nov;58(13):2078-2091. [doi: 10.1002/asi.20672]
74. Wei P, Lu H. An examination of the celebrity endorsements and online customer reviews influence female consumers' shopping behavior. Comput Hum Behav 2013 Jan;29(1):193-201 [FREE Full text] [doi: 10.1016/j.chb.2012.08.005]
75. Metzger M, Flanagin A, Medders R. Social and heuristic approaches to credibility evaluation online. J Commun 2010;60(3):413-439. [doi: 10.1111/j.1460-2466.2010.01488.x]
76. Leibenstein H. Bandwagon, snob, and veblen effects in the theory of consumers' demand. Q J Econ 1950 May;64(2):183. [doi: 10.2307/1882692]
77. Derwin J. Instagram is Getting Rid of the Total Number of Likes in Australia – And It Could Soon Be the New Normal. Business Insider. 2019. URL: https://www.businessinsider.com.au/instagram-is-getting-rid-of-the-total-number-of-likes-in-australia-and-it-could-soon-be-the-new-normal-2019-7 [accessed 2019-06-18]
78. Facebook is Hiding the Number of 'likes' on Posts in Australia. ABC (Australian Broadcasting Corporation). 2019. URL: https://www.abc.net.au/news/science/2019-09-27/facebook-follows-instagram-and-hides-likes-in-australia/11554046 [accessed 2019-10-18]
79. Hellmueller L, Trilling D. The Credibility of Credibility Measures: A Meta-Analysis in Leading Communication Journals, 1951 to 2011. Handle Net Registry. 2012. URL: https://hdl.handle.net/11245/1.372698 [accessed 2019-06-18]
80. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
81. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010 Jun;33(2-3):61-83; discussion 83. [doi: 10.1017/S0140525X0999152X] [Medline: 20550733]
82. Hanel PH, Vione KC. Do student samples provide an accurate estimate of the general public? PLoS One 2016;11(12):e0168354 [FREE Full text] [doi: 10.1371/journal.pone.0168354] [Medline: 28002494]
83. Hoskin R. The Dangers of Self-Report. Science Brainwaves. 2012. URL: https://www.sciencebrainwaves.com/the-dangers-of-self-report/ [accessed 2019-08-01]

Abbreviations

CDC: Centers for Disease Control and Prevention
ELM: elaboration likelihood model
FDA: Food and Drug Administration
MAIN: modality, agency, interactivity, and navigability
SDT: self-determination theory
SM: social media
SMI: social media influencer

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by M Jordan-Marsh, Y Zhang; comments to author 30.01.20; revised version received 16.03.20; accepted 20.04.20; published 23.07.20.

Please cite as:
Jenkins EL, Ilicic J, Barklamb AM, McCaffrey TA
Assessing the Credibility and Authenticity of Social Media Content for Applications in Health Communication: Scoping Review
J Med Internet Res 2020;22(7):e17296
URL: http://www.jmir.org/2020/7/e17296/
doi:10.2196/17296
PMID:32706675

XSL•FO
RenderX

©Eva L Jenkins, Jasmina Ilicic, Amy M Barklamb, Tracy A McCaffrey. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Influence of Social Media on Sexualized Drug Use and Chemsex Among Chinese Men Who Have Sex With Men: Observational Prospective Cohort Study

Zixin Wang[1,2], PhD; Xue Yang[1,2], PhD; Phoenix K H Mo[1], PhD; Yuan Fang[3], PhD; Tsun Kwan Mary Ip[1], MSc; Joseph T F Lau[1,2], PhD

[1]Centre for Health Behaviours Research, JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

[3]Department of Early Childhood Education, Faculty of Education and Human Development, The Education University of Hong Kong, Hong Kong, Hong Kong

Corresponding Author:
Zixin Wang, PhD
Centre for Health Behaviours Research
JC School of Public Health and Primary Care, Faculty of Medicine
The Chinese University of Hong Kong
Room 508, School of Public Health
Prince of Wales Hospital
Hong Kong, 666888
Hong Kong
Phone: 852 22528740
Email: wangzx@cuhk.edu.hk

Abstract

Background: Sexualized drug use (SDU; the use of any psychoactive substance before or during sexual intercourse) is prevalent among men who have sex with men (MSM) and may aggravate the worsening HIV epidemic in this key population.

Objective: This observational prospective cohort study investigated factors predicting the occurrence of SDU within a 6-month follow-up period among a sample of MSM in Hong Kong. We hypothesized that perceptions related to SDU would mediate the association between the influence of social media/gay social networking apps and SDU during the follow-up period.

Methods: Participants were Chinese-speaking men in Hong Kong, China who had anal intercourse with at least one man in the past year. Among 600 participants who completed the baseline telephone survey, 407 (67.8%) completed another telephone survey 6 months later. Logistic regression models and path analysis were fitted.

Results: At Month 6, 6.9% (28/407) and 4.4% (18/407) of participants reported SDU and chemsex during the follow-up period. After adjustment for significant baseline background variables (use of pre-exposure prophylaxis; history of HIV and other sexually transmitted infections; anal intercourse with nonregular male sex partners, condomless anal intercourse with men, multiple male sex partnerships, and SDU at baseline), three constructs of the Theory of Planned Behavior (TPB) were significantly associated with SDU during the follow-up period: (1) positive attitudes toward SDU (adjusted odds ratio [AOR] 1.19, 95% CI 1.05-1.36), (2) perceived support for SDU from significant others (AOR 1.15, 95% CI 1.01-1.30), and (3) perceived behavioral control of refraining from SDU (AOR 0.76, 95% CI 0.59-0.98). Exposure to information supporting SDU on social media and gay social networking apps was also significantly associated with SDU (AOR 1.11, 95% CI 1.01-1.22). Bootstrapping analyses indicated that social media influence was indirectly associated with SDU through TPB-related perceptions of SDU (β=.04; B=.002, 95% CI 0.001-0.01).

Conclusions: Social media and gay social networking apps may be a major source of influence on MSM's perceptions and actual behaviors related to SDU.

(J Med Internet Res 2020;22(7):e17894) doi:10.2196/17894

XSL•FO
RenderX

KEYWORDS

influence of social media; sexualized drug use; chemsex; men who have sex with men; prospective observational cohort study

Introduction

Psychoactive substance use and HIV are intertwined epidemics among MSM [1]. Sexualized drug use (SDU) refers to the use of any psychoactive substance before or during sexual intercourse [2]. Chemsex is considered a subset of SDU, which is commonly defined as the use of specific drugs (methamphetamine, mephedrone, γ-hydroxybutyrate [GHB/GBL], ketamine and cocaine) before or during sexual intercourse [3-5]. Two published systematic reviews addressed SDU and chemsex among MSM and found these practices prevalent across settings [3,6]. A recent study found that 54% of MSM visiting sexually transmitted infection (STI) clinics in the Netherlands reported SDU [4]. The prevalence of using alcohol or illicit psychoactive substances before or during sexual intercourse was 41% among an online sample of MSM in the United Kingdom [5]. In Hong Kong, China, 12.0% of MSM reported SDU in 2017 [7]. The prevalence of chemsex among MSM ranged from 3% to 35% across countries [3,8-10]. Psychoactive substances such as those used for SDU/chemsex adversely affect users' capacity to perceive and respond to risks during sexual encounters, and may prompt them to engage in high-risk sexual practices [11]. Studies have consistently shown that SDU and chemsex were associated with a higher risk of condomless anal intercourse (CAI), group sex, and fisting, resulting in a higher risk of HIV and other STIs [3,8,12]. SDU/chemsex among MSM in Hong Kong may aggravate the worsening HIV epidemic in this group (6.54% of MSM were HIV-positive in 2017) [13]. However, there is a lack of effective behavioral interventions targeting SDU/chemsex [14-17].

Factors associated with SDU/chemsex among MSM included those related to sociodemographics (eg, age, ethnicity, income, education level, and employment status), sexual orientation, and utilization of HIV testing and other HIV/STI prevention services [3,5,9]. Studies also showed that MSM with SDU/chemsex experience were more likely to access pre-exposure prophylaxis (PrEP) [10,12]. These factors were considered in this study. Perceptions related to SDU also play important roles, and are important for theory-based interventions to address. Qualitative studies found that a key reason MSM engaged in SDU/chemsex was the belief that psychoactive substances could increase stamina and arousal levels, so that an individual could engage in sex for a sustained period [3]. Other reasons given included the following: overcoming underconfidence, enhancing the emotional connection with sex partners, and coping with stress [3]. Although a qualitative study suggested that MSM are under peer pressure to have SDU and chemsex [18], other social network–related factors (such as fear of loss of partners or friends) deterred MSM from using psychoactive substances before or during sexual intercourse [3]. Only one quantitative study showed that a perceived lower confidence of practicing safe sex consistently was associated with a higher likelihood of chemsex among MSM in the United Kingdom [5]. In this study, we chose the Theory of Planned Behavior (TPB) as the framework to guide our examination of perceptions related to SDU [19]. The TPB has commonly been applied to explain various types of risk behaviors [20] and design various health-related interventions [21]. The TPB specifies that attitudes, subjective norms (ie, whether others approve of certain behaviors), and perceived behavioral control (perceived availability of resources and opportunities that enable a person to perform behaviors) determine health behaviors [19]. This framework is potentially applicable for understanding the factors related to SDU/chemsex among MSM, given findings from previous studies that noted the importance of attitudes, norms, and perceived control on this behavior.

Constructs from the TPB and subsequent behaviors may be shaped by social network–related factors as well as the use of social media and gay social networking apps. In Hong Kong, 96% of residents have access to a smartphone [22]; 85% and 57% use Facebook and Instagram, respectively [23]; and about 70% of local MSM use gay social networking apps to seek sex partners [24]. Social media (eg, Facebook, Twitter, Instagram, Weibo) and location-aware gay social networking apps (eg, Grindr, Jack'd, Blued) have facilitated drug purchasing and seeking SDU and chemsex in one's geographical location [18,25,26]. Across countries, MSM reported that it was common to encounter information related to SDU/chemsex when browsing these social media and apps [18,25,26]. Many MSM openly solicit chemsex in their online profiles [18,25,26], while others reported being asked to use psychoactive substances or engage in SDU/chemsex by other MSM through these apps [18,25-28]. In addition, posts on social media such as Facebook or Twitter usually portray psychoactive substance use as normative, which may increase intentions to practice SDU or chemsex [29,30]. Despite the potential for social media and apps to increase SDU and chemsex among MSM in Hong Kong, no quantitative studies have investigated the influence of these factors.

In addition to social media and gay social networking apps, peer behaviors may also influence SDU and chemsex practices among MSM. Compared to other HIV key populations, MSM tend to have large and dense social networks [31]. Most Chinese MSM are heavily involved in the MSM community, and have strong ties to each other [32]. As MSM are a highly marginalized community in Hong Kong, individuals within it may perceive a high level of similarity to their MSM peers and may consequently find their peers' experiences particularly valuable and influential. Prior studies have found that peers within an individual's social network are the most common source for obtaining psychoactive substances [33,34]. Qualitative studies have suggested that MSM often introduce SDU and chemsex to their peers. Previous studies suggested that a higher number of psychoactive substance users in one's social network was positively associated with substance use in various populations [35-37], including MSM [36].

The Social Learning Theory posits that people learn by both direct experience and observation [38]. Observation of peers is a major source of influence on people's health attitudes and

XSL•FO
RenderX

behaviors [39]. Given the close relationships and high level of perceived similarity among MSM, observing peer behaviors plays an important role in influencing MSM's attitudes and behaviors related to SDU/chemsex. Such interactions and observations may occur online and offline. In this study, we hypothesized that being exposed to information supporting SDU on social media/gay apps and observing SDU among peers would influence MSM's perceptions related to SDU, which would influence the occurrence of SDU.

To address these gaps, this study investigated predictors of SDU within a 6-month follow-up period among a sample of MSM in Hong Kong. Potential predictors measured at baseline included sociodemographics and behavioral characteristics of the participants, perceptions related to SDU based on the TPB, and the influence of social media/gay social networking apps and peers. We further tested the hypothesis that perceptions of SDU would mediate the associations between the influence of social media/gay social networking apps and peers on SDU.

Methods

Participants and Data Collection

We conducted an observational prospective cohort study among MSM in Hong Kong, China from April 2018 to July 2019. Inclusion criteria for participants of the cohort study were the following: (1) Chinese-speaking men in Hong Kong, (2) aged 18 years or above, and (3) had anal intercourse with at least one man in the last 12 months. Participants were recruited through multiple sources. A recent mapping exercise was conducted by the government, which identified 12 gay bars and 16 gay saunas in Hong Kong. Upon obtaining the approval of the owners, trained and experienced fieldworkers approached prospective MSM participants in these venues at different time slots during weekdays and weekends. They briefed prospective participants about the study details and gave them an information sheet. The research team also conducted online outreach by posting information about the study periodically as discussion topics on the two gay websites with the highest traffic in Hong Kong. If prospective participants were interested in this study, they could contact the interviewers through private messaging or other means (eg, WhatsApp, telephone, email). Recruitment was supplemented by peer referrals. Participants were guaranteed anonymity during the study, and had the right to end participation in the study at any time. Their refusal or withdrawal from the study would not affect their access to any future services. Verbal consent was obtained instead of written consent to allow participants to maintain anonymity, and the fieldworkers signed a form pledging that the participants had been fully informed about the study. Multiple forms of contact information were obtained to make an appointment to conduct a baseline telephone interview. Trained telephone interviewers confirmed participants' eligibility and consent to participate in the study, and conducted telephone interviews that took approximately 30 minutes to complete. Another telephone survey was conducted 6 months later. At least 5 follow-up calls were made during different time slots during weekdays and weekends before considering the participant lost to follow-up. Upon completion of both surveys, a HK $50 (US $6.45) supermarket or café coupon was mailed to participants as compensation for their time. Telephone numbers and addresses were cross-checked to avoid repetition. Of the 1131 prospective participants approached through outreach in gay venues (n=211), online recruitment (n=607), and peer referral (n=313), 711 were eligible for participation after screening (venues: n=117; online: n=401; referral: n=193). Of these, 600 provided verbal informed consent and completed the baseline telephone interview (venues: n=85; online: n=345; referral: n=170). The response rate was 84.3%. At Month 6, 407 (67.8%) participants completed the follow-up survey. The Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong approved this study (reference number 160051).

Measures

Design of the Questionnaires

A panel consisting of a public health researcher, an epidemiologist, a psychologist, an MSM community member, and a community-based organization worker was formed to develop the questionnaires. The questionnaires were tested by 10 local MSM. Based on their feedback, the questionnaires were then finalized by the panel.

Baseline Background Characteristics

Information collected included sociodemographics (age, highest education level attained, current marital status, monthly personal income, and current employment status), sexual orientation; utilization of HIV prevention services and PrEP; history of HIV and other STIs; and sexual behaviors with regular and nonregular male sex partners in the last year. A regular male sex partner (RP) was defined as a stable boyfriend, while a nonregular male sex partner (NRP) was defined as a casual sex partner or a male sex worker.

Patterns of SDU/Chemsex

In this study, SDU is defined as the use of any of the following psychoactive substances before or during sexual intercourse: ketamine, methamphetamine, cocaine, cannabis, ecstasy, Dormicum/Halcion/Erimin 5/nonprescription hypnotic drugs, heroin, cough suppressant (not for curing cough), amyl nitrite (popper), GHB/GBL, 5-methoxy-N, N-diisopropyltryptamine (Foxy), and mephedrone. We defined chemsex as the use of ketamine, methamphetamine, cocaine, GHB/GBL, or mephedrone before or during sexual intercourse. Such definitions have been used in a number of studies [3-5]. Participants were asked whether they had used any of the aforementioned psychoactive substances at baseline and Month 6 during different reference periods (baseline: lifetime and past year; Month 6: in the past 6 months).

Participants with SDU experience were asked about behavioral details, including the following: (1) types of psychoactive substance used during sexual intercourse, (2) poly-use of psychoactive substances during sexual intercourse, (3) time since the first episode of SDU, (4) frequency of SDU, (5) CAI during SDU, and (6) sexual risk behaviors during their most recent episode of SDU (ie, the number of participants; the use of alcohol and erectile dysfunction drugs; and the occurrence of CAI or group sex).

XSL•FO
RenderX

Influence of Social Media/Gay Social Networking Apps and Peers Measured at Baseline

Participants were asked to report the frequency of their exposure to the following information related to SDU on social media (eg, Facebook, Twitter, Instagram, Weibo) or gay social networking apps (eg, Grindr, Jack'd, Blued) in the last 12 months (eg, "Frequency of exposure to others' personal experiences supporting psychoactive substance use") with the following response categories: 1=Never, 2=Seldom, 3=Sometimes, and 4=Always. The Influence of Social Media Scale was formed by summing up individual item scores, with higher scores indicating higher levels of exposure to information supporting SDU from social media and gay social networking apps. The Cronbach α for this scale was .70, and a single factor was identified by exploratory factor analysis, explaining 70.1% of total variance. Additionally, the questionnaire also measured the number of peers in one's social network who had experience with SDU (response categories: None, 1-2, 3-5, 6-10, and >10).

Perceptions Related to SDU Based on the TPB

Five scales derived from the TPB were constructed [40]. Positive attitudes toward SDU were measured by four items (eg, "It is easier for you to find sex partners when using psychoactive substances before/during sexual intercourse"). The Positive Attitude Scale was formed by summing up individual item scores (from 1=Strongly disagree to 5=Strongly agree). Higher scores on the scale indicated more positive attitudes toward SDU. The Cronbach α of the Positive Attitude Scale was .82; one factor was identified by exploratory factor analysis, explaining 65.1% of total variance.

Four items (with response options ranging from 1=Strongly disagree to 5=Strongly agree) were used to measure negative attitudes toward SDU (eg, "Using psychoactive substances before/during sexual intercourse would increase your risk of having CAI"). The Negative Attitude Scale was formed by summing up individual item scores. Higher scores on the scale indicated more negative attitudes toward SDU. The Cronbach α of the Negative Attitude Scale was .62; one factor was identified by exploratory factor analysis, explaining 63.0% of total variance.

Four items were used to measure participants' perceived support from their significant others (referring to regular and nonregular male sex partners and MSM friends) for SDU. Items were measured on a 5-point Likert Scale from 1=Strongly disagree to 5=Strongly agree. The Perceived Subjective Norms for SDU among MSM Scale was constructed by summing up individual item scores. Higher scores indicated perceived subjective norms that were more supportive of SDU. The Cronbach α of the Perceived Subjective Norm Scale was .90; one factor was identified by exploratory factor analysis, explaining 76.8% of total variance.

Perceived behavioral control in refusing SDU and perceived behavioral control in refraining from SDU were measured by two items each. The Perceived Behavioral Control of Refusing SDU Scale and the Perceived Behavioral Control of Refraining from SDU Scale were formed by summing up individual item scores (from 1=Strongly disagree to 5=Strongly agree). Higher scores of the Perceived Behavioral Control of Refusing SDU Scale indicated lower levels of control of refusing SDU, while higher scores in the Perceived Behavioral Control of Refraining from SDU indicated higher levels of control of refraining from SDU. The Cronbach α of the Perceived Behavioral Control of Refusing SDU Scale and the Perceived Behavioral Control of Refraining from SDU Scale were .93 and .95, respectively; single factors were identified by exploratory factor analysis, explaining 93.5% and 95.6% of total variance, respectively.

Data Analyses

Baseline background characteristics of participants who were followed up at Month 6 and those who were lost to follow-up were compared using chi-square tests (for categorical variables) or independent sample *t* tests (for continuous variables). The subsequent analysis was performed among those who had completed both surveys. Using SDU during the follow-up period as the dependent variable, and background characteristics measured at baseline as independent variables, univariate odds ratios (OR) predicting the dependent variable were obtained using logistic regression models. After adjusting for variables with $P<.05$ in the univariate analysis, associations between independent variables of interest (ie, perceptions, the influence of social media and gay social networking apps and peers) and the dependent variable were then assessed by adjusted OR (AOR). Each AOR was obtained by fitting a single logistic regression model, which involved one of the independent variables of interest and the significant background variables.

Path analysis was conducted to test the mediation model. The means for social media influence and peer influence were used as independent variables, while SDU during the follow-up period was used as the dependent variable. The means of the Positive Attitude Scale, the Negative Attitude Scale, the Perceived Subjective Norm Scale, the Perceived Behavioral Control for Refusing SDU Scale, and the Perceived Behavioral Control for Refraining from SDU Scale were used as indicators to represent the latent variable of perceptions related to SDU based on the TPB; confirmatory factor analysis was then conducted to test the goodness of fit of this construct. All significant background variables were controlled for in the model. Goodness of fit was tested by using the chi-square test, the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI), and the root mean square error of approximation (RMSEA). A CFI and NNFI greater than .90 and RMSEA lower than .08 indicated acceptable goodness of fit [41]. Standardized path coefficients (β) and unstandardized path coefficients (B) were reported. We tested the mediation analyses using bootstrapping. The 95% CI of the indirect effects were obtained from 5000 bootstrap samples. A statistically significant mediation effect was observed when the CI did not include zero. The level of statistical significance was .05. SPSS (Version 18.0, IBM Corp) and AMOS were used.

Results

Baseline Background Characteristics

Most of the 600 participants were 18 to 30 years old (n=342, 57.0%), currently single (n=504, 84.0%), employed full-time (n=498, 83.0%), with a monthly personal income of HK $20,000 (US $2580) or higher (n=336, 56.0%), and had attained at least

XSL•FO
RenderX

a college education (n=505, 84.2%). Over half had received HIV antibody testing (n=430, 71.7%) and utilized other HIV prevention services (n=335, 55.8%) in the last year; 3.8% (n=23) were using PrEP at the time of the baseline survey. Among the participants, 3.3% (n=20) self-reported as living with HIV and 20.5% (n=123) had a history of other STIs. In the last year, 85.0% (n=510) and 60.5% (n=363) had had anal intercourse with RP and NRP, respectively. Furthermore, 40.2% (n=241) and 69.0% (n=414) reported CAI with men and multiple sex partnerships, respectively. The lifetime prevalence of SDU and chemsex were 19.3% (n=116) and 8.7% (n=52), respectively.

Means and standard deviation of items/scales related to the influence of social media/gay apps, the influence of peers, and perceptions related to SDU based on the TPB were described in Table 1. No significant difference was found in baseline characteristics between those with follow-up data available (n=407) and those who were lost to follow-up at Month 6 (n=193), with the exception of self-reported HIV status (*P*=.04), the frequency of exposure to commentary/discussion about psychoactive substance use on social media/gay apps (*P*=.001), and the score on the Perceived Behavioral Control of Refraining from SDU Scale (*P*=.04).

XSL•FO
RenderX

Table 1. Baseline characteristics of the participants.

Characteristics	All participants (n=600)	Follow-up at Month 6 (n=407)	Loss to follow-up (n=193)	*P* value
Sociodemographics				
Age group (years), n (%)				**.29**
18-24	124 (20.7)	80 (19.7)	44 (22.8)	
25-30	218 (36.3)	146 (35.9)	72 (37.3)	
31-40	187 (31.2)	126 (31.0)	61 (31.6)	
>40	71 (11.8)	55 (13.5)	16 (8.3)	
Highest educational level attained, n (%)				**.73**
Senior high school or below	95 (15.8)	63 (15.5)	32 (16.6)	
College or above	505 (84.2)	344 (84.5)	161 (83.4)	
Current marital status, n (%)				**.38**
Currently single	504 (84.0)	344 (84.5)	160 (82.9)	
Married/cohabiting with a man	93 (15.5)	60 (14.7)	33 (17.1)	
Married/cohabiting with a woman	3 (0.5)	3 (0.7)	0 (0.0)	
Monthly personal income (HK $), n (%)				**>.99**
<10,000 (<US $1290)	84 (14.0)	57 (14.0)	27 (14.0)	
10,000-19,999 (US $1290-2580)	174 (29.0)	120 (29.5)	54 (28.0)	
20,000-39,999 (US $2580-3870)	220 (36.7)	148 (36.4)	72 (37.3)	
>40,000 (>US $5161)	116 (19.3)	78 (19.2)	38 (19.7)	
Refuse to disclose	6 (1.0)	4 (1.0)	2 (1.0)	
Current employment status, n (%)				**.61**
Full-time	498 (83.0)	340 (83.5)	158 (81.9)	
Part-time/unemployed/retired/students	102 (17.0)	67 (16.5)	35 (18.1)	
Sexual orientation, n (%)				**.35**
Homosexual	546 (91.0)	375 (92.1)	171 (88.6)	
Bisexual	52 (8.7)	31 (7.6)	21 (10.9)	
Heterosexual	2 (0.3)	1 (0.3)	1 (0.5)	
Service utilization, n (%)				
HIV testing in the last 12 months				**.15**
No	170 (28.3)	108 (26.5)	62 (32.1)	
Yes	430 (71.7)	299 (73.5)	131 (67.9)	
Other HIV prevention services in the last 12 months (eg, condom distribution, peer education, pamphlet and lectures)				**.46**
No	265 (44.2)	184 (45.2)	81 (42.0)	
Yes	335 (55.8)	223 (54.8)	112 (58.0)	
Currently on PrEP (pre-exposure prophylaxis)				**.78**
No	577 (96.2)	392 (96.3)	185 (95.9)	
Yes	23 (3.8)	15 (3.7)	8 (4.1)	
History of HIV/sexually transmitted infections, n (%)				
Self-reported HIV status				**.04**
Never tested for HIV	26 (4.3)	11 (2.7)	15 (7.8)	
Negative	549 (91.5)	380 (93.4)	169 (87.6)	
Positive	20 (3.3)	13 (3.2)	7 (3.6)	

XSL•FO
RenderX

Characteristics	All participants (n=600)	Follow-up at Month 6 (n=407)	Loss to follow-up (n=193)	*P* value
Refuse to disclose	5 (0.8)	3 (0.7)	2 (1.0)	
History of other sexually transmitted infections				**.76**
No	477 (79.5)	325 (79.9)	152 (78.8)	
Yes	123 (20.5)	82 (20.1)	41 (21.2)	
Sexual behaviors in the last 12 months, n (%)				
Has had anal intercourse with regular male sex partners				**.14**
No	90 (15.0)	55 (13.5)	35 (18.1)	
Yes	510 (85.0)	352 (86.5)	158 (81.9)	
Has had anal intercourse with nonregular male sex partners				**.69**
No	237 (39.5)	163 (40.0)	74 (38.3)	
Yes	363 (60.5)	244 (60.0)	119 (61.7)	
Condomless anal intercourse with men				**.33**
No	359 (59.8)	249 (61.2)	110 (57.0)	
Yes	241 (40.2)	158 (38.8)	83 (43.0)	
Multiple male sex partnerships				**.88**
No	186 (31.0)	127 (31.2)	59 (30.6)	
Yes	414 (69.0)	280 (68.8)	134 (69.4)	
Experience of sexualized drug use (SDU)[a], n (%)				
SDU in lifetime				**.61**
No	484 (80.7)	326 (80.1)	158 (81.9)	
Yes	116 (19.3)	81 (19.9)	35 (18.1)	
SDU in the past year				**.68**
No	512 (85.3)	349 (85.7)	163 (84.5)	
Yes	88 (14.7)	58 (14.3)	30 (15.5)	
Influence of social media related to SDU				
Frequency of exposure to information supporting SDU on social media/gay social networking apps in the past year, % Sometimes/Always				
Sharing of personal experiences to support MSM using psychoactive substances	253 (42.2)	168 (41.3)	85 (44.0)	.52
Sharing of personal experiences against MSM using psychoactive substances (reverse coded)	180 (30.0)	131 (32.2)	49 (25.4)	.09
Receiving personal invitations to use psychoactive substances from MSM friends	93 (15.5)	58 (14.3)	35 (18.1)	.22
Receiving personal invitations to have SDU/chemsex from MSM friends	90 (15.0)	55 (13.5)	35 (18.1)	.14
Receiving personal invitations to use psychoactive substances from strangers	212 (35.3)	145 (35.6)	67 (34.7)	.83
Receiving personal invitations to have SDU/chemsex from strangers	218 (36.3)	150 (36.9)	68 (35.2)	.70
Commentary/discussion about psychoactive substance use	168 (28.0)	97 (23.8)	71 (36.8)	.001
The Influence of Social Media Scale, mean (SD)	7.7 (4.4)	7.5 (4.3)	8.1 (4.4)	.13
Influence of peers related to SDU				
Regular male sex partners had experience of SDU/chemsex, % Yes	48 (8.0)	39 (9.6)	9 (4.7)	.04

XSL•FO
RenderX

Characteristics	All participants (n=600)	Follow-up at Month 6 (n=407)	Loss to follow-up (n=193)	*P* value
Nonregular male sex partners had experience of SDU/chemsex, % Yes	109 (18.2)	72 (17.7)	37 (19.2)	.66
Close friends had experience of SDU/chemsex, % Yes	129 (21.5)	94 (23.1)	35 (18.1)	.17
Other friends had experience of SDU/chemsex, % Yes	229 (38.2)	151 (37.1)	78 (40.4)	.44
Number of peers in one's social network who had ever engaged in SDU/chemsex, n (%)				**.62**
0	178 (29.7)	121 (29.7)	57 (29.5)	
1-2	148 (24.7)	104 (25.6)	44 (22.8)	
3-5	143 (23.8)	92 (22.6)	51 (26.4)	
6-10	48 (8.0)	36 (8.8)	16 (8.2)	
>10	83 (13.8)	54 (13.3)	29 (15.0)	
Perceptions related to SDU based on the Theory of Planned Behavior				
Positive attitudes toward SDU, % Agree/Strongly agree				
SDU allows you temporary escape from reality	106 (17.7)	70 (17.2)	36 (18.7)	.66
SDU increases your sexual pleasure	139 (23.2)	99 (24.3)	40 (20.7)	.33
It is easier to find sex partners during SDU	74 (12.3)	47 (11.5)	27 (14.0)	.40
SDU would heighten euphoria	131 (21.8)	93 (22.9)	38 (19.7)	.38
Positive Attitude Scale, mean (SD)	8.0 (4.2)	8.8 (4.1)	8.8 (4.3)	.93
Negative attitudes toward SDU, % Agree/Strongly agree				
SDU would harm your cognitive function	558 (94.0)	379 (93.1)	185 (95.9)	.19
SDU would have negative impact on your relationship with sex partners	384 (64.0)	265 (65.1)	119 (61.7)	.41
SDU would increase your risk of having condomless anal intercourse	381 (63.5)	258 (63.4)	123 (63.7)	.94
SDU would increase your risk of HIV infection	511 (85.2)	341 (83.8)	170 (88.1)	.17
Negative Attitude Scale, mean (SD)	16.8 (2.9)	16.7 (2.9)	16.8 (2.8)	.72
Perceived subjective norms related to SDU, % Agree/Strongly agree				
Your male sex partners support you using psychoactive substances	29 (4.8)	19 (4.7)	10 (5.2)	.78
Your other MSM friends support you using psychoactive substances	17 (2.8)	12 (2.9)	5 (2.6)	.81
Your male sex partners support you having SDU	21 (3.5)	14 (3.4)	7 (3.6)	.91
Your other MSM friends support you having SDU	11 (1.8)	8 (2.0)	3 (1.6)	.73
Perceived Subjective Norms Scale, mean (SD)	5.9 (3.1)	5.8 (3.1)	6.0 (3.0)	.40
Perceived behavioral control of refusing SDU, % Agree/Strongly agree				
If your sex partner asks you to use psychoactive substances, it is difficult for you to refuse	60 (10.0)	36 (8.8)	24 (12.4)	.17
If your sex partner asks you to have SDU, it is difficult for you to refuse	55 (9.2)	31 (7.6)	24 (12.4)	.06
Perceived Behavioral Control of Refusing SDU Scale, mean (SD)	3.5 (2.2)	3.4 (2.1)	3.7 (2.3)	.10
Perceived behavioral control of refraining from SDU, % Agree/Strongly agree				
You can exercise self-control to stop using psychoactive substance	541 (90.2)	371 (91.2)	170 (88.1)	.24
You can exercise self-control to stop having SDU	550 (91.7)	378 (92.9)	172 (89.1)	.12

XSL•FO
RenderX

Characteristics	All participants (n=600)	Follow-up at Month 6 (n=407)	Loss to follow-up (n=193)	*P* value
Perceived Behavioral Control of Refraining from SDU Scale, mean (SD)	9.2 (1.7)	9.3 (1.5)	9.0 (2.0)	.04

[a]SDU: sexualized drug use. Sexualized drug use is defined as the use of any of the following psychoactive substances before/during anal intercourse: ketamine, methamphetamine, cocaine, cannabis, ecstasy, Dormicum/Halcion/Erimin 5/nonprescription hypnotic drugs, heroin, cough suppressant (not for curing cough), amyl nitrite (popper), GHB/GBL (γ-hydroxybutyrate), 5-methoxy-N, N-diisopropyltryptamine (Foxy), and mephedrone.

Patterns of SDU and Chemsex at Baseline and Month 6

At baseline, 14.7% (88/600) and 6.7% (40/600) of the participants reported SDU and chemsex in the past year, respectively. During the 6-month follow-up period, their prevalence was 6.9% (28/407) and 4.4% (18/407), respectively. Among 58 participants with experience of SDU who completed the Month 6 follow-up survey, 36% (n=21) reported SDU during the follow-up period. Among 26 MSM who had chemsex at baseline and completed the Month 6 follow-up survey, 14 (54%) reported chemsex during the follow-up period.

Patterns of SDU one year prior to the baseline and during the 6-month follow-up period were similar. Amyl nitrite, methamphetamine, and GHB/GBL were the most commonly used psychoactive substances during sexual intercourse. About half of MSM who reported SDU reported poly-use of psychoactive substances and CAI. SDU commonly involved more than two people, the use of erectile dysfunction drugs, and group sex (Table 2).

XSL•FO
RenderX

Table 2. Patterns of sexualized drug use in different reference periods.

Variables	One year prior to baseline survey (among MSM[a] who reported sexualized drug use in the past year at baseline, n=88), n (%)	During the 6-month follow-up period (among MSM who reported sexualized drug use during the follow-up, n=28), n (%)
Types of psychoactive substance used during chemsex		
Ketamine	2 (2.3)	2 (7.1)
Methamphetamine	32 (36.4)	11 (39.3)
Cocaine	0 (0.0)	0 (0.0)
Cannabis	7 (8.0)	1 (3.6)
Ecstasy	2 (2.3)	2 (7.1)
Dormicum/Halcion/Erimin 5/Hypnotic drugs (nonprescription)	0 (0.0)	1 (3.6)
Heroin	0 (0.0)	0 (0.0)
Cough suppressant (not for curing cough)	1 (1.1)	2 (7.1)
Amyl nitrite	69 (78.4)	20 (71.4)
γ-hydroxybutyrate (GHB/GBL)	29 (33.0)	10 (35.7)
5-methoxy-N, N-diisopropyltryptamine	5 (5.7)	1 (3.6)
Mephedrone	0 (0.0)	1 (3.6)
Poly-use of psychoactive substances		
No	57 (64.8)	14 (50.0)
Yes	31 (35.2)	14 (50.0)
Frequency of sexualized drug use		
<1 episode/month	0 (0.0)	13 (46.4)
1 episode/month	35 (39.8)	11 (39.3)
1-2 episodes/month	24 (27.3)	2 (7.1)
≥3 episodes/month	29 (33.0)	2 (7.1)
Condomless anal intercourse during sexualized drug use		
No	41 (46.6)	10 (35.7)
Yes	47 (53.4)	18 (64.3)
Characteristics of most recent episode of sexualized drug use		
Number of participants		
2	67 (76.1)	19 (67.9)
≥3	21 (23.9)	9 (32.1)
Alcohol consumption		
No	77 (87.5)	28 (100.0)
Yes	11 (12.5)	0 (0.0)
Use of erectile dysfunction drugs		
No	62 (70.5)	16 (57.9)
Yes	26 (29.5)	12 (42.1)
Group sex		
No	18 (20.5)	11 (39.3)
Yes	70 (79.5)	17 (60.7)
Condomless anal intercourse		
No	45 (51.1)	15 (53.6)
Yes	43 (48.9)	13 (46.4)

XSL•FO
RenderX

[a]MSM: men who have sex with men.

Baseline Factors Predicting SDU During the Follow-Up Period

Baseline background variables that were significantly associated with SDU during the follow-up period included the following: (1) currently on PrEP, (2) self-reported living with HIV, (3) history of other STI, (4) any anal intercourse with NRP, (5) CAI with any male sexual partners, (6) multiple male sex partnerships, and (7) SDU in the past year (Table 3).

After adjusting for significant variables, three constructs of the TPB measured at baseline were significantly associated with SDU during the follow-up period. These constructs included: (1) positive attitudes toward SDU (AOR 1.19, 95% CI 1.05-1.36), (2) perceived support for SDU from significant others (AOR 1.15, 95% CI 1.01-1.30), and (3) perceived behavioral control of refraining from SDU (AOR 0.76, 95% CI 0.59-0.98). A higher level of exposure to information supporting SDU on social media/gay apps as measured at baseline was associated with a higher likelihood of SDU during the follow-up period (AOR 1.11, 95% CI 1.01-1.22). The association between the number of peers in one's social network who had ever engaged in SDU and the dependent variable was of marginal statistical significance (AOR 1.43, 95% CI 0.95-2.16, *P*=.06; Table 4.

XSL•FO
RenderX

Table 3. Baseline background variables associated with sexualized drug use during the follow-up period for MSM in Hong Kong (among those being followed up at Month 6, n=407).

Variables	Participants, n (%)	OR[a] (95% CI)
Sociodemographics measured at baseline		
Age group (years)		
18-24	4 (5.0)	1.0
25-30	10 (6.8)	1.40 (0.42-4.61)
31-40	11 (8.7)	1.82 (0.56-5.92)
>40	3 (5.5)	1.10 (0.24-5.10)
Highest educational level attained		
Senior high school or below	7 (11.1)	1.0
College or above	21 (6.1)	0.52 (0.21-1.28)
Current marital status		
Currently single	24 (7.0)	1.0
Married/cohabiting with a man	3 (5.0)	0.70 (0.21-2.41)
Married/cohabiting with a woman	1 (33.3)	6.67 (0.58-76.18)
Monthly personal income (HK $)		
<10,000 (<US $1290)	5 (8.8)	1.0
10,000-19,999 (US $1290-2580)	6 (5.0)	0.55 (0.16-1.88)
20,000-39,999 (US $2580-3870)	12 (8.1)	0.92 (0.31-2.73)
>40,000 (>US $5161)	5 (6.4)	0.71 (0.20-2.59)
Refuse to disclose	0 (0.0)	N/A[b]
Current employment status		
Full-time	21 (6.2)	1.0
Part-time/unemployed/retired/student	7 (10.4)	1.77 (0.72-4.35)
Sexual orientation		
Homosexual	27 (7.2)	1.0
Bisexual	1 (3.2)	0.43 (0.05-3.27)
Heterosexual	0 (0.0)	N/A
Service utilization measured at baseline		
HIV testing in the last 12 months		
No	4 (3.7)	1.0
Yes	24 (8.0)	2.27 (0.77-6.70)
Other HIV prevention services in the last 12 months (eg, condom distribution, peer education, pamphlets, lectures)		
No	12 (6.5)	1.0
Yes	16 (7.2)	1.11 (0.51-2.41)
Currently on PrEP (pre-exposure prophylaxis)		
No	23 (5.9)	1.0
Yes	5 (33.3)	8.02 (2.53-25.42)[c]
History of HIV/sexual transmitted infections		
Self-reported HIV status		
Negative/never tested for HIV/refuse to disclose	24 (6.1)	1.0
Positive	4 (30.8)	6.85 (1.97-23.87)[d]

XSL•FO
RenderX

Variables	Participants, n (%)	OR[a] (95% CI)
History of other sexually transmitted infections		
No	16 (4.9)	1.0
Yes	12 (14.6)	3.31 (1.50-7.31)[d]
Sexual behaviors in the last 12 months as measured at baseline		
Anal intercourse with regular male sex partners		
No	2 (3.6)	1.0
Yes	26 (7.4)	2.11 (.49-9.17)
Anal intercourse with nonregular male sex partners		
No	1 (.6)	1.0
Yes	27 (11.1)	20.16 (2.71-149.88)[d]
Condomless anal intercourse with men		
No	8 (3.2)	1.0
Yes	20 (12.7)	4.37 (1.87-10.18)[d]
Multiple male sex partnerships		
No	0 (0)	N/A
Yes	28 (10.0)	N/A[e]
Experience of sexualized drug use as measured at baseline		
Sexualized drug use in the past year		
No	7 (2.0)	1.0
Yes	21 (36.2)	27.73 (11.05-69.60)[c]

[a]OR: univariate odds ratio.
[b]N/A: not applicable.
[c]*P*<.001.
[d]*P*<.01.
[e]*P*<.05.

XSL•FO
RenderX

Table 4. Factors associated with sexualized drug use during the follow-up period (among those being followed up at Month 6, n=407)[a].

Factors	OR[b] (95% CI)	AOR[c] (95% CI)
Influence of social media/gay apps related to sexualized drug use		
Influence of Social Media Scale	1.23 (1.13, 1.34)[d]	1.11 (1.01, 1.22)[e]
Influence of peers related to sexualized drug use		
Number of peers in one's social network who had ever engaged in SDU/chemsex	2.29 (1.67, 3.14)[d]	1.43 (0.95, 2.16)[f]
Perceptions related to sexualized drug use based on the Theory of Planned Behavior		
Positive Attitude Scale	1.33 (1.19, 1.47)[d]	1.19 (1.05, 1.36)[g]
Negative Attitude Scale	0.86 (0.76, 0.97)[e]	0.99 (0.84, 1.16)
Perceived Subjective Norm Scale	1.29 (1.18, 1.42)[d]	1.15 (1.01, 1.30)[e]
Perceived Behavioral Control of Refusing Sexualized Drug Use Scale	1.33 (1.15, 1.55)[d]	1.08 (0.89, 1.33)
Perceived Behavioral Control of Refraining from Sexualized Drug Use Scale	0.76 (0.64, 0.91)[g]	0.76 (0.59, 0.98)[e]

[a]6.9% of participants reported sexualized drug use during the follow-up period.

[b]OR: univariate odds ratio. The OR represents the increase in the odds of sexualized drug use caused by a one-unit increase in the item/scale score.

[c]AOR: adjusted odds ratio. The AOR is adjusted for the significant background variables listed in Table 3 and multiple sex partnerships in the past year.

[d]$P<.001$.

[e]$P<.05$.

[f]$.05<P<.10$.

[g]$P<.01$.

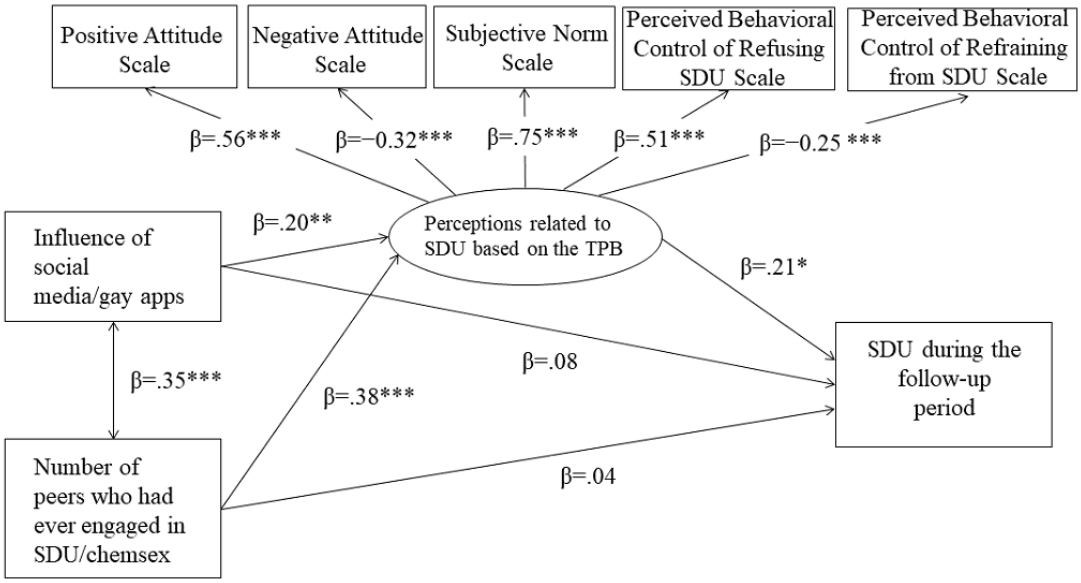
Testing the Mediation Effects of TPB Constructs in the Association Between Social Media/Peers Influences and SDU

Model Testing

The results of the confirmatory factor analysis found that perceptions related to SDU based on the TPB did not show adequate model fit to the data ($\chi^2_5=19.01$, $P<.01$; CFI=.94; NNFI=.89; RMSEA=.08). Modification indices suggested adding a covariance between the Perceived Behavioral Control of Refusing SDU Scale's error and the Perceived Behavioral Control of Refraining from SDU Scale's error (B=–0.58, $P<.001$). Since the two scales are related constructs [19], we added the covariance accordingly. The modified model showed excellent fit ($\chi^2_4=3.81$, $P=.43$; CFI=.99; NNFI=.99; RMSEA=.01). The chi-square test showed that the model fit change was significant (χ^2 change=15.20, degrees of freedom change=1, $P<.05$). All factor loadings were significant at $P<.001$, with absolute values of standardized coefficients greater than .25. The mediation model fit the data well ($\chi^2_{147}=188.71$, $P<.05$; CFI=.92; NNFI=.91; RMSEA=.06; Figure 1).

XSL•FO
RenderX

Figure 1. The proposed mediation models with standardized path coefficients. *P<.05; **P<.01; ***P<.001. SDU: sexualized drug use; TPB: Theory of Planned Behavior.



Path Coefficients

The influence of social media/gay apps was positively associated with perceptions related to SDU based on the TPB (β=.20, B=.11, *P*=.001), while it was not significantly associated with SDU (β=.08, B=.004, *P*=.11). Peer influence was positively associated with perceptions related to SDU based on the TPB (β=.38, B=.68, *P*<.001), while it was not significantly associated with SDU (β=.04, B=.01, *P*=.50). Perceptions related to SDU based on the TPB was positively associated with SDU (β=.21, B=.02, *P*=.02).

Mediation Effects

Bootstrapping analyses indicated that social media influence had a significant indirect effect on SDU through perceptions of SDU in the hypothesized direction (β=.04, B=.002, 95% CI 0.001-0.01). However, peer influence was not significantly associated with SDU through perceptions of SDU (β=.08, B=.01, 95% CI –0.002 to 0.04).

Discussion

We found that 6.9% and 4.4% of MSM in Hong Kong practiced SDU and chemsex at the end of the 6-month follow-up period. Such prevalence was lower than that reported among MSM in Western Europe. At Month 6, about half of participants having SDU or chemsex at baseline refrained from these behaviors, while very few initiated these behaviors. Studies with longer follow-up periods are needed in the future to understand trends in SDU and chemsex among MSM in Hong Kong. Such studies are useful to understand the effectiveness of SDU/chemsex interventions conducted by local community-based organizations.

Similar to studies conducted in Western countries, amyl nitrite, methamphetamine, and GHB/GBL were the most commonly used psychoactive substances during SDU at both baseline and Month 6 [7]. The prevalence of risky practices during SDU (ie, poly-use of psychoactive substance, CAI, and group sex) was high at both time points. Effective risk reduction strategies should be explored to address these risky practices.

According to social marketing approaches, careful segmentation improves the effectiveness of health promotion programs [42]. First, heath care workers and interventions should tailor SDU and chemsex reduction measures to MSM living with HIV. Similar to findings conducted in mainland China [43] and other countries [3], MSM living with HIV were more likely to practice SDU compared to HIV-negative or status unknown MSM. They have a high risk of HIV transmission through SDU. For MSM living with HIV, SDU may be a potential maladaptive coping method for stressful situations caused by perceived stigma and experiences of violence and prejudice [44]. Reducing substance use during sex should be an essential part of care for MSM living with HIV. Second, health care workers and interventionists should focus on MSM with multiple risk behaviors (eg, anal intercourse with NRP, CAI, and multiple male sex partnerships), as these participants were more likely to practice SDU. Those with experience of SDU at baseline also had a much higher risk of practicing SDU during the 6-month follow-up period compared to those who had never practiced SDU. Moreover, MSM who were on PrEP reported higher likelihood of SDU than those not on PrEP. Studies conducted both locally and internationally consistently demonstrated the significant association between SDU and PrEP use [3,24]. This association may be due to the fact that some MSM may continue or initiate SDU as PrEP has minimized their risk of HIV infection. Health care providers supervising MSM on PrEP should continue to provide education about other sexual risks of SDU, such as acquiring STIs, or reduced adherence to PrEP.

Our results showed that MSM in Hong Kong are frequently exposed to information about psychoactive substances and received invitations to use such substances or engage in SDU/chemsex on social media and gay social networking apps. This exposure was associated with a higher prevalence of SDU during the follow-up period. Previous studies consistently showed that social media and gay social networking apps have

XSL•FO
RenderX

facilitated access to psychoactive substances, SDU, and chemsex for MSM [18,25,26]. One of the important contributions of this study is that it examined the potential mechanism of the associations between social media and peer influences and SDU. Our results suggest that social media exposure may enhance positive perceptions of SDU, which in turn increase the risk of SDU. The significant mediation effect supports mechanisms proposed by the Social Learning Theory [38] to understand how social media may influence one's behavior related to SDU. Most social media sites allow users to leave comments, and express approval or disapproval of the contents. Therefore, users can see how many others (and sometimes exactly who) expressed approval when they view social media. This may explain why MSM find social media content to be a believable and influential source of information. Gay social networking apps provide a quick and convenient way to locate and connect with other MSM nearby, and MSM commonly use these apps to seek SDU or chemsex [18,25-28]. Gay social networking app users may perceive SDU as normative as many users may openly solicit SDU or chemsex in their online profiles [18,25,26]. Additionally, receiving a personal invitation of SDU through these apps may make MSM less likely to refuse. Interventions should consider disseminating messages on reducing harm associated with SDU and chemsex on gay apps platforms to reduce perceptions of SDU as normative.

The association between the number of peers in their social network who had experience with SDU/chemsex and SDU was of marginal statistical significance. Previous studies suggested that peers are common sources of obtaining psychoactive substances among MSM [33,34]. It is also common for MSM to introduce and invite peers to have SDU or chemsex [33,34]. However, the TPB perceptions of SDU were not a significant mediator between peer influence and SDU. Other potential theory-based mediators should be explored in future work. We found significant correlations between peer influence and social media influence, and between peer influence and TPB constructs. MSM in Hong Kong contact each other mainly through social media and gay apps. Some of these peers may be online friends that they have never physically met. Since MSM have very close connections to each other, they may find their peers' experiences particularly valuable [32]. Furthermore, knowing more peers who practice SDU may make MSM perceive such behavior as normative.

These results highlight the importance of modifying perceptions related to SDU based on the TPB. About 20% of the participants held positive attitudes toward SDU. They perceived that SDU would allow them escape from reality, increase their sexual pleasure, and facilitate finding sex partners. Given our findings about the indirect effect of social media influence on SDU behavior, interventions should use social media to shape these attitudes. Health communication messages about reducing substance use during sex and increasing MSM's awareness of the harms associated with SDU and chemsex can be disseminated using the same keywords used to promote SDU and chemsex, allowing the messages to reach MSM at high risk of SDU and chemsex. Gay social networking apps are also useful for delivering such health communication messages. Some apps widely used in Hong Kong (eg, Blued) have incorporated HIV prevention information and referral to HIV testing services [45]. Although less than 5% of the participants perceived support for SDU from significant others, such perceptions were positively associated with SDU during the follow-up period. Health promotion efforts led by influential peers may be useful to cultivate subjective norms against SDU. Health promotion campaigns may also consider sponsoring support groups led by peers living healthy lifestyles to reduce the influence of substance-using peers. Perceived behavioral control of refraining from SDU needs to be further strengthened as it was a protective factor. Enhancement in self-control skills is warranted and rehearsals may be a useful component of future health promotion programs.

This study had some limitations. First, the results were self-reported and subject to social desirability bias, although anonymity likely reduced the bias compared to nonanonymously collected data. Second, participants were recruited by nonprobabilistic sampling in the absence of a sampling frame. As compared to a representative MSM survey in Hong Kong, our participants had a lower prevalence of HIV and sexual risk behaviors, but higher levels of HIV testing. Third, we were not able to obtain the characteristics of participants who refused to join the study; selection bias might exist. However, the response rate for our study was higher than in other published studies involving MSM in China. Fourth, we did not ask behavioral intention to have SDU or chemsex at baseline. Behavioral intention is an important construct of the TPB that predicts actual behaviors [46]. Furthermore, attrition bias might exist. Those who had lower perceived behavioral control in refraining from SDU were more likely to drop out. Perceived behavioral control in refraining from SDU was a protective factor of SDU. Therefore, the prevalence of SDU during the follow-up period is expected to be higher. Finally, the Negative Attitude Scale had a relatively low Cronbach α (.62) in our sample. Although previous studies suggested that Cronbach $\alpha \geq .60$ was acceptable for exploratory research [47,48], caution is still needed when interpreting the results. Future studies are needed to validate this scale.

In sum, MSM in Hong Kong reported a lower prevalence of SDU and chemsex than that of their counterparts in Western countries. Social media and gay social networking apps may be a major source of influence on MSM's perceptions and actual behaviors related to SDU, and interventions delivered on these platforms may be especially effective.

Acknowledgments

The authors would like to thank Ms Willa Dong at the Department of Health Behavior, Gillings School of Global Public Health, the University of North Carolina at Chapel Hill for proofreading and correcting the manuscript.

XSL•FO
RenderX

This study was supported by the Beat Drugs Fund, Narcotics Division, Security Bureau of Hong Kong SAR, China (reference number 160051). The funder had no role in study design; collection, analysis, or interpretation of the data; writing the manuscript, or the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. Halkitis PN, Parsons JT, Stirratt MJ. A Double Epidemic. Journal of Homosexuality 2001 Jul 10;41(2):17-35. [doi: 10.1300/j082v41n02_02] [Medline: 11482426]
2. Edmundson C, Heinsbroek E, Glass R, Hope V, Mohammed H, White M, et al. Sexualised drug use in the United Kingdom (UK): A review of the literature. International Journal of Drug Policy 2018 May;55:131-148. [doi: 10.1016/j.drugpo.2018.02.002] [Medline: 29625796]
3. Maxwell S, Shahmanesh M, Gafos M. Chemsex behaviours among men who have sex with men: A systematic review of the literature. International Journal of Drug Policy 2019 Jan;63:74-89. [doi: 10.1016/j.drugpo.2018.11.014] [Medline: 30513473]
4. Evers YJ, Van Liere GAFS, Hoebe CJPA, Dukers-Muijrers NHTM. Chemsex among men who have sex with men living outside major cities and associations with sexually transmitted infections: A cross-sectional study in the Netherlands. PLoS ONE 2019 May 14;14(5):e0216732. [doi: 10.1371/journal.pone.0216732] [Medline: 31086390]
5. Hibbert MP, Brett CE, Porcellato LA, Hope VD. Psychosocial and sexual characteristics associated with sexualised drug use and chemsex among men who have sex with men (MSM) in the UK. Sex Transm Infect 2019 Apr 12;95(5):342-350. [doi: 10.1136/sextrans-2018-053933] [Medline: 30979782]
6. Tomkins A, George R, Kliner M. Sexualised drug taking among men who have sex with men: a systematic review. Perspect Public Health 2018 May 30;139(1):23-33. [doi: 10.1177/1757913918778872] [Medline: 29846139]
7. Wong NS, Kwan TH, Lee KCK, Lau JYC, Lee SS. Delineation of chemsex patterns of men who have sex with men in association with their sexual networks and linkage to HIV prevention. International Journal of Drug Policy 2020 Jan;75:102591. [doi: 10.1016/j.drugpo.2019.10.015] [Medline: 31756695]
8. Kohli M, Hickson F, Free C, Reid D, Weatherburn P. Cross-sectional analysis of chemsex drug use and gonorrhoea diagnosis among men who have sex with men in the UK. Sex Health 2019;16(5):464. [doi: 10.1071/sh18159] [Medline: 30760386]
9. Sewell J, Cambiano V, Speakman A, Lampe FC, Phillips A, Stuart D, et al. Changes in chemsex and sexual behaviour over time, among a cohort of MSM in London and Brighton: Findings from the AURAH2 study. International Journal of Drug Policy 2019 Jun;68:54-61. [doi: 10.1016/j.drugpo.2019.03.021] [Medline: 30999243]
10. Logan L, Fakoya I, Howarth A, Murphy G, Johnson AM, Rodger AJ, et al. Combination prevention and HIV: a cross-sectional community survey of gay and bisexual men in London, October to December 2016. Euro Surveill 2019 Jun;24(25):1800312 [FREE Full text] [doi: 10.2807/1560-7917.ES.2019.24.25.1800312] [Medline: 31241039]
11. Kurtz SP. Post-Circuit Blues: Motivations and Consequences of Crystal Meth Use Among Gay Men in Miami. AIDS Behav 2005 Mar;9(1):63-72. [doi: 10.1007/s10461-005-1682-3] [Medline: 15812614]
12. Stevens O, Moncrieff M, Gafos M. Chemsex-related drug use and its association with health outcomes in men who have sex with men: a cross-sectional analysis of Antidote clinic service data. Sex Transm Infect 2019 Jun 06;96(2):124-130. [doi: 10.1136/sextrans-2019-054040] [Medline: 31171592]
13. Centre for Health Protection. Factsheet: PRiSM --- HIV prevalence and risk behavioural survey of men who have sex with men in Hong Kong 2017. 2018. URL: https://www.aids.gov.hk/english/surveillance/sur_report/prism2017e.pdf [accessed 2020-06-30]
14. Coffin PO, Santos G, Colfax G, Das M, Matheson T, DeMicco E, et al. Adapted Personalized Cognitive Counseling for Episodic Substance-Using Men Who Have Sex with Men: A Randomized Controlled Trial. AIDS Behav 2014 Feb 9;18(7):1390-1400. [doi: 10.1007/s10461-014-0712-4] [Medline: 24510401]
15. Santos G, Coffin PO, Vittinghoff E, DeMicco E, Das M, Matheson T, et al. Substance use and drinking outcomes in Personalized Cognitive Counseling randomized trial for episodic substance-using men who have sex with men. Drug and Alcohol Dependence 2014 May;138:234-239. [doi: 10.1016/j.drugalcdep.2014.02.015] [Medline: 24641808]
16. Kurtz SP, Stall RD, Buttram ME, Surratt HL, Chen M. A Randomized Trial of a Behavioral Intervention for High Risk Substance-Using MSM. AIDS Behav 2013 Jun 4;17(9):2914-2926. [doi: 10.1007/s10461-013-0531-z] [Medline: 23732957]
17. Mansergh G, Koblin BA, McKirnan DJ, Hudson SM, Flores SA, Wiegand RE, et al. An Intervention to Reduce HIV Risk Behavior of Substance-Using Men Who Have Sex with Men: A Two-Group Randomized Trial with a Nonrandomized Third Group. PLoS Med 2010 Aug 24;7(8):e1000329. [doi: 10.1371/journal.pmed.1000329] [Medline: 20811491]
18. Tan RKJ, Wong CM, Chen MI, Chan YY, Bin Ibrahim MA, Lim OZ, et al. Chemsex among gay, bisexual, and other men who have sex with men in Singapore and the challenges ahead: A qualitative study. International Journal of Drug Policy 2018 Nov;61:31-37. [doi: 10.1016/j.drugpo.2018.10.002] [Medline: 30388567]

XSL•FO
RenderX

19. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
20. Roberto AJ, Shafer MS, Marmo J. Predicting substance-abuse treatment providers' communication with clients about medication assisted treatment: A test of the theories of reasoned action and planned behavior. Journal of Substance Abuse Treatment 2014 Nov;47(5):307-313. [doi: 10.1016/j.jsat.2014.06.002] [Medline: 25092045]
21. Vederhus J, Zemore SE, Rise J, Clausen T, Høie M. Predicting patient post-detoxification engagement in 12-step groups with an extended version of the theory of planned behavior. Addict Sci Clin Pract 2015 Jun 20;10(1):15. [doi: 10.1186/s13722-015-0036-3] [Medline: 26092327]
22. Kao E. 96% of Hongkongers use their smartphone to go online everyday. URL: http://www.scmp.com/news/hong-kong/article/1298330/hong-kong-smartphone-use-doubles-two-years [accessed 2020-06-30]
23. Statista Research Department. Hong Kong social network penetration Q2-Q3 2018. URL: https://www.statista.com/statistics/412500/hk-social-network-penetration/ [accessed 2020-06-30]
24. Kwan TH, Lee SS. Bridging Awareness and Acceptance of Pre-Exposure Prophylaxis Among Men Who Have Sex With Men and the Need for Targeting Chemsex and HIV Testing: Cross-Sectional Survey. JMIR Public Health Surveill 2019 Jul 03;5(3):e13083 [FREE Full text] [doi: 10.2196/13083] [Medline: 31271148]
25. Bourne A, Reid D, Hickson F, Torres RS, Weatherburn P. The Chemsex Study: drug use in sexual settings among gay and bisexual men in Lambeth, Southwark & Lewisham. Technical Report. Sigma Research, London School of Hygiene & Tropical Medicine, London. URL: https://www.lambeth.gov.uk/sites/default/files/ssh-chemsex-study-final-main-report.pdf [accessed 2020-06-30]
26. Van Hout MC, Crowley D, O'Dea S, Clarke S. Chasing the rainbow: pleasure, sex-based sociality and consumerism in navigating and exiting the Irish Chemsex scene. Culture, Health & Sexuality 2019 Jan 02;21(9):1074-1086. [doi: 10.1080/13691058.2018.1529336] [Medline: 30600787]
27. Yeo TED, Ng YL. Sexual risk behaviors among apps-using young men who have sex with men in Hong Kong. AIDS Care 2015 Oct 07;28(3):314-318. [doi: 10.1080/09540121.2015.1093597] [Medline: 26444390]
28. Young SD, Nianogo RA, Chiu CJ, Menacho L, Galea J. Substance use and sexual risk behaviors among Peruvian MSM social media users. AIDS Care 2015 Sep 14;28(1):112-118. [doi: 10.1080/09540121.2015.1069789] [Medline: 26324405]
29. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014 Jun 27;16(6):e157 [FREE Full text] [doi: 10.2196/jmir.3247] [Medline: 24974893]
30. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter Chatter About Marijuana. Journal of Adolescent Health 2015 Feb;56(2):139-145. [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
31. Kelly JA, Amirkhanian YA, Seal DW, Galletly CM, DiFranceisco W, Glasman LR, et al. Levels and Predictors of Sexual HIV Risk in Social Networks of Men who Have Sex with Men in the Midwest. AIDS Education and Prevention 2010 Dec;22(6):483-495. [doi: 10.1521/aeap.2010.22.6.483] [Medline: 21204625]
32. Yun Gao M, Wang S. Participatory communication and HIV/AIDS prevention in a Chinese marginalized (MSM) population. AIDS Care 2007 Jun 14;19(6):799-810. [doi: 10.1080/09540120601114832] [Medline: 17573601]
33. Harrison PA, Fulkerson JA, Park E. The Relative Importance of Social Versus Commercial Sources in Youth Access to Tobacco, Alcohol, and Other Drugs. Preventive Medicine 2000 Jul;31(1):39-48. [doi: 10.1006/pmed.2000.0691] [Medline: 10896842]
34. McCabe SE, Boyd CJ. Sources of prescription drugs for illicit use. Addictive Behaviors 2005 Aug;30(7):1342-1350. [doi: 10.1016/j.addbeh.2005.01.012] [Medline: 16022931]
35. Tyler KA, Handema R, Schmitz RM, Phiri F, Kuyper KS, Wood C. Multi-Level Risk and Protective Factors for Substance Use Among Zambian Street Youth. Substance Use & Misuse 2016 May 04;51(7):922-931. [doi: 10.3109/10826084.2016.1156702] [Medline: 27144671]
36. Janulis P, Birkett M, Phillips G, Mustanski B. Substance use network characteristics and drug and alcohol use behaviors among young men who have sex with men (YMSM). Drug and Alcohol Dependence 2015 Dec;157:188-191. [doi: 10.1016/j.drugalcdep.2015.10.003] [Medline: 26483355]
37. Scalco MD, Meisel SN, Colder CR. Misperception and accurate perception of close friend substance use in early adolescence: Developmental and intervention implications. Psychology of Addictive Behaviors 2016 May;30(3):300-311. [doi: 10.1037/adb0000175] [Medline: 27214169]
38. Moreno MA, Whitehill JM. Influence of Social Media on Alcohol Use in Adolescents and Young Adults. Alcohol Res 2014;36(1):91-100 [FREE Full text] [Medline: 26259003]
39. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ, USA: Prentice-Hall; 1986.
40. Ajzen I. Constructing a TPB questionnaire: conceptual and methodological considerations. Constructing a TPB questionnaire: conceptual and methodological considerations storrs, CT; 2002. URL: https://pdfs.semanticscholar.org/0574/b20bd58130dd5a961f1a2db10fd1fcbae95d.pdf [accessed 2020-06-30]
41. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychological Methods 2002;7(1):64-82. [doi: 10.1037/1082-989x.7.1.64]

XSL•FO
RenderX

42. Valente TW, Fosados R. Diffusion of Innovations and Network Segmentation: The Part Played by People in Promoting Health. Sexually Transmitted Diseases 2006;33(Supplement):S23-S31. [doi: 10.1097/01.olq.0000221018.32533.6d]
43. Chen X, Mo PKH, Li J, Lau JTF. Factors Associated with Drug Use Among HIV-Infected Men Who Have Sex with Men in China. AIDS Behav 2019 Sep 4;24(6):1612-1620. [doi: 10.1007/s10461-019-02660-z]
44. Li J, Mo PKH, Wu AMS, Lau JTF. Roles of Self-Stigma, Social Support, and Positive and Negative Affects as Determinants of Depressive Symptoms Among HIV Infected Men who have Sex with Men in China. AIDS Behav 2016 Feb 20;21(1):261-273. [doi: 10.1007/s10461-016-1321-1]
45. Muessig KE, Bien CH, Wei C, Lo EJ, Yang M, Tucker JD, et al. A mixed-methods study on the acceptability of using eHealth for HIV prevention and sexual health care among men who have sex with men in China. J Med Internet Res 2015 Apr 21;17(4):e100 [FREE Full text] [doi: 10.2196/jmir.3370] [Medline: 25900881]
46. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
47. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
48. Hair J. Multivariate data analysis with readings (5th ed). Englewood Cliffs, NJ, USA: Prentice-Hall; 1998.

Abbreviations

AOR: adjusted odds ratio
CAI: condomless anal intercourse
CFI: Comparative Fit Index
GHB/GBL: γ-hydroxybutyrate
MSM: men who have sex with men
NNFI: Non-Normed Fit Index
NRP: nonregular partner
PrEP: pre-exposure prophylaxis
RMSEA: root mean square error of approximation
RP: regular partner
SDU: sexualized drug use
STI: sexually transmitted infection
TPB: Theory of Planned Behavior

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by HF Sit, Q Zhao; comments to author 06.04.20; revised version received 16.05.20; accepted 03.06.20; published 24.07.20.

Please cite as:
Wang Z, Yang X, Mo PKH, Fang Y, Ip TKM, Lau JTF
Influence of Social Media on Sexualized Drug Use and Chemsex Among Chinese Men Who Have Sex With Men: Observational Prospective Cohort Study
J Med Internet Res 2020;22(7):e17894
URL: http://www.jmir.org/2020/7/e17894/
doi:10.2196/17894
PMID:32706705

©Zixin Wang, Xue Yang, Phoenix K H Mo, Yuan Fang, Tsun Kwan Mary Ip, Joseph T F Lau. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Blended Care-Cognitive Behavioral Therapy for Depression and Anxiety in Real-World Settings: Pragmatic Retrospective Study

Anita Lungu[1], PhD; Janie Jihee Jun[1], PhD; Okhtay Azarmanesh[1], PhD; Yan Leykin[2], PhD; Connie E-Jean Chen[1], MD

[1]Lyra Health, Burlingame, CA, United States

[2]Department of Psychology, Palo Alto University, Palo Alto, CA, United States

Corresponding Author:
Anita Lungu, PhD
Lyra Health
287 Lorton Avenue
Burlingame, CA, 94010
United States
Phone: 1 6505673136
Email: anita@lyrahealth.com

Abstract

Background: The past few decades saw considerable advances in research and dissemination of evidence-based psychotherapies, yet available treatment resources are not able to meet the high need for care for individuals suffering from depression or anxiety. Blended care psychotherapy, which combines the strengths of therapist-led and internet interventions, can narrow this gap and be clinically effective and efficient, but has rarely been evaluated outside of controlled research settings.

Objective: This study evaluated the effectiveness of a blended care intervention (video-based cognitive behavior therapy and internet intervention) under real-world conditions.

Methods: This is a pragmatic retrospective cohort analysis of 385 participants with clinical range depression and/or anxiety symptoms at baseline, measured using Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7), who enrolled in blended care psychotherapy treatment. Participants resided in the United States and had access to the blended care intervention as a mental health benefit offered through their employers. Levels of depression and anxiety were tracked throughout treatment. Hierarchical linear modeling was used to examine the change in symptoms over time. The effects of age, gender, and providers on participants' symptom change trajectories were also evaluated. Paired sample t-tests were also conducted, and rates of positive clinical change and clinically significant improvement were calculated.

Results: The average depression and anxiety symptoms at 6 weeks after the start of treatment were 5.94 and 6.57, respectively. There were significant linear effects of time on both symptoms of depression and anxiety (β=–.49, *P*<.001 and β=–.64, *P*<.001). The quadratic effect was also significant for both symptoms of depression and anxiety (β=.04, P<.001 for both), suggesting a decelerated decrease in symptoms over time. Approximately 73% (n=283) of all 385 participants demonstrated reliable improvement, and 83% (n=319) recovered on either the PHQ-9 or GAD-7 measures. Large effect sizes were observed on both symptoms of depression (Cohen d=1.08) and of anxiety (d=1.33).

Conclusions: Video blended care cognitive behavioral therapy interventions can be effective and efficient in treating symptoms of depression and anxiety in real-world conditions. Future research should investigate the differential and interactive contribution of the therapist-led and digital components of care to patient outcomes to optimize care.

(J Med Internet Res 2020;22(7):e18723) doi:10.2196/18723

KEYWORDS

cognitive behavior therapy; blended psychotherapy; dissemination; implementation; depression; anxiety/anxiety disorders; internet; web based; video psychotherapy

Introduction

Depression and anxiety are the leading causes of disability worldwide [1]. Within work environments, depression and anxiety have been associated with decreased productivity for employees through absenteeism, presenteeism, increased disability leave, or sick leave [2-4]. Although effective psychological treatments for depression and anxiety exist, they

XSL•FO
RenderX

are difficult to access for many individuals [5,6]. Approximately half of the individuals meeting criteria for major depressive disorder in a given year remain untreated or under-treated [7]. Cognitive behavior therapy (CBT) has been rigorously tested and consistently shown to be efficacious in randomized controlled trials and effective in real-world applications for treating depression and anxiety, emerging as the initial treatment of choice [8]. Barriers to large scale dissemination of CBT include insufficient numbers of specialized mental health providers [9], long waitlists to see a provider [10], high cost of therapist delivered care, the stigma associated with seeing a therapist and receiving treatment for mental health disorders [11], as well as travel time required to see a provider in person.

Internet-administered automated cognitive behavior therapy (iCBT) has emerged as a promising cost-effective solution that could narrow the gap between clinical demand and availability of CBT for adults, adolescents, and children [12,13]. iCBT involves the delivery of clinical CBT content via the internet and can involve different formats such as text, video, audio files, and interactive elements. iCBT interventions can be delivered alone or with additional support and guidance from a therapist or a coach [14]. Among the advantages of iCBT over traditional therapist-delivered care are scalability, guaranteed treatment fidelity, increased geographical reach, full temporal availability, ability to progress at one's own pace, savings in travel time, and cost-effectiveness. Several meta-analyses have found iCBT interventions to be clinically effective for a wide range of clinical disorders, including depression, anxiety disorders, and eating disorders [12,15].

iCBT interventions with human support lead to better clinical outcomes compared to unsupported ones [16]. Results of iCBT interventions with human support are sustained in the long term, and supported interventions have higher adherence rates compared to unsupported ones [12,16]. For unsupported interventions, lack of support to sustain motivation for change and of accountability towards a professional can lead to decreased clinical efficacy and higher dropout rates [17,18]. Many iCBT interventions are relatively static and lack the sophistication to be able to adapt to the client's initial and evolving clinical presentation.

Blended care CBT treatments (BC-CBT) have emerged as a newer approach that integrates regular therapist-led CBT sessions with iCBT modules into an integrated treatment, taking advantage of the benefits of both approaches while mitigating their disadvantages. Blended interventions can take place with therapist-led sessions delivered in a face-to-face format [19] or via video [20]. Given that part of the treatment is taking place via Internet-based modules, BC-CBT has the potential to decrease the number of sessions with therapists while achieving similar outcomes to therapist-only treatments [21], which can improve the clinical efficiency and scalability of treatment. BC-CBT may also lead to faster improvements when weekly therapist-led sessions are supplemented with iCBT, resulting in a more intensive therapy experience [22]. BC-CBT addresses the fundamental disadvantages for iCBT, specifically the low initial engagement and high dropout rates, through its human component, by enabling the development of a therapeutic alliance found to be associated with higher motivation to initiate and sustain engagement in care [23,24]. Importantly, having the therapist in charge of clinical assessment, treatment plan, and delivery in BC-CBT allows for more personalization of care compared to iCBT. Specifically, the therapist selects digital modules most relevant to the client's presenting concerns and explains the rationale to the client, linking digital tools to the client's goals, thereby increasing motivation and compliance to treatment. BC-CBT treatments are clinically efficacious in controlled studies for multiple mental health conditions such as depression, anxiety, and substance abuse [19] and multiple settings such as primary care [25], specialized mental health clinics [26], and inpatient [27]. The vast majority of BC-CBT applications described to date consist of in-person face-to-face therapy sessions supplemented with iCBT modules. However, by moving face-to-face BC-CBT therapy sessions to the telehealth modality (ie, the delivery of care via video or audio), the reach of the treatment can be extended, offering treatment to more individuals in need. It has consistently been found that telehealth mental health services are just as effective as in-person care [28].

To date, there have been very few large-scale studies examining the effectiveness of BC-CBT interventions delivered via telehealth on a large scale in real-world settings [20,29]. Dissemination of most interventions outside of tightly controlled research environments introduces challenges, including therapist drift, potentially more complex clinical presentations, as well as possible lower treatment engagement [30], and BC-CBT would likely face similar challenges. Additionally, for BC-CBT delivered by telehealth, acceptability of the video communication channel, as well as the iCBT care components by both therapists and clients, are likely necessary to prevent dropout and preserve clinical effectiveness.

The present study uses data gathered as part of routine care for clients who received BC-CBT treatment as a benefit offered through their employer. The BC-CBT program combines live video-based sessions with a therapist with technology and evidence-based care tools for clients to use in between sessions. BC-CBT therapists are supported in their work to implement BC-CBT via regular individual and group consultations; clients have access to and are encouraged to use evidence-based care tools at any time. These elements can make the psychotherapy experience more clinically efficient, enabling more clients to improve faster. To our knowledge, to date, no study in the United States has examined BC-CBT interventions at a large scale under real-world conditions.

Methods

Study Design

This pragmatic retrospective cohort study used data collected for quality control of a BC-CBT program. Participants resided in the United States and had access to the BC-CBT program as a mental health benefit at no cost to them from their employer companies. Lyra Health partners with Lyra Clinical Associates to offer a behavioral health benefit to companies through which employees and dependents have access to a video BC-CBT program. Employees and their dependents learned of this mental health benefit through information from their employer and

XSL•FO
RenderX

could then access the benefit through registering online with Lyra Health, searching for a provider, and directly enrolling in the BC-CBT program. The program's rigorous quality assurance elements included ongoing individual and team-based clinical consultations and clinical case reviews informed by routine outcome monitoring of depression and anxiety symptomatology.

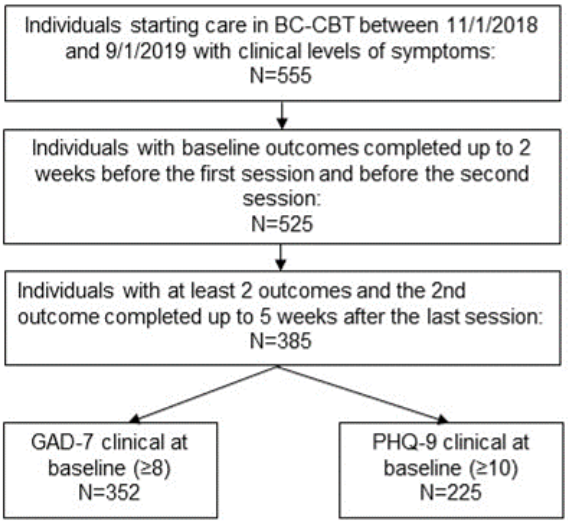
All participants who engaged in the BC-CBT program were asked to complete electronically secure, standardized measures of depression and anxiety every week as well as a satisfaction measure at the end of treatment. No specific length of care was defined. Participants had access to a minimum of 12 sessions, depending on the benefit offered by the sponsoring company. This post-factum analysis of deidentified data gathered from treatment offered by Lyra Clinical Associates was determined to be not human subject research by the Palo Alto University Institutional Review Board.

Participants and Data Inclusion

Participants were individuals who started BC-CBT treatment between November 1, 2018, and September 1, 2019. Participants who sought psychotherapy went through an online onboarding process and answered questions about their symptoms, the impact of symptoms on their general functioning, and their interest in receiving care via video. Those who were not open to seeing a provider via video were referred to in-person therapy and not offered BC-CBT. Exclusion criteria for the BC-CBT program were: being under 18 years of age, active suicidality/self-harm, active homicidality, a current diagnosis of a mental health disorder with psychotic features not stabilized on medications, unstable bipolar disorder, and a current diagnosis of severe alcohol or substance use disorder. Participants also had to score above the clinical cut-off for either the Patient Health Questionnaire-9 (PHQ-9≥10) or the Generalized Anxiety Disorder-7 (GAD-7≥8) on a baseline assessment (n=555). No additional diagnostic assessment was conducted for the study, though each therapist conducted independent clinical interviews for treatment purposes. We considered a baseline assessment to be invalid if it was collected more than 2 weeks before the first therapy session or after the second therapy session. We excluded 30 records with invalid baselines from analyses based on this definition. Participants for whom no assessment was available within 5 weeks after the last therapy session were excluded. We excluded from the analysis 140 records with invalid second assessments per this definition, leaving 385 participant records for analysis. The final scores were the last valid available datapoint after baseline.

For two-thirds of the included records (63%, n=242), final scores were collected after the final therapy session (but <5 weeks after the final session), and the remainder was collected from an earlier session (Figure 1).

Figure 1. Participant flow.



Self-Report Measures

Assessments consisted of the PHQ-9 [31] and the GAD-7 [32], well-validated measures of depression and anxiety, respectively. Clinical cut-offs were PHQ-9≥10 and GAD-7≥8, as research suggests that participants who score at or above these cut-offs are very likely to meet criteria for major depression [31] or to have an anxiety disorder diagnosis [33].

Treatment

The BC-CBT program combined live video-based sessions with a therapist plus digital care tools (lessons and exercises) that clients could access and were encouraged to use between sessions (Multimedia Appendix 1). The digital therapy platform supporting BC-CBT had separate portals with customized functions for clients and therapists, respectively.

Individual Video Psychotherapy

Therapy sessions were conducted on a secure, proprietary HIPAA-compliant video platform developed by Lyra Health. Therapy staff consisted of 49 licensed therapists (licensed clinical psychologists, licensed marriage and family therapists, licensed clinical social workers, or licensed professional counselors) vetted for their commitment to and proficiency in CBT via extensive application reviews and clinical interviews. Approximately 41% (n=20) of the therapists had less than 5 years of experience, 32% (n=16) had between 5 and 10 years of experience, and 27% (n=13) had over 10 years of experience. All therapists were trained on-site with a 2-day workshop on

XSL•FO
RenderX

BC-CBT. The mean number of clinical cases per therapist was 8.2 (SD 7.1). Participants engaged in BC-CBT met with their therapists, either weekly or bi-weekly. A core focus of the program was to ensure a high quality of evidence-based clinical care via providing therapists with peer supervision. Providing clinicians with ongoing consultation and support has been identified as a critical component of treatment dissemination [34] that helps maintain and enhance clinicians' skills [35,36]. Sessions were recorded with the participant's consent. Quality adherence of clinical care was ensured via session review, bi-weekly individual consultation video calls, and video consultation groups with other licensed therapists. The weekly assessment of depression and anxiety symptoms represented important information helping to focus the quality assurance process on clients who were not making progress as expected.

Digital Lessons and Exercises

Following each session, therapists assigned educational digital lessons and exercises to be completed by clients before their next session. The digital treatment components were developed by Lyra Health based on transdiagnostic treatment approaches, such as the Unified Treatment Protocol [37], Acceptance Commitment Therapy [38], and Dialectical Behavior Therapy [39,40]. Example principles and skills taught in the digital lessons and exercises included: clarifying values, understanding emotions, mindful awareness, cognitive restructuring, challenging avoidance (via behavioral activation and exposure), and communication skills.

The digital lessons consisted of animated videos and quizzes to test comprehension and provide corrective feedback. The digital lessons utilized a storytelling approach wherein viewers followed the therapy journey for characters presenting with symptoms of depression or anxiety. This approach has been used in other efficacious iCBT interventions [41] and found to have a normalizing effect for clients. Therapists personalized digital tools in several ways. First, therapists selected digital lessons and exercises to match clients' case formulation (clinician's understanding of clients' presenting concerns viewed from a CBT perspective). Second, to increase motivation for completion of the personalized digital tools, therapists introduced them by linking them to issues immediately relevant to the client and discussed during that specific session, as well as to clients' broader goals. Third, therapists received alerts when clients completed assigned digital lessons and exercises and could send personalized feedback to clients.

The BC-CBT Therapy Platform for Providers

The BC-CBT platform enabled therapists to conduct a variety of BC-CBT therapy tasks before, during, after, and in between therapy sessions seamlessly in a single environment.

Before therapy sessions, therapists could check whether clients had submitted PHQ-9 and GAD-7 outcomes to track their progress in treatment. They could also review clients' completion of assigned practices, clinical outcomes, and communicate asynchronously with clients via secure messaging to encourage them or support them in applying therapy principles and skills.

Therapists utilized the platform to hold their video sessions in a HIPAA secure environment and to record sessions for quality assurance with clients' consent. During sessions, therapists could also share their screen to collaboratively complete digital exercises with the client or preview specific exercises and digital lessons they intended to assign to the client. Modeling such activities was done to increase the likelihood that the client would later complete the exercise or lesson on their own. Providing users with the ability to perform a "virtual rehearsal" of a desired new behavior within a technology tool (such as completing a psychotherapy digital exercise) has been proposed as a feature that can change behavior in the real world [42]. Screen sharing also allowed therapists to review the clinical outcomes responses and progress with their clients, which has been found to make treatment more efficacious and efficient [43].

After sessions, therapists completed therapy notes in the BC-CBT platform. The notes were designed as a decision support system [42] to a) help clinicians create case conceptualizations for clients that would guide treatment goals and treatment plans, b) ensure that treatment goals were indeed addressed during session time, c) increase adherence to the digital elements of BC-CBT like assigning and reviewing digital lessons and exercises, and d) increase consistency and thoroughness of conducting more complex clinical tasks, such as safety risk assessments.

The BC-CBT Therapy Platform for Clients

The digital platform behind BC-CBT supported clients in completing clinical outcomes as well as assigned digital lessons and exercises in several ways. First, clients received automatic notifications encouraging them to complete the assigned outcome measures and digital lessons and exercises if they had not already, and a new session was approaching. Second, clients were also alerted if their therapist sent them feedback on a completed exercise. Third, "tunneling" was utilized as a digital persuasion technique to increase the completion of digital assignments [42]. For example, the client was guided through a digital lesson composed of multiple parts (animated videos, comprehension quiz) such that subsequent parts automatically started if the client did not explicitly stop the experience. Last, clients could revisit all digital lessons and materials assigned for the duration of treatment.

Data Analyses

Two analytical methods were employed to estimate the impact of BC-CBT on depression and anxiety symptomatology. Hierarchical linear modeling (HLM) was used to examine change over time. HLM is a statistical technique that is capable of modeling incomplete data over time. The routinely administered outcomes measures were included as weekly observations for participants. In cases when participants completed multiple outcomes measures during the same week in treatment, an average of the scores for that week was calculated. We included all participants who provided at least two outcome measures meeting all other inclusion criteria. In order to estimate the change in outcomes, we tested for a linear effect of time and a quadratic effect of time, modeling the slowing of the rate of change. HLM estimates the individual

XSL•FO
RenderX

slopes and intercepts for all participants, as well as the sample's average slope and intercept. Participants' age, gender, and provider were added to the analysis as level-2 predictors of the intercept, slope, and quadratic component. For both dependent variables we added level 2 predictors, age, gender, provider, at the intercept, slope, and quadratic effect. As a supplementary statistical analysis, we conducted paired samples *t*-tests between baseline and last available assessment scores for each measure. We calculated Cohen *d*, a conservative measure of effect size for within-subjects designs that controls for the correlation between measurements. We then assessed whether participants attained reliable clinical change (a measure of symptom change beyond what could be attributed to measure error alone), recovery (moving from the clinical range to the subclinical range), and clinically significant change (demonstrated reliable change and recovery; [44]) on either the PHQ-9 or GAD-7.

We used the Chi-square test to assess differences in the distribution of genders between records with a valid baseline assessment that were not included in the analyses (n=140) and those that were included (n=385). Independent samples *t*-tests were used to evaluate differences between those groups in age as well as baseline depression and anxiety symptoms scores. To provide more detailed information on rates of improvement, we performed the paired samples t-test and reliable clinical change and recovery analyses for the entire sample of clients but also separately for individuals scoring in the clinical range of symptoms for depression, anxiety, or both.

Results

Overview

There were no statistically significant differences between records with a valid baseline assessment versus those that were not included in the analyses by gender (χ^2_2=1.13, *P*=.56), age (t_{523}=1.08, *P*=.28), baseline depression severity (t_{523}=0.98, *P*=.33) and baseline anxiety severity (t_{523}=0.86, *P*=.39).

Participants' mean age was 32.8 years (SD 8.0). The majority of participants were between 25 and 34 years old (55%; n=212), 25.4%% (n=98) between 35 and 44 years old, 11.2% (n=43) were between 18 and 24 years old, and 8.3% (32) were above 45 years old. The majority of participants were women (63.4%; n=244), 35.8% (n=138) were men, and 0.8% (n=3) did not specify their gender. The mean number of sessions delivered during a course of BC-CBT was 5.2 (SD 2.9; range 1-17). The mean number of weeks participants spent in treatment was 6.4 (SD 5.3; range 1-25).

Hierarchical Linear Models

Based on prior work that found a negatively accelerated relationship between the number of therapy sessions and improvement in care [45] as well as the observed sample mean in our dataset, we fitted an HLM with intercept, slope, and quadratic effect of time. The time variable was centered at the point of the average length of care (week 6) to reduce collinearity between the linear and quadratic components while keeping the interpretation of the model's parameters more meaningful [46].

Likelihood ratio testing revealed that a quadratic effect of time provided a significant improvement in model fit over a simple linear effect of time. A cubic effect of time was also attempted but found not to improve model fit significantly. The results of the model indicate that all growth effects were significant (Table 1). Specifically, the slopes were negative, suggesting that symptoms of depression and anxiety decrease with time in care. The quadratic slope effects were positive, suggesting that rapid progress initially in care tapers off as clients advance in treatment. There was also significant variability in the growth parameters, indicating that there were significant differences between individuals in how participant's depression and anxiety symptoms changed during care. Likelihood ratio testing revealed a better fit for models allowing both the linear and quadratic components to vary. In other words, participants varied in the trajectory and rate of change in treatment.

For depression symptoms, at 6 weeks after starting treatment (which represents the average length of care), the mean level of symptoms in the sample was 5.94. There was a significant linear effect in depression symptoms (β=–.49 $t_{99.24}$=–20.64, *P*<.001), indicating a .49 decrease in depression symptoms with each week in care. The quadratic effect was also significant (β=.04 $t_{90.41}$=10.72, *P*<.001), suggesting a decelerated decrease in depression symptoms over time (Table 1).

For anxiety symptoms at 6 weeks, the average length of care, the mean level of symptoms in the sample was 6.57. There was a significant linear effect in anxiety symptoms (β=–.64, $t_{274.62}$=–22.43, *P*<.001), indicating a .64 decrease of anxiety symptoms with each week in care. The quadratic effect was also significant (β=.04, $t_{89.79}$=12.17, *P*<.001), suggesting a decelerated decrease in anxiety symptoms over time (Table 2).

Table 1. Unconditional growth model for depression symptoms.

PHQ-9[a]	Estimate	SE	*t*	95% CI	*P* value
Intercept	5.94	.20	28.34	5.53 to 6.36	<.001
Slope	–.49	.02	-20.64	–.53 to –.44	<.001
Quadratic	.04	.004	10.72	.03 to .05	<.001

[a] Patient Health Questionnaire-9

XSL•FO
RenderX

Table 2. Unconditional growth model for anxiety symptoms.

GAD-7[a]	Estimate	SE	*t*	95% CI	*P* value
Intercept	6.57	.19	33.61	6.19 to 6.96	<.001
Slope	–.64	.02	–22.43	–.70 to –.58	<.001
Quadratic	.04	.004	12.17	.04 to .05	<.001

[a] Generalized Anxiety Disorder-7

Age, gender, and provider identifier were included in the models for anxiety and depression symptoms as level-2 predictors of the intercept, slope, and quadratic slope in order to determine whether age, gender, and providers moderated the outcomes for participants. Likelihood ratio testing revealed a better fit for models including age, gender, and provider identifiers as level-2 predictors of the intercept only, thus that model was retained. For both depression and anxiety symptoms, neither the age, gender, or provider were statistically significant, suggesting that the intercept, slope, and quadratic slope were not related to these variables (Tables 3 and 4).

Table 3. Age, gender, and provider as predictors of the intercept, slope, and quadratic components for depression symptoms.

PHQ-9[a]	Estimate	SE	*t*	95% CI	*P* value
Intercept	6.51	.87	7.48	4.80 to 8.22	<.001
Slope	–.58	.02	–20.53	–.63 to –.52	<.001
Quadratic	.04	.004	10.72	.03 to .05	<.001
Age	–.01	.02	–.67	–.05 to .02	.49
Gender	.18	.37	.48	–.55 to .91	.62
Provider	–.007	.01	–.43	–.04 to .02	.66

[a] Patient Health Questionnaire-9

Table 4. Age, gender, and provider as predictors of the intercept, slope, and quadratic components for anxiety symptoms.

GAD-7[a]	Estimate	SE	t	95% CI	p-value
Intercept	6.75	.78	8.63	5.21 - 8.29	<.001
Slope	-.64	.02	-22.25	-.70 - -.58	<.001
Quadratic	.04	.003	12.10	.04 - .05	<.001
Age	-.002	.02	-.14	-.04 - .03	.88
Gender	.28	.33	.85	-.37 - .94	.39
Provider	-.01	.01	-1.01	-.04 - .01	.31

[a] Generalized Anxiety Disorder-7

Paired samples t-tests

Participants Starting With Depression or Anxiety Symptoms in Clinical Range

The mean (SD) scores of pre-treatment GAD-7 across all participants with baseline clinical range of depression or anxiety symptoms (n=385) was 11.72 (3.89) and of PHQ-9 was 10.77 (4.71), corresponding to moderate anxiety and moderate depression (Table 5).

Table 5. Changes in depression and anxiety symptoms for participants starting at clinical levels of depression or anxiety symptoms (n=385).

Measure	Baseline score, mean (SD)	Follow-up score, mean (SD)	Paired differences, mean (SD)	95% CI of the difference	t-value (*df*)	*P* value	Cohen *d*
PHQ-9[a]	10.77 (4.71)	5.57 (4.92)	5.20 (5.52)	4.52-5.88	18.49 (384)	<.001	1.08
GAD-7[b]	11.72 (3.89)	6.07 (4.55)	5.65 (5.40)	5.06-6.25	20.55 (384)	<.001	1.33

[a] Patient Health Questionnaire-9

[b] Generalized Anxiety Disorder-7

XSL•FO
RenderX

Reliable clinical change on either the PHQ-9 and/or GAD-7 measures was observed in 283 participants (73.5%), 319 participants (82.8%) recovered on either measure, and 258 participants (67%) demonstrated clinically significant improvement in either depression or anxiety (Table 6).

Table 6. Reliable improvement and recovery for participants grouped by their baseline levels of depression and anxiety symptoms.

Participant sub-group	Reliable improvement, n (%)	Recovery, n (%)	Reliable improvement and recovery, n (%)	Reliable improvement or recovery, n (%)
Baseline depression (PHQ-9 ≥10; n=225)	149 (66.2)	167 (74.2)	141 (62.7)	175 (77.8)
Baseline anxiety (GAD-7≥8; n=352)	245 (69.6)	247 (70.2)	217 (61.7)	275 (78.1)
Baseline depression and anxiety (PHQ-9≥10 and GAD-7≥8; n=192)	115 (59.8)[a]	120 (62.5)[b]	105 (54.6)[c]	130 (67.7)[d]
Baseline depression or anxiety (PHQ-9≥10 or GAD-7≥8; n=385)	283 (73.5)[e]	319 (82.8)[f]	258 (67.0)[g]	336 (87.2)[h]

[a] Calculated as reliable improvement on PHQ-9 AND GAD-7

[b] Calculated as recovery on PHQ-9 AND GAD-7

[c] Calculated as (reliable improvement and recovery on PHQ-9) OR (reliable improvement and recovery on GAD-7)

[d] Calculated as (reliable improvement OR recovery on PHQ9) OR (reliable improvement OR recovery on GAD7)

[e] Calculated as reliable improvement on PHQ-9 OR GAD-7

[f] Calculated as recovery on PHQ-9 OR GAD-7

[g] Calculated as (reliable improvement AND recovery on PHQ-9) OR (reliable improvement AND recovery on GAD-7)

[h] Calculated as (reliable improvement OR recovery on PHQ-9) OR (reliable improvement OR recovery on GAD-7)

Participants Starting With Depression Symptoms in Clinical Range

At pre-treatment, 225 participants (58.4%) scored in the clinical range on the PHQ-9. The mean of the pre-treatment PHQ-9 score was 13.88 (SD 3.40), corresponding to moderate depression, while the post-treatment score was 6.76 (SD 5.35). Results of paired samples *t*-tests revealed that for this group, depression scores decreased significantly, with an average reduction of 7.12 (SD 5.74) points, (t_{224}=18.62, *P*<.001, Cohen *d*=1.59), suggesting a large effect of treatment on depression symptoms.

Reliable clinical change in depression scores was observed in 149 participants (66.2%), and 167 participants (74.2%) recovered on the PHQ-9. A total of 141 participants (62.7%) demonstrated clinically significant improvement on the PHQ-9, ie, meeting criteria for both reliable clinical change and recovery (Table 6).

Participants Starting With Anxiety Symptoms in Clinical Range

At pre-treatment, 352 participants (91.4%) scored in the clinical range on the GAD-7. The mean of pre-treatment GAD-7 score for this group was 12.32 (SD 3.49), corresponding to moderate anxiety, while the post-treatment score was 6.24 (SD 4.56).

Anxiety scores decreased significantly by an average of 6.08 (SD 5.31) points, (t_{351}=21.47, *P*<.001, Cohen *d*=1.50), suggesting a substantial effect of treatment on anxiety. Reliable clinical change was observed in 245 (69.6%) participants, 247 participants (70.2%) recovered, and 217 (61.7%) participants demonstrated clinically significant improvement (Table 5).

Participants Starting With Depression and Anxiety Symptoms in Clinical Range

A total of 192 participants (49.9%) had pre-treatment PHQ-9 and GAD-7 scores in the clinical range, suggesting comorbid symptoms for both depression and anxiety. In this group, depression scores decreased significantly by an average of 7.13 (SD 5.93) points, (t_{191}=16.67, *P*<.001, Cohen *d*=1.56), and anxiety by an average of 6.79 (SD 5.92) points, (t_{191}=15.88, *P*<.001, Cohen *d*=1.55), suggesting large and similar effects of treatment on both anxiety and depression (Table 7).

Reliable clinical change in both measures was observed for 115 participants (59.8%), 120 participants (62.5%) recovered on both measures, and 105 participants (54.6%) demonstrated clinically significant improvement in both depression and anxiety (Table 6).

XSL•FO
RenderX

Table 7. Changes in depression and anxiety symptoms for participants starting at clinical levels of depression and anxiety symptoms (n=192).

Measure	Baseline score (SD)	Follow-up score (SD)	Paired differences mean (SD)	95% CI of the difference	*t* value (df)	*P* value	Cohen *d*
PHQ-9[a]	14.14 (3.49)	7.01 (5.43)	7.13 (5.93)	6.21-8.05	16.67 (191)	<.001	1.56
GAD-7[b]	13.60 (3.61)	6.82 (5.04)	6.79 (5.92)	5.91-7.67	15.88 (191)	<.001	1.55

[a] Patient Health Questionnaire-9

[b] Generalized Anxiety Disorder-7

Discussion

To our knowledge, this is the first study to examine the effectiveness of BC-CBT delivered via video at a large scale in a real-world setting within the US. The BC-CBT program evaluated combined live video-based sessions with a therapist plus technology-based care tools that clients had access to and were encouraged to use in between sessions. Our results suggest that BC-CBT can be effective in significantly reducing symptoms of depression and anxiety. Age, gender, and specific psychotherapists did not moderate the changes in symptoms observed. The within-subjects effect sizes were similar and similarly large for both depression (d=1.59) and anxiety (d=1.50) symptom improvements for individuals starting at clinical levels of depression and anxiety symptoms at baseline. These results were similar to those reported in various randomized trials for adult depression [47] and similar to those reported for effectiveness trials of adult anxiety disorders [48]. Notably, these results were observed after treatment lasting an average of 5.2 sessions and 6.4 weeks, which is significantly shorter than the conventional 12 to 16 weekly sessions protocol of most CBT treatments [49,50]. Outcomes collected more than 2 weeks before the first therapy session or more than 5 weeks after the last therapy session were considered invalid. If baseline outcomes were collected too early, or too much time had passed between the end of care and the last assessment, such data likely no longer reflected actual symptoms at the beginning and end of care. Our findings suggest that the BC-CBT program and its intense psychotherapy experience, resulting from combining video therapy sessions with personalized digital tools, were successful in delivering a clinically efficient treatment, as demonstrated by participants making great progress in a short duration of time. Thus, CBT can be delivered via alternative modalities, ie, via video and blended with technologically based approaches, in an ecologically valid manner, and remain both effective and efficient.

An essential component of CBT is engaging in homework between therapy sessions. Indeed, homework compliance has been consistently associated with better outcomes for CBT [51,52], yet compliance is commonly low [53]. In our BC-CBT program, therapists assigned homework such as digital lessons and exercises through a digital platform, which enabled them to monitor whether and when the participant did the homework. If the participant struggled or procrastinated to complete the homework, the therapist could send a personalized reminder and encouragement about the homework, which could increase the client's motivation and compliance. The combination of individual provider-led therapy sessions with personalized digital components to foster consolidation of CBT skills allowed clients to receive personalized care and encouragement for increased treatment engagement and completion. Future research should investigate the specific impact of digital homework such as digital lessons and exercises, and personalized messages from the provider to encourage homework completion, on the effectiveness and efficacy of the BC-CBT treatment.

This study has several limitations. We did not utilize an RCT design, and our results should thus be interpreted with caution. Although an RCT design might offer a more precise understanding regarding the effectiveness of a treatment, randomization is not feasible in a setting where care is offered as an employer-sponsored benefit given ethical complexities related to limiting care options for participants. Given that no control group was included, we cannot be entirely certain that the results reported are indeed due to the BC-CBT intervention rather than due to the passage of time or other factors. However, comparing our outcomes to well-conducted RCTs, we have observed that our outcomes are comparable, lending support for the validity of our results. Another study limitation is that though we assessed for meaningful change by calculating reliable clinical change, recovery, and clinically significant change, whether symptom improvement was maintained in the long-term is not known. Participants were either employees or dependents of employed individuals, mostly younger, and without symptoms of severe mental illness or acute suicidality, or homicidality; thus, the results may not generalize to other populations or those with acute safety concerns. Outcomes were based on self-reported measures, which can be less reliable in assessing levels of symptoms compared to clinician-administered interviews due to possible misinterpretation of the questions by clients and the lack of follow-up clarifying questions from the provider.

In conclusion, this study demonstrates that BC-CBT can effectively reduce depressive and anxiety symptoms outside of the clinical trial environment. Future research should examine the separate and synergistic contributions of the therapist-led versus the digital components of care to treatment outcomes. Having a better understanding of the optimal ratio of the two modalities, as well as the matching of this ratio to client-specific variables in order to maximize treatment outcomes for different participants and presenting issues, will allow us to enhance and personalize BC-CBT.

XSL•FO
RenderX

Acknowledgments

The authors would like to express their appreciation for the Lyra therapists who work every day to support clients in leading more fulfilling, productive lives. The authors would like to thank Chelsey Wilks PhD, for assistance with the HLM analysis.

Conflicts of Interest

AL and OA are employed by Lyra Health, receive income from Lyra Health, and have been granted equity in Lyra Health. JJ and CC are employed by Lyra Health and Lyra Clinical Associates, receive income from Lyra Health and Lyra Clinical Associates, and have been granted equity in Lyra Health. YL has nothing to disclose.

Multimedia Appendix 1
Sample BC-CBT components: secure messaging, video therapy, and digital lessons.
[PNG File , 697 KB - jmir_v22i7e18723_app1.png]

References

1. Friedrich M. Depression Is the Leading Cause of Disability Around the World. JAMA 2017 Apr 18;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28418490]
2. Goetzel RZ, Long SR, Ozminkowski RJ, Hawkins K, Wang S, Lynch W. Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. J Occup Environ Med 2004 Apr;46(4):398-412. [doi: 10.1097/01.jom.0000121151.40413.bd] [Medline: 15076658]
3. Plaisier I, Beekman A, de Graaf R, Smit J, van Dyck R, Penninx B. Work functioning in persons with depressive and anxiety disorders: the role of specific psychopathological characteristics. J Affect Disord 2010 Sep;125(1-3):198-206. [doi: 10.1016/j.jad.2010.01.072] [Medline: 20185180]
4. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. JAMA 2003 Jun 18;289(23):3135-3144. [doi: 10.1001/jama.289.23.3135] [Medline: 12813119]
5. Harvey AG, Gumport NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. Behav Res Ther 2015 May;68:1-12 [FREE Full text] [doi: 10.1016/j.brat.2015.02.004] [Medline: 25768982]
6. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]
7. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
8. David D, Cristea I, Hofmann SG. Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
9. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
10. Mireau R, Inch R. Brief solution-focused counseling: a practical effective strategy for dealing with wait lists in community-based mental health services. Soc Work 2009 Jan 01;54(1):63-70. [doi: 10.1093/sw/54.1.63] [Medline: 19205258]
11. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
12. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
13. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015 Mar;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
14. Cuijpers P, Riper H, Karyotaki E. Internet-Based Cognitive-Behavioral Therapy in the Treatment of Depression. Focus (Am Psychiatr Publ) 2018 Oct;16(4):393-394 [FREE Full text] [doi: 10.1176/appi.focus.20180019] [Medline: 31975931]
15. Pittock A, Hodges L, Lawrie SM. The effectiveness of internet-delivered cognitive behavioural therapy for those with bulimic symptoms: a systematic review : A review of iCBT treatment for bulimic symptoms. BMC Res Notes 2018 Oct 22;11(1):748 [FREE Full text] [doi: 10.1186/s13104-018-3843-2] [Medline: 30348226]
16. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]

XSL•FO
RenderX

17. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an Internet-based cognitive behaviour therapy program. Aust N Z J Psychiatry 2006 Jan;40(1):59-62. [doi: 10.1080/j.1440-1614.2006.01743.x] [Medline: 16403040]
18. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
19. Erbe D, Eichert H, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
20. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
21. Jacmon J, Malouff J, Taylor N. Treatment of Major Depression: Effectiveness of Cognitive Behavior Therapy with an Internet Course as a Central Component. EJAP 2010 Jan 03;5(2). [doi: 10.7790/ejap.v5i2.153]
22. Cuijpers P, Huibers M, Ebert DD, Koole SL, Andersson G. How much psychotherapy is needed to treat depression? A metaregression analysis. J Affect Disord 2013 Jul;149(1-3):1-13. [doi: 10.1016/j.jad.2013.02.030] [Medline: 23528438]
23. Rumpold G, Doering S, Smrekar U, Schubert C, Koza R, Schatz DS, et al. Changes in motivation and the therapeutic alliance during a pretherapy diagnostic and motivation-enhancing phase among psychotherapy outpatients. Psychotherapy Research 2005 Jan;15(1-2):117-127. [doi: 10.1080/10503300512331327092]
24. Anderson SR, Tambling R, Yorgason JB, Rackham E. The mediating role of the therapeutic alliance in understanding early discontinuance. Psychother Res 2019 Oct 05;29(7):882-893. [doi: 10.1080/10503307.2018.1506949] [Medline: 30079816]
25. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013 Aug 05;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
26. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
27. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online Self-Help as an Add-On to Inpatient Psychotherapy: Efficacy of a New Blended Treatment Approach. Psychother Psychosom 2017 Nov;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
28. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
29. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: The study protocol for the MasterMind project. Internet Interventions 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
30. Teachman BA, Drabick DAG, Hershenberg R, Vivian D, Wolfe BE, Goldfried MR. Bridging the gap between clinical research and clinical practice: introduction to the special section. Psychotherapy (Chic) 2012 Jun;49(2):97-100 [FREE Full text] [doi: 10.1037/a0027346] [Medline: 22642515]
31. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
33. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
34. Edmunds JM, Beidas RS, Kendall PC. Dissemination and Implementation of Evidence-Based Practices: Training and Consultation as Implementation Strategies. Clin Psychol (New York) 2013 Jun 1;20(2):152-165 [FREE Full text] [doi: 10.1111/cpsp.12031] [Medline: 24072959]
35. Beidas RS, Edmunds JM, Marcus SC, Kendall PC. Training and consultation to promote implementation of an empirically supported treatment: a randomized trial. Psychiatr Serv 2012 Jul;63(7):660-665 [FREE Full text] [doi: 10.1176/appi.ps.201100401] [Medline: 22549401]
36. Mannix KA, Blackburn IM, Garland A, Gracie J, Moorey S, Reid B, et al. Effectiveness of brief training in cognitive behaviour therapy techniques for palliative care practitioners. Palliat Med 2006 Sep;20(6):579-584. [doi: 10.1177/0269216306071058] [Medline: 17060250]
37. Barlow D, Farchione T, Sauer-Zavala S. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders. In: Barlow DH, editor. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders. Oxford, England: Therapist Guide. Oxford University Press; 2017.

XSL•FO
RenderX

38. Hayes S, Strosahl K, Wilson K. Acceptance and Commitment Therapy The Process and Practice of Mindful Change. In: Watkins B, editor. Acceptance and Commitment Therapy The Process and Practice of Mindful Change. New York, NY: The Guilford Press; 2012.
39. Linehan MM. DBT Skills Training Handouts and Worksheets. In: Watkins B, editor. DBT Skills Training Handouts and Worksheets. New York, NY: The Guilford Press; 2014.
40. Neacsiu AD, Eberle JW, Kramer R, Wiesmann T, Linehan MM. Dialectical behavior therapy skills for transdiagnostic emotion dysregulation: A pilot randomized controlled trial. Behaviour Research and Therapy 2014 Aug;59:40-51. [doi: 10.1016/j.brat.2014.05.005]
41. Williams AD, O'Moore K, Mason E, Andrews G. The effectiveness of internet cognitive behaviour therapy (iCBT) for social anxiety disorder across two routine practice pathways. Internet Interventions 2014 Oct;1(4):225-229. [doi: 10.1016/j.invent.2014.11.001]
42. Fogg BJ. Persuasive Technology Using Computers to Change What We Think and Do. In: Buehler M, Drew J, editors. Persuasive Technology Using Computers to Change What We Think and Do. Burlington, Massachusetts: Morgan Kaufmann; 2002.
43. Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. Psychotherapy 2018 Dec;55(4):520-537. [doi: 10.1037/pst0000167]
44. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
45. Howard KI, Kopta SM, Krause MS, Orlinsky DE. The dose–effect relationship in psychotherapy. American Psychologist 1986;41(2):159-164. [doi: 10.1037/0003-066x.41.2.159]
46. Biesanz JC, Deeb-Sossa N, Papadakis AA, Bollen KA, Curran PJ. The Role of Coding Time in Estimating and Interpreting Growth Curve Models. Psychological Methods 2004;9(1):30-52. [doi: 10.1037/1082-989x.9.1.30]
47. Rubin A, Yu M. Within-Group Effect Size Benchmarks for Cognitive-Behavioral Therapy in the Treatment of Adult Depression. Soc Work Res 2017;41(3):135-144. [doi: 10.1093/swr/svx011]
48. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009 Aug;77(4):595-606. [doi: 10.1037/a0016032] [Medline: 19634954]
49. APA Division 12. Cognitive Therapy for Depression. Society of Clinical Psychology. 2016. URL: https://www.div12.org/treatment/cognitive-therapy-for-depression/ [accessed 2020-01-05]
50. APA Division 12. Cognitive Behavioral Therapy for Social Anxiety Disorder. Society of Clinical Psychology. 2016. URL: https://www.div12.org/treatment/cognitive-behavioral-therapy-for-social-anxiety-disorder/ [accessed 2020-01-05]
51. Kazantzis N, Deane F, Ronan K. Homework assignments in cognitive and behavioral therapy: A meta-analysis. Clin Psychol Sci Pract 2000;7(2):189-202. [doi: 10.1093/clipsy.7.2.189]
52. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The Relationship Between Homework Compliance and Therapy Outcomes: An Updated Meta-Analysis. Cognit Ther Res 2010 Oct;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]
53. Helbig S, Fehm L. PROBLEMS WITH HOMEWORK IN CBT: RARE EXCEPTION OR RATHER FREQUENT? Behav. Cognit. Psychother 1999;32(3):291-301. [doi: 10.1017/S1352465804001365]

Abbreviations

BC-CBT: blended care cognitive behavioral therapy
CBT: cognitive-behavioral therapy
GAD-7: Generalize Anxiety Disorder-7
HIPAA: Health Insurance Portability and Accountability Act
HLM: hierarchical linear modeling
iCBT: internet-administered automated cognitive behavior therapy
PHQ: Patient Health Questionnaire-9
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.03.20; peer-reviewed by K Mathiasen, D Erbe; comments to author 30.03.20; revised version received 20.05.20; accepted 14.06.20; published 06.07.20.

Please cite as:
Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CEJ
Blended Care-Cognitive Behavioral Therapy for Depression and Anxiety in Real-World Settings: Pragmatic Retrospective Study
J Med Internet Res 2020;22(7):e18723
URL: https://www.jmir.org/2020/7/e18723
doi:10.2196/18723
PMID:32628120

XSL•FO
RenderX

©Anita Lungu, Janie Jihee Jun, Okhtay Azarmanesh, Yan Leykin, Connie E-Jean Chen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Transdiagnostic Guided Internet-Delivered Protocol for Emotional Disorders Versus Treatment as Usual in Specialized Care: Randomized Controlled Trial

Alberto González-Robles[1], PhD; Amanda Díaz-García[1], PhD; Azucena García-Palacios[1,2], PhD; Pablo Roca[3], MSc; Josep Antoni Ramos-Quiroga[4,5,6,7], PhD; Cristina Botella[1,2], PhD

[1]Department of Basic and Clinical Psychology, and Psychobiology, Universitat Jaume I, Castellón de la Plana, Spain

[2]CIBER Fisiopatología Obesidad y Nutrición (CIBERObn), Instituto Carlos III, Madrid, Spain

[3]Department of Personality, Assessment, and Clinical Psychology, Universidad Complutense de Madrid, Madrid, Spain

[4]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[5]Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Barcelona, Spain

[6]Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain

[7]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

Corresponding Author:
Alberto González-Robles, PhD
Department of Basic and Clinical Psychology, and Psychobiology
Universitat Jaume I
Sos Baynat Avenue, Campus del Riu Sec
Research II Building, Labpsitec
Castellón de la Plana, 12071
Spain
Phone: 34 964387646 ext 7646
Email: vrobles@uji.es

Abstract

Background: Anxiety disorders and depression (emotional disorders) are highly prevalent mental disorders. Extensive empirical evidence supports the efficacy of cognitive behavioral therapy (CBT) for the treatment of these disorders. However, there are still some barriers related to their dissemination and implementation, which make it difficult for patients to receive these treatments, especially in public health care settings where resources are limited. Recent advances in improving CBT dissemination encompass different perspectives. One is the transdiagnostic approach, which offers treatment protocols that can be used for a range of emotional disorders. Another approach is the use of the internet to reach a larger number of people who could benefit from CBT.

Objective: This study aimed to analyze the effectiveness and acceptability of a transdiagnostic internet-delivered protocol (EmotionRegulation) with human and automated guidance in patients from public specialized mental health care settings.

Methods: A 2-armed randomized controlled trial (RCT) was conducted to compare the effectiveness of EmotionRegulation with treatment as usual (TAU) in specialized mental health care. In all, 214 participants were randomly assigned to receive either EmotionRegulation (n=106) or TAU (n=108). Measurement assessments were conducted at pre- and postintervention and at a 3-month follow-up.

Results: The results revealed the superiority of EmotionRegulation over TAU on measures of depression (d=0.41), anxiety (d=0.35), and health-related quality of life (d=−0.45) at posttreatment, and these gains were maintained at the 3-month follow-up. Furthermore, the results for expectations and opinions showed that EmotionRegulation was well accepted by participants.

Conclusions: EmotionRegulation was more effective than TAU for the treatment of emotional disorders in the Spanish public mental health system. The implications of this RCT, limitations, and suggestions for future research are discussed.

Trial Registration: ClinicalTrials.gov NCT02345668; https://clinicaltrials.gov/ct2/show/NCT02345668

(J Med Internet Res 2020;22(7):e18220) doi:10.2196/18220

KEYWORDS

transdiagnostic; internet; cognitive behavioral therapy; emotional disorders; depression; anxiety; specialized care

XSL•FO
RenderX

Introduction

Disorder-Specific Cognitive Behavioral Therapy for Emotional Disorders

Anxiety and depressive disorders, also known as emotional disorders (EDs) [1], have the highest prevalence rates among psychological disorders [2,3], and are associated with substantial costs [4,5] and disability [6,7]. In the past three decades, research efforts to decrease the burden of these disorders have led to the development and evaluation of cognitive behavioral treatments in randomized controlled trials (RCTs) for each ED (ie, disorder-specific protocols), such as depression [8,9] and several anxiety disorders, including generalized anxiety disorder (GAD) [10,11], panic disorder (PD) and agoraphobia (AG) [12], social anxiety disorder (SAD) [13], and obsessive compulsive disorder (OCD) [14].

Although there is a large body of evidence showing the efficacy and effectiveness of *disorder-specific* cognitive behavioral therapy (CBT), in the past 15 years, an increasing number of researchers have agreed that there are some problems that hinder the optimal deployment of these treatments. The main drawback of disorder-specific treatments stems from the high comorbidity rates observed among anxiety disorders and between anxiety disorders and depressive disorders, with comorbidity estimates for these disorders ranging between 40% and 80% [15,16]. Thus, because disorder-specific treatment protocols focus on treating a specific diagnosis, the accompanying comorbid disorders do not receive therapeutic attention [17,18]. This problem becomes clearer when we take into account research linking comorbidity to aspects such as greater severity [3], increased chronicity rates [19], and a worse clinical course [20]. Another problem with disorder-specific treatments is that subthreshold symptoms that do not meet diagnostic thresholds for a particular disorder, but might be important to treat, normally go untreated [17]. Similarly, these protocols do not address diagnoses that do not fit any specific category, despite their clinical relevance, that is, *not otherwise specified* (NOS) anxiety and depressive disorders [15]. Finally, each disorder-specific treatment requires the use of different handbooks and protocols, which increases the economic costs and the amount of training needed to gain the knowledge and skills necessary to cover the array of anxiety and depressive disorders [17].

Transdiagnostic Treatments for Anxiety and Depression

Transdiagnostic treatments have emerged as an alternative to the traditional disorder-specific approach that has dominated CBT research for the past 30 years. Transdiagnostic treatments have been developed and tested in several RCTs for anxiety disorders [21-24] and anxiety and depressive disorders [25-27], and their number continues to grow. Moreover, the efficacy and effectiveness of transdiagnostic treatments have been shown in different meta-analytic reviews, comparing them with different control groups, such as waiting list, attention control, and treatment as usual (TAU) [28-31], with pooled effect sizes (Hedges *g*) in the medium to large range for overall measures of anxiety (0.65-0.82) and depression (0.79-0.84). Moreover, an additional meta-analysis reported equivalent effects of transdiagnostic treatments (g=1.06) and disorder-specific treatments (d=0.95) on anxiety outcomes [32]. However, these meta-analyses include a number of mixed studies of different orientations, such as theory-based transdiagnostic treatments and tailored CBT. Therefore, these studies make it difficult to determine the effectiveness of each orientation (eg, the effectiveness of theory-based transdiagnostic treatments). The effectiveness of transdiagnostic treatments has also been shown in a meta-analysis by García-Escalera et al [33], with pooled effect sizes of g=0.80 for anxiety and g=0.72 for depression. Unlike the abovementioned meta-analyses, this study has the particularity that it only included theory-based transdiagnostic treatments. The main characteristic of theory-based transdiagnostic treatments, also known as *mechanistically transdiagnostic* treatments or transdiagnostic treatments based on *shared mechanisms* [34], is that they are designed to address the common psychopathological processes underlying anxiety and depression. Among the mechanistically transdiagnostic treatments for EDs, the Unified Protocol (UP) [35,36] stands out as one of the most empirically supported transdiagnostic protocols for anxiety and depression [23,37,38]. The UP is a CBT transdiagnostic protocol developed to address the underlying psychopathological processes that are common to anxiety and depressive disorders, with a particular focus on neuroticism, (low) extraversion, and emotion dysregulation, which have been shown to play a key role in the onset and maintenance of these disorders [39,40]. Thus, the main goal of the UP is to teach patients strategies to regulate their emotions in a more adaptive way through the following core treatment modules: (1) present-focused emotional awareness, (2) cognitive flexibility, (3) identification and prevention of emotional avoidance patterns, (4) increasing awareness and tolerance to emotion-elicited physical sensations, and (5) graded (interoceptive and situational) exposure procedures. The overall efficacy of the UP was first shown in an RCT where it was compared with a waitlist control group [23] and, more recently, in a larger RCT where it was compared with well-established disorder-specific CBT protocols for anxiety disorders [37]. In addition, some research shows the long-term effects of the UP [41] and its ability to produce changes in the temperament dimensions of behavioral inhibition (BI) and behavioral activation (BA) [42]. BI and BA have been conceptualized as 2 neurological systems representing motivational tendencies that are sensitive to threat and reward environmental cues, respectively [43]. These aspects have been intimately linked to neuroticism and negative affect and extraversion and positive affect [42]. Another treatment approach that includes principles or components that could be useful to target EDs is dialectical behavioral therapy (DBT) [44,45]. DBT initially emerged as a theoretical model and a treatment approach for the treatment of suicidal behaviors and borderline personality disorder, with the general aim of teaching strategies to change patterns of emotion dysregulation [46]. Of the range of strategies, DBT places a special emphasis on increasing experiential awareness and acceptance (mindfulness *what* and *how* techniques), but it also includes behavioral strategies such as the *opposite action*, designed to diminish distress by engaging in behaviors or actions opposite to those associated with negative emotions [44]. More

XSL•FO
RenderX

recently, DBT skills have been adapted and successfully applied to several anxiety and depressive disorders [47,48], which suggests that they could be used in a transdiagnostic manner to improve the symptomatology of these disorders. For example, DBT principles may be combined with other evidence-based components (eg, components of the UP) to strengthen their effectiveness. The strategy of integrating CBT principles and components from different evidence-based therapies and orientations is consistent with the notion of process-based CBT, defined by Hofmann and Hayes [49]. As the authors state, "modern CBT places much less focus on protocols for syndromes and more focus on evidence-based processes linked to evidence-based procedures" [49].

Internet-Delivered Interventions

In the past two decades, one of the most evident efforts made by researchers has been to take advantage of the possibilities offered by information and communication technologies to improve the assessment and treatment of psychological disorders. A clear example would be the use of the internet to increase the dissemination of empirically supported psychological treatments to anyone in need [50]. Research has shown promising and compelling evidence that internet-delivered psychological interventions are effective for a variety of psychosocial problems, including anxiety and depressive disorders [51,52]. The main advantages of internet-delivered treatments over traditional delivery methods (eg, face-to-face therapy) include widespread access and dissemination [53], a nonstigmatizing way of receiving psychological treatment [54], and increased cost-effectiveness [55].

Research has shown that transdiagnostic internet-delivered treatments are more effective than control groups [29] and that these treatments are at least as effective as individual and group face-to-face transdiagnostic treatments [33]. However, most of the existing literature on transdiagnostic treatments is limited to studies conducted in community settings, with few studies carried out in public contexts such as primary or specialized care [21]. Indeed, to our knowledge, no transdiagnostic internet-delivered treatments for anxiety and depression have been conducted in specialized public mental health care. This is somewhat surprising because transdiagnostic internet-delivered treatments in this particular setting could have several advantages for both clinicians and the patients attending these centers. First, anxiety and depressive disorders are disorders with the highest prevalence rates [2,3]. Second, resources in these settings are usually scarce, which affects both the quantity and the quality of the mental health care provided [56]. Third, a large percentage of patients with anxiety and depressive disorders do not receive treatment in mental health care centers [57]. In the specific case of Spain, most patients attending public mental health units suffer from anxiety and depressive disorders [58], the ratio of clinical psychologists to patients is one of the lowest in Europe [59], and patients have to endure long waitlists to receive treatment [60]. Fourth, transdiagnostic treatments can be provided at a lower cost (eg, in terms of training) [61]. Finally, the use of the internet can help improve mental health services, for instance, by reducing the waiting period to receive face-to-face treatment [62] or by implementing these treatments as part of a stepped-care model that takes patients' profiles and needs into account. Thus, each patient can be assigned the most appropriate treatment [63], leaving face-to-face therapy for those patients who are less likely to benefit from internet-delivered interventions.

This Study

Taking all of this into consideration, in this study, an RCT was conducted to test the effectiveness of a transdiagnostic internet-delivered protocol for ED (EmotionRegulation), compared with TAU provided in Spanish public specialized mental health care. EmotionRegulation includes components of the UP and the skills from DBT (eg, mindfulness *what* and *how* techniques), and it was designed to target a wide range of EDs, including major depressive disorder (MDD), dysthymic disorder (DD), PD, AG, SAD, GAD, OCD, anxiety NOS, and depression NOS. It was hypothesized that (1) the EmotionRegulation group would outperform the TAU group on measures of overall anxiety and depression, temperament (ie, BI and BA), and health-related quality of life (QoL) at posttreatment; (2) these posttreatment changes would be maintained at follow-up; (3) a significantly greater clinical change would be observed in EmotionRegulation compared with TAU; and (4) participants in the EmotionRegulation group would rate the treatment as acceptable (scores on expectations and opinion ≥7/10).

Methods

Study Design

A 2-armed RCT was conducted. Participants were randomly assigned in a 1:1 ratio to one of the following 2 conditions: (1) EmotionRegulation and (2) TAU. Participants were stratified by principal diagnosis, performing block randomization in blocks of 4 to ensure that all the principal diagnoses were equally represented across conditions. Computer-generated random number sequences were obtained using statistical software (Epidat 4.1, SourceForge). This task was performed by an independent researcher who was unaware of the characteristics of the study.

The study was conducted in compliance with the study protocol, the Consolidated Standards of Reporting Trials (CONSORT) statement [64,65], the CONSORT of Electronic and Mobile HEalth Applications and onLine TeleHealth guidelines [66], and the Declaration of Helsinki and good clinical practice. A full description of the study protocol has been reported elsewhere [67]. The RCT obtained ethical approval from the Ethics Committee of Universitat Jaume I (Castellón, Spain) and the Clinical Research Ethics Committees of 3 hospitals (Consorcio Hospitalario Provincial de Castellón, Hospital Universitario de la Ribera, and Hospital Universitario Vall d'Hebron). The study protocol was registered at ClinicalTrials.gov (NCT02345668) on July 27, 2015. The duration of the intervention period was 18 weeks for participants in both conditions, and participants' assessments were conducted at pre- and posttreatment and at 3- and 12-month follow-ups. Both the intervention (EmotionRegulation) and the assessment instruments, except the diagnostic interview, were delivered through a web platform designed by our research group [68].

XSL•FO
RenderX

All transferred data were secured via Advanced Encryption Standard-256 encryption. Study researchers conducting posttreatment and follow-up assessments (ie, diagnostic interviews) were blinded to the participants' treatment conditions. To ensure blinding of the evaluators, participants were informed that an independent researcher would contact them to conduct follow-up assessments, and they were asked not to disclose the treatment condition to which they had been allocated. This study reports pre- to posttreatment data and data at 3-month follow-up.

Sample Size

Several studies were considered for the calculation of the expected sample size [23,69,70]. On the basis of a minimum power of 0.80 in a 1-tailed test (ie, *t* test for differences between 2 independent means), an α of .05, and an estimated dropout rate of approximately 30%, a sample size of 78 participants per condition was determined to detect a posttreatment effect size of 0.40 (Cohen *d*) between the 2 conditions. In addition, based on the literature [71,72], an estimated dropout rate of approximately 30% was expected. Thus, the final sample size was set at 100 participants per condition (total of 200 participants). The G*Power software (version 3.1.9.4, Heinrich-Heine-Universität) was used to calculate the sample size [73].

Participants

Participants were recruited from adult outpatients attending Spanish public specialized mental health care services (mental health units) to seek psychological and/or psychiatric treatment between July 2015 and June 2019. Initial recruitment was performed by clinical psychologists and psychiatrists working in these centers, and it took place in 3 different hospitals: Consorcio Hospitalario Provincial de Castellón (Castellón de la Plana), Hospital Universitario de la Ribera (Valencia), and Hospital Universitario Vall d'Hebrón (Barcelona). Recruitment was performed as follows: (1) once psychiatrists and clinical psychologists had identified a potential candidate, they offered the patient the possibility of participating in the study and described the study characteristics to him or her; (2) patients who were interested in participating gave their informed written consent, and the clinician filled out a document with the participants' sociodemographic and clinical characteristics (moreover, in this stage, participants were provided with a document containing information about the study); (3) one of the researchers involved in the study contacted the participants by phone to schedule a face-to-face appointment to evaluate eligibility criteria using a structured diagnostic interview; and (4) whenever a participant met the eligibility criteria, an independent researcher (unaware of the study characteristics) was contacted to implement randomization, and participants completed the remaining assessment instruments (self-reported questionnaires) through web-based surveys.

Participants were selected based on the following inclusion criteria: (1) aged 18 years or older; (2) ability to understand and read Spanish; (3) having access to the internet at home and an email address; (4) meeting Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnostic criteria [74] for ED (ie, MDD, DD, depression NOS, PD, AG, SAD, GAD, anxiety NOS, and OCD); (5) providing written informed consent; (6) not suffering from a severe mental disorder (schizophrenia, bipolar disorder, and alcohol and/or substance dependence disorder); (7) not presenting a high risk of suicide; (8) not suffering from a disabling medical disease that prevented the participant from carrying out the psychological treatment; and (9) not receiving another psychological treatment during the study (in the experimental group). Pharmacological treatment was allowed, but participants had to be taking the same dose during the 2 months before enrolling in the study. In addition, participants in the experimental group whose medication was increased or changed during the study period were excluded from the trial (decreases in pharmacological treatment were accepted). There was no monetary compensation for participation in the study under any of the treatment conditions.

Instruments

Clinical Outcomes

Diagnosis Interview

Clinical diagnoses were obtained using the mini-international neuropsychiatric interview (MINI) version 5.00 [75,76], a brief structured diagnostic psychiatric interview for the assessment of key DSM-IV and International Classification of Diseases, 10th Revision, diagnoses.

Principal Outcomes

Beck Depression Inventory, Second Edition

Beck depression inventory, second edition (BDI-II) [77,78], is a self-report questionnaire with 21 items about the different symptoms characterizing MDD, added together to obtain the total score, which can be a maximum of 63 points. The instrument has shown good internal consistency (α=.76-.95). The Spanish version also showed high internal consistency (α=.87) for both the general and clinical populations (α=.89). Cronbach α for the BDI-II in this study was .90.

Beck Anxiety Inventory

The Beck anxiety inventory (BAI) [79,80] is a 21-item self-report scale that assesses anxiety, with a maximum score of 63 points. Each item has a 4-point severity scale (from *not at all* to *severely*) that addresses symptoms experienced during the previous week. Previous validation studies have shown an internal consistency ranging from 0.85 to 0.94 as well as convergent and divergent validity. The Spanish version of the BAI has demonstrated high internal consistency (α=.93). Cronbach α for the BAI in this study was .92.

Secondary Outcomes

Behavioral Inhibition Scale and Behavioral Activation Scale

The behavioral inhibition scale (BIS) and behavioral activation scale (BAS) [81,82] contains 20 items rated from 1 to 4, with 7 BIS subscale items that evaluate emotional responses of individuals to impending negative events and 13 BAS items that evaluate the behavioral and emotional responses of individuals to potentially positive events. The BIS and BAS have shown good reliability in individuals with EDs (α=.73-.92) and good convergent and discriminant validity as indicators of

XSL•FO
RenderX

temperament. The internal consistency of the Spanish version ranges between 0.65 and 0.82. Cronbach α for the BIS and BAS subscales in this study were .61 and .80, respectively.

Quality of Life EuroQoL-5D-3L Questionnaire

Quality of life EuroQoL-5D-3L (EQ-5D-3L) questionnaire [83,84] is a generic instrument that measures health-related QoL and consists of 2 parts. Part 1 assesses self-reported problems in each of the following 5 domains: mobility, self-care, daily activities, pain/discomfort, and anxiety/depression. Each domain is divided into 3 levels of severity corresponding to no problems, some problems, and extreme problems, yielding a population-based preference score or societal index (SI). A total of 243 theoretically possible health states can be obtained, and the SI is calculated on the basis of these health states. Values range from 1 (best health state) to 0 (death). However, this index may also provide negative values that correspond to health states perceived as worse than death. Utility scores for these health states were assigned using the available Spanish population tariffs. Part 2 records the subjects' self-assessed health on a visual analog scale (VAS), a 10-cm vertical line on which the best and worst imaginable health states score 100 and 0, respectively. In this study, health-related QoL was assessed using the VAS.

Disorder-Specific Measures

Disorder-specific symptoms were evaluated using 4 different self-report questionnaires. Symptoms of GAD were assessed using the Penn State Worry Questionnaire [85,86]. PD and AG symptoms were evaluated using the Panic Disorder Severity Scale, Self-Reported [87,88]. SAD symptoms were evaluated using the Social Interaction Anxiety Scale [89,90], and OCD symptoms were assessed using the Obsessive Compulsive Inventory-Revised [91,92]. All 4 instruments have shown adequate psychometric properties in both the original and Spanish validations. Cronbach αs for these scales in this study were .78, .89, .80, and .91, respectively. More details about these assessment instruments have been described elsewhere [67].

Treatment Acceptability

Expectations and Opinions of Treatment Scales

These questionnaires were adapted from the study by Borkovec and Nau [93]. Each scale is made up of 5 items, rated from 0 (*nothing at all*) to 10 (*completely*), which ask how logical the treatment seems to be (*How logical do youthink this treatment is?*), to what extent it satisfies the patient (*How satisfied are you with the treatment?*), whether the patient would recommend it to a person with the same problem (*To what extent do you feel confident recommending this treatment to a friend who has the same problems?*), whether it could be used to treat other psychological problems (*To what extent do you think this treatment could be useful in treating other psychological problems?*), and its usefulness for the patient's problem (*To what extent do you think this treatment will be/was helpful to you?*). The expectation scale was applied after the treatment rationale was explained. Its objective is to measure subjective patient expectations regarding this treatment. The opinion scale was administered when the patient had completed the treatment, and it was designed to assess satisfaction with this treatment.

Treatments

EmotionRegulation

Following randomization, participants in the treatment condition were contacted via telephone by a researcher who provided free access to EmotionRegulation, a 12-module transdiagnostic internet-delivered protocol for the treatment of ED, namely, MDD, DD, depression NOS, PD, AG, GAD, SAD, anxiety NOS, and OCD. The protocol is delivered through a web platform [68] designed by our research group. Access to the web platform is through a unique username-password combination and is available 24 hours a day. The treatment was first developed as a manualized transdiagnostic treatment protocol with handbooks for both patient and therapist and then adapted to be delivered through a web-based platform. The web platform has successfully been used in previous RCTs exploring the efficacy of internet-delivered treatments for several disorders, such as depression [94] and flying phobia [95].

The main core components are based on the UP [35,36], but the protocol also contains treatment strategies derived from DBT [46]. The principal aim of the treatment components in EmotionRegulation is to learn and practice adaptive ways to regulate emotions from a transdiagnostic perspective, with the following treatment components: present-focused emotional awareness, cognitive flexibility, emotional avoidance and emotion-driven behaviors, and exposure procedures (interoceptive and situational). The protocol also contains a module to facilitate the patient's engagement with the therapy (motivation for change), a module with psychoeducation about emotions, and a relapse prevention module. Regarding the DBT components, greater emphasis was placed on the present-focused emotional awareness component by adapting and including strategies such as the *what* and *how* techniques and the concept of radical acceptance. Moreover, the opposite action was integrated into the treatment as a way to address emotion-driven behaviors. EmotionRegulation includes a Welcome module that contains general information about the protocol and its goals as well as recommendations for benefiting from it and 12 treatment modules (described in Table 1).

The modules are sequential to enable step-by-step movement through the program. The program duration can vary among the users, and participants had access to the protocol for a maximum period of 18 weeks. Moreover, participants were allowed to use the program any time they wanted during the trial period (ie, during the follow-up periods).

Regarding guidance, all participants in this condition received therapist and automated support. Therapist support consisted of (1) an initial face-to-face session to explain the characteristics of the study and administer the diagnostic interview to confirm the eligibility criteria, (2) an initial phone call encouraging participants to start the intervention after the baseline assessments had been completed, (3) 1 weekly brief phone call (maximum of 10 min) during the treatment period, and (4) a final phone call (once the treatment had ended) to remind participants that they would be allowed to use the program at

XSL•FO
RenderX

any time during the trial period and that they would be contacted for follow-up assessments. Automated support consisted of 2 weekly text messages reminding participants about the importance of completing the homework tasks and encouraging them to review the treatment modules. Text messages were sent through a secure web platform [96]. This web platform was only used to send text messages (unidirectionally) with predefined contents, and they did not include personal information that could have identified the participants. Additional details about the treatment and support protocols, as well as other functionalities of the web-based platform, have been published elsewhere [67]. Finally, it is important to note that all participants in this condition were allowed to continue to receive TAU, but only for monitoring of pharmacological treatment. Participants who received other forms of treatment during the study period (eg, psychological treatment) or who experienced increases or changes in pharmacological treatment were excluded from the analyses.

Table 1. Treatment modules and their objectives.

Module	Objective
1. Introduction to treatment	Provides a framework about the role of emotion regulation in ED[a].
2. Motivation for change and goal setting	To analyze pros and cons of changing, emphasize the importance of being motivated, and help to establish significant life goals.
3. Understanding the role of emotions	Provides psychoeducation about the roles and functions of emotions and trains the patient to track the 3 components of emotional experiences.
4. Nonjudgmental emotional awareness and acceptance of emotional experiences	Aims to train the patient in nonjudgmental emotional awareness (ie, mindfulness *what* and *how* skills) and the acceptance of emotional experiences.
5. Practicing present-focused awareness	To continue to practice the acceptance of emotional experiences and increase awareness of physical sensations, thoughts, emotions, and daily activities.
6. Learning to be flexible	Focuses on the identification of maladaptive ways of thinking (ie, thinking traps).
7. Practicing cognitive flexibility	Aims to teach the patients strategies to modify thinking traps (ie, cognitive reappraisal). It also provides information about intrusive thoughts and how to deal with them.
8. Emotional avoidance	Aims to teach the patients to identify the emotion avoidance strategies that contribute to the maintenance of ED.
9. Emotion-driven behaviors	To learn the concept of EDBs[b] and replace their maladaptive EDB with other more adaptive behaviors.
10. Accepting and facing physical sensations	To teach the role of physical sensations in the emotional response and provide training in interoceptive exposure.
11. Facing emotions in the contexts in which they occur	To build exposure hierarchies to help the patients begin to face situation-elicited avoided emotions.
12. Relapse prevention	To review what patients have learned throughout the program, schedule the future practice of the learned strategies, and teach the patient how to identify and cope with future high-risk situations.

[a]ED: emotional disorder.

[b]EDBs: emotion-driven behaviors.

Treatment as Usual

TAU was treatment as delivered in current daily practice by psychiatrists and clinical psychologists in the mental health centers in Spain. TAU in this study was provided by 3 hospitals: Consorcio Hospitalario Provincial de Castellón (Castellón de la Plana), Hospital Universitario de la Ribera (Valencia), and Hospital Universitario Vall d'Hebron (Barcelona). To maximize the external validity of this RCT, participants in this condition were allowed to receive either psychiatric treatment (ie, prescription and monitoring of antidepressant and/or anxiolytic medication), psychological treatment (including case management, group psychotherapy, empathic listening, and/or supportive counseling), or a combination of both. The frequency of visits during the 18-week treatment period varied depending on the type of treatment (ie, psychiatric or psychological) provided to the participant. Patients in the TAU condition who were already receiving any of the aforementioned treatments at the time of enrollment were informed that they would continue to receive these services during the treatment period. Furthermore, participants receiving a treatment other than those provided in the mental health unit were excluded from the trial. All participants allocated to TAU were offered free access to the treatment platform after the study ended.

Therapists and Treatment Fidelity

The treatment and support protocols were administered by doctoral students with at least two years of experience in the diagnosis, psychological assessment, and application of CBT for different ED. Several steps were taken to ensure treatment fidelity. First, therapists had previously been trained in the application of the treatment modules. Second, a support protocol (ie, weekly phone calls and automated text messages) was developed to be applied to all the participants in the EmotionRegulation condition. This support protocol has been briefly described earlier, but more details can be found in the study by González-Robles et al [67]. Third, to increase diagnosis reliability, all therapists involved in the participants' assessment

XSL•FO
RenderX

were trained in the application of the diagnostic interview (MINI).

Data Analysis Plan

All analyses were performed using the Statistical Package for Social Sciences version 25.

First, chi-square tests for categorical data and independent samples *t* tests for continuous data were performed to confirm that there were no significant differences between the groups at baseline on any of the sociodemographic and clinical variables.

Intention-to-treat (ITT) analyses were performed following Newman's guidelines [97], using maximum likelihood (ML) estimation through the expectation maximization imputation method. To handle missing data, we followed the procedure suggested by Hair et al [98]. First, we explored the types of missing data and determined that data were missing at the construct level. On the basis of this, we concluded that the data were susceptible to imputation. Second, the quantity of missing data was analyzed to ensure that none of the measures exceeded the recommended limits to implement this method [99]. Third, Little's missing completely at random (MCAR) tests were carried out to analyze the pattern of missing data, concluding that all missing data were MCAR (X^2_{25}=28.7; *P*=.28). Finally, a sensitivity analysis was performed on the main outcomes to compare the results of the per-protocol sample (ie, completers) with the imputed values. This analysis revealed that the ML estimation was not likely to produce biased estimations in the main analyses, reaching the same conclusions in both the completers and the imputed dataset (per-protocol: $F_{BAI\ (1,127)}$=6.1; *P*=.02; and $F_{BDI\ (1,127)}$=15.54; *P*<.001; ITT: $F_{BAI\ (1,197)}$=4.79; *P*=.03; and $F_{BDI\ (1,197)}$=12.97; *P*<.001).

To test the first hypothesis and control for baseline differences, analyses of covariance (ANCOVAs) were performed to compare the effects of the groups on measures of anxiety, depression, temperament, and health-related QoL, taking condition as the between-subject variable and the pretreatment scores as covariates. The use of ANCOVAs for the analysis has been recommended by several authors as a more powerful tool to analyze data in studies with randomized designs [100,101].

To test the second hypothesis, a 2 (condition: EmotionRegulation vs TAU) × 3 (time: pretreatment vs posttreatment vs 3-month follow-up) mixed analysis of variance (ANOVA) was performed to test whether the differences between EmotionRegulation and TAU (ie, between-subjects factors) were maintained at follow-up (ie, within-subject factor). The following assumptions for the mixed ANOVA were analyzed: normality (Shapiro-Wilk test), homoscedasticity (Levene test), independence (nonparametric Runs test), and sphericity (Mauchly test). The degrees of freedom were corrected using Greenhouse-Geisser whenever the sphericity assumption was violated. Moreover, pairwise Bonferroni-corrected tests were used for posthoc comparisons.

To compute the magnitude of both within-group and between-group changes, effect sizes (Cohen *d*) were calculated by dividing the differences between means by the pooled SD. Effect sizes were interpreted according to Cohen convention: effect sizes of 0.20 are considered low, effect sizes of 0.50 are considered medium, and effect sizes of 0.80 and above are considered large [102].

To test the third hypothesis, we explored the clinical significance of the changes achieved by the participants as well as potential deterioration rates using Jacobson and Truax's reliable change index (RCI) [103] for the main outcome measures (BDI-II and BAI) in the completer sample for posttreatment and follow-up measurements. First, the cutoff points for the posttreatment and follow-up scores were determined to be within the range of a functional distribution. The RCI was then calculated to test the clinically significant change, with an RCI of |1.96| or greater (*P*<.05). Finally, both criteria were taken into account to classify participants into the following 4 categories: (1) recovered: when the change is significantly reliable (RCI≥|1.96|; *P*<.05) and the posttreatment score is located within the range of the functional distribution (mean [SD 2]), (2) improved: when the change is significantly reliable but the posttreatment score is below the functional level, (3) not changed: when the change is not significantly reliable and the posttreatment score does not reach the functional level, and (4) deteriorated: when the change is significantly reliable but the posttreatment score is worse than the pretreatment score.

Finally, to test the fourth hypothesis, the scores on expectations and opinions were analyzed by calculating means and SDs for each of the items on the expectation and opinion of treatment scales. In addition, 1-way ANOVAs were performed to analyze the significance of the differences between expectations and opinions.

Results

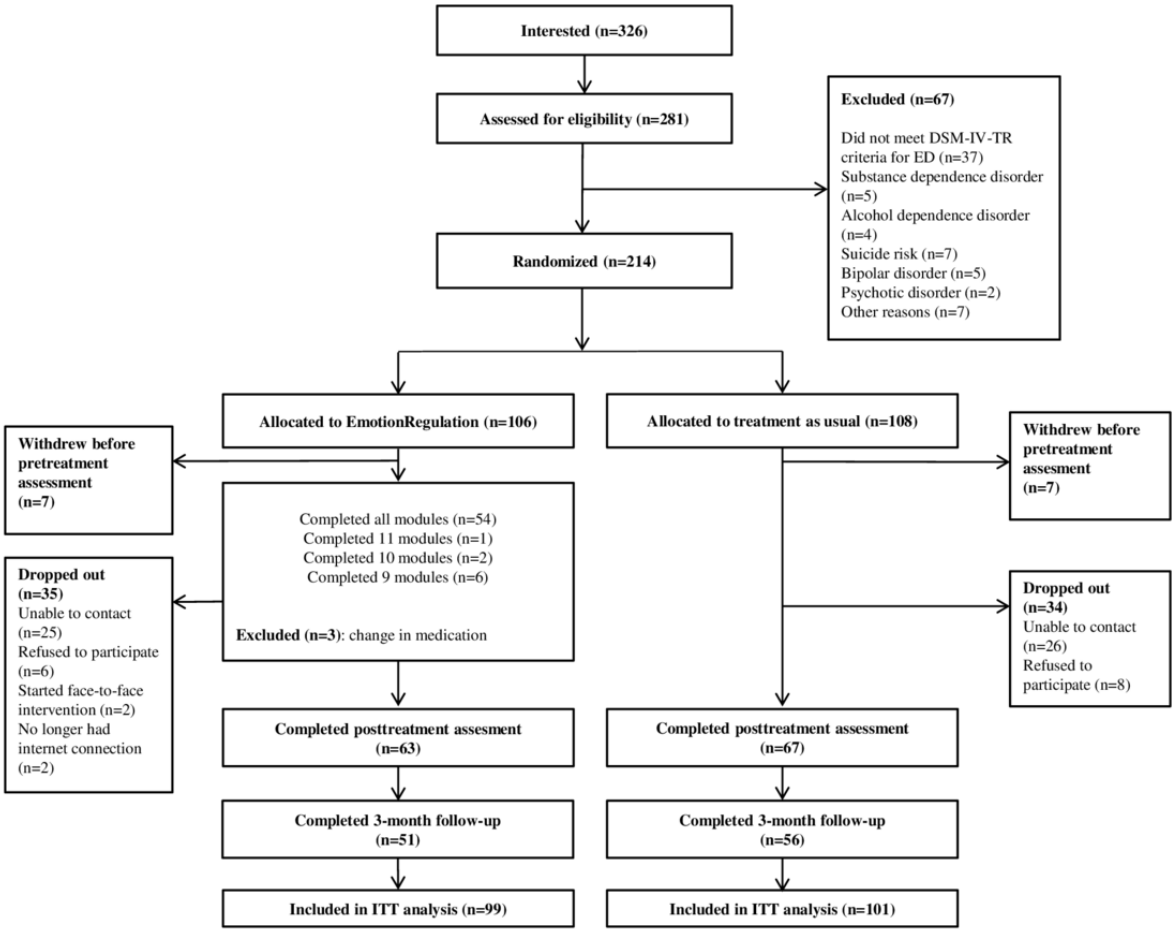
Participant Flow and Attrition

A flowchart of the study participants is displayed in Figure 1. A total of 326 patients expressed interest in the study, 281 of whom were assessed for eligibility. Of these 281, 67 participants were excluded from the study. A total of 214 participants were randomized to either EmotionRegulation (n=106) or TAU (n=108). In addition, 7 patients in each condition withdrew from the study before the pretreatment assessment. Consequently, these participants were not included in any of the analyses.

Regarding attrition, 35 participants in the EmotionRegulation condition (35/106, 33.0%) and 34 in the TAU condition (34/108, 31.5%) dropped out of the study (reasons for dropout are shown in Figure 1). In addition, 3 participants in the EmotionRegulation condition had to be excluded from the trial because of a change in their pharmacological treatment during the treatment period. Posttreatment data were obtained from 63 participants (63/99, 64%) in the EmotionRegulation condition and from 67 participants (67/101, 66.3%) in the TAU condition. Follow-up data were collected from 51 participants (51/99, 52%) in the EmotionRegulation condition and 56 participants (56/101, 55.4%) in the TAU condition. Finally, 99 participants in the EmotionRegulation group and 101 participants in the TAU condition were included in the ITT analysis.

XSL•FO
RenderX

Figure 1. Flowchart of participants. ED: emotional disorder; DSM-IV-TR; Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision; ITT: intention-to-treat.



Baseline Characteristics

Participants (N=200) had a mean age of 38.44 years (SD 10.80; range 18-68), and they were mostly females (138/200, 69.0%). Table 2 provides the sociodemographic and clinical characteristics for both conditions at baseline. There were no significant differences between the EmotionRegulation and TAU groups at baseline on any of the sociodemographic and clinical characteristics. Moreover, no significant differences were found for medication, principal diagnosis, number of comorbid diagnoses, or clinical severity on any of the measures.

XSL•FO
RenderX

Table 2. Demographic and clinical characteristics of the sample at baseline (N=200).

Variable	EmotionRegulation regulation (n=99)	Treatment as usual (n=101)	X^2 (*df*)	*t* test (*df*)	*P* value
Age (years), mean (SD)	38.64 (10.61)	38.25 (11.03)	N/A[a]	0.25 (198)	.80
Sex, n (%)			**1.3 (1)**	**N/A**	**.26**
Female	72 (72)	66 (65.3)			
Male	27 (27)	35 (34.7)			
Marital status, n (%)			**1.1 (3)**	**N/A**	**.78**
Single	22 (22)	26 (25.7)			
Married or partnered	63 (63)	65 (64.4)			
Divorced or widowed	14 (14)	10 (9.9)			
Education, n (%)			**2.1 (2)**	**N/A**	**.35**
Basic studies	26 (26)	36 (35.6)			
Secondary studies	41 (41)	35 (34.7)			
University studies	32 (32)	30 (29.7)			
Occupation, n (%)			**3.3 (6)**	**N/A**	**.77**
Student	9 (9)	11 (10.9)			
Housekeeper	6 (6)	9 (8.9)			
Employed	45 (45)	36 (35.6)			
Unemployed	23 (23)	22 (21.8)			
Off work	13 (13)	17 (16.8)			
Retired	3 (3)	6 (5.9)			
Monthly income (€), n (%)			**1.0 (4)**	**N/A**	**.91**
None	27 (27)	28 (27.7)			
<641.40 (US $699.45)	20 (20)	16 (15.8)			
641.40-1282.80 (US $699.46-1398.89)	32 (32)	38 (37.6)			
1282.81-2565.60 (US $1398.90-2797.78)	18 (18)	17 (16.8)			
>2565.60 (US $2798.78)	2 (2)	2 (2.0)			
Principal diagnosis, n (%)			**2.7 (8)**	**N/A**	**.95**
GAD[b]	23 (23)	26 (27.7)			
AG[c]	16 (16)	13 (12.9)			
PD[d]	9 (9)	5 (5.0)			
SAD[e]	4 (4)	4 (4.0)			
OCD[f]	8 (8)	12 (12.0)			
MDD[g]	20 (20)	22 (21.8)			
DD[h]	7 (7)	6 (5.9)			
Anxiety NOS[i]	10 (10)	9 (8.9)			
Depression NOS	2 (2)	3 (3.0)			
Comorbid diagnoses, n			**N/A**	**N/A**	**N/A**
GAD	10	18			
PD	6	5			

XSL•FO
RenderX

Variable	EmotionRegulation regulation (n=99)	Treatment as usual (n=101)	X^2 (*df*)	*t* test (*df*)	*P* value
AG	18	22			
SAD	7	10			
OCD	2	4			
MDD	15	18			
DD	15	5			
Anxiety NOS	3	1			
Depression NOS	1	0			
Alcohol abuse	1	2			
Substance abuse	0	2			
Number of comorbid disorders, n (%)			**2.3 (3)**	**N/A**	**.50**
0	49 (49)	41 (40.6)			
1	29 (29)	38 (37.6)			
2	15 (15)	13 (12.9)			
≥ 3	6 (6)	8 (7.9)			
Medication, n (%)			**5.2 (3)**	**N/A**	**.16**
None	29 (29)	18 (17.8)			
Antidepressant	22 (22)	20 (19.8)			
Anxiolytic	10 (10)	17 (16.8)			
Both	38 (38)	46 (45.5)			

[a]N/A: not applicable.
[b]GAD: generalized anxiety disorder.
[c]AG: agoraphobia.
[d]PD: panic disorder.
[e]SAD: social anxiety disorder.
[f]OCD: obsessive compulsive disorder.
[g]MDD: major depressive disorder.
[h]DD: dysthymic disorder.
[i]NOS: not otherwise specified.

Amount of Support Provided

Participants in the EmotionRegulation condition were provided a mean of 49.97 min (SD 41.20 min) of clinician support delivered through phone calls. In addition, an initial face-to-face session was scheduled with all the patients in both conditions (EmotionRegulation and TAU) to explain the study and perform the screening assessment, with an approximate duration of 60 min for each participant. Regarding automated support, that is, weekly text messages, participants in the EmotionRegulation group were sent a mean of 24.61 text messages (SD 8.80).

Effectiveness of EmotionRegulation on Primary and Secondary Outcome Measures

Table 3 provides the means and SDs for the 2 conditions at pre- and posttreatment and at 3-month follow-up on both primary and secondary outcome measures. Additional data on change scores and within-group effect sizes for diagnosis-specific measures are reported in Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Descriptive statistics for EmotionRegulation and treatment as usual at pretreatment, posttreatment, and 3-month follow-up.

Instrument	EmotionRegulation (n=99), mean (SD)			Treatment as usual (n=101), mean (SD)		
	Pre-T[a]	Post-T[b]	F/U[c]	Pre-T	Post-T	F/U
Beck depression inventory-II	23.49 (11.01)	15.54 (10.9)	15.70 (11.97)	24.08 (11.69)	19.85 (12.85)	17.90 (13.23)
Beck anxiety inventory	20.00 (11.88)	15.08 (10.12)	15.41 (10.50)	22.27 (12.93)	18.88 (11.31)	18.11 (11.21)
Behavioral inhibition and behavioral activation scale—behavioral inhibition system	23.32 (2.76)	22.30 (2.67)	21.81 (2.67)	23.40 (2.87)	22.87 (2.44)	22.44 (2.42)
Behavioral inhibition scale and behavioral activation scale—behavioral activation system	35.26 (5.93)	36.26 (5.31)	34.94 (5.27)	35.84 (5.58)	35.01 (5.98)	34.07 (6.04)
EuroQoL-5D-3L questionnaire	55.86 (16.72)	65.38 (14.63)	63.12 (15.18)	53.56 (18.25)	58.02 (17.46)	57.81 (17.28)

[a]Pre-T: pretreatment.

[b]Post-T: posttreatment.

[c]F/U: follow-up.

Pretreatment to Posttreatment Effects

Principal Outcome Measures

The ANCOVAs of the baseline-corrected postintervention scores revealed a significant condition effect on anxiety (BAI: $F_{1,197}$=4.79; P=.03; η^2 partial=0.02) and depression (BDI-II: $F_{1,197}$=12.97; P<.001; η^2 partial=0.06), reflecting that the EmotionRegulation group showed significantly lower posttreatment anxiety and depression scores than the TAU group.

Secondary Outcome Measures

Regarding the measures of BI and BA (BISBAS), the ANCOVAs yielded a significant condition effect for the BA dimension (BAS: $F_{1,197}$=9.66; P=.002; η^2 partial=0.05), indicating that patients in the EmotionRegulation group had significantly higher BA scores than those in the TAU group. Although patients in the EmotionRegulation group showed greater improvements in the BI dimension, no significant differences between groups were observed for this dimension (BIS: $F_{1,197}$=2.44; P=.12; η^2 partial=0.01). On the other hand, the ANCOVA revealed a significant condition effect on health-related QoL (EQ-5D-3L: $F_{1,197}$=10.38; P=.001; η^2 partial=0.05), indicating that health-related QoL scores were significantly higher in the EmotionRegulation group posttreatment than in the TAU group.

Follow-Up Effects

Principal Outcome Measures

For depression, a significant condition×time interaction effect was found ($F_{2,396}$=6.18; P=.01; η^2 partial=0.03). There was a significant time effect ($F_{2,396}$=102.07; P<.001; η^2 partial=0.34) and a nonsignificant condition effect on anxiety scores ($F_{1,198}$=2.25; P=.14; η^2 partial=0.01). In the EmotionRegulation condition, Bonferroni tests indicated that the differences between pre- and posttreatment were significant (P<.001), but the differences between posttreatment and follow-up were not significant (P>.99), revealing that the reductions in depression scores were maintained at the 3-month follow-up.

The analyses showed no condition×time interaction effect on anxiety ($F_{1.62,319.88}$=0.75; P=.45; η^2 partial=0.004). However, there was a significant time effect ($F_{1.62,319.88}$=29.35; P<.001; η^2 partial=0.13) and a significant condition effect on anxiety scores ($F_{1,198}$=4.20; P=.04; η^2 partial=0.02). Although there was no interaction effect, we decided to perform posthoc tests to preliminarily explore the direction of the changes. In the EmotionRegulation group, posthoc comparison tests revealed significant differences between pre- and posttreatment (P<.001), but no significant differences between posttreatment and follow-up in the EmotionRegulation group (P>.99).

Secondary Outcome Measures

Regarding temperament measures, no significant condition×time interaction effect was found for the BIS subscale ($F_{1.49,295.29}$=0.93; P=.34; η^2 partial=0.005). A significant time effect was found ($F_{1.49,295.29}$=15.33; P<.001; η^2 partial=0.07), but the effect of condition was not significant ($F_{1,198}$=2.51; P=.12; η^2 partial=0.01).

For the BAS subscale, a significant condition×effect interaction effect was found ($F_{1.84,363.40}$=5.62; P=.02; η^2 partial=0.03). In the EmotionRegulation group, posthoc tests showed that BAS scores were significantly higher at posttreatment than at baseline (P=.05). However, these differences vanished at follow-up, as shown by the comparison between pretreatment and follow-up scores (P>.99).

Regarding health-related QoL, the analyses did not reveal a condition×time interaction effect ($F_{1.7,336.93}$=2.73; P=.08; η^2partial=0.01). However, there was a significant time effect ($F_{2,396}$=23.34; P=.001; η^2 partial=0.11) and a significant condition effect ($F_{1,198}$=6.3; P=.01; η^2 partial=0.03), which indicated a significant improvement in health-related QoL at posttreatment in both EmotionRegulation and TAU and

XSL•FO
RenderX

generally higher health-related QoL in the EmotionRegulation group. Table 4 provides the effect sizes for the within- and between-group comparisons.

Table 4. Within- and between-group effect sizes and 95% CIs.

Instrument	EmotionRegulation (n=99), *d* (95% CI)		TAU[a] (n=101), *d* (95% CI)		EmotionRegulation versus TAU, *d* (95% CI)	
	Pre-post	Pre-F/U[b]	Pre-post	Pre-F/U	Posttreatment	F/U
Beck depression inventory-II	0.72[c] (0.54 to 0.90)	0.70 (0.54 to 0.87)	0.36 (0.23 to 0.49)	0.52 (0.39 to 0.66)	0.41 (0.13 to 0.69)	0.24 (−0.04 to 0.52)
Beck anxiety inventory	0.41 (0.25 to 0.57)	0.38 (0.20 to 0.56)	0.26 (0.09 to 0.43)	0.32 (0.13 to 0.51)	0.35 (0.07 to 0.63)	0.25 (−0.03 to 0.53)
Behavioral inhibition scale and behavioral activation scale—behavioral inhibition system	0.37 (0.10 to 0.63)	0.54 (0.26 to 0.82)	0.18 (−0.07 to 0.44)	0.33 (0.06 to 0.61)	0.22 (−0.06 to 0.50)	0.25 (−0.03 to 0.52)
Behavioral inhibition scale and behavioral activation scale—behavioral activation system	−0.17[c] (−0.31 to −0.03)	0.05 (−0.09 to 0.20)	0.15 (0.01 to 0.29)	0.31 (0.15 to 0.48)	−0.22 (−0.50 to 0.06)	−0.15 (−0.43 to 0.12)
EuroQoL-5D-3L questionnaire	−0.57 (−0.76 to −0.37)	−0.43 (−0.64 to −0.22)	−0.24 (−0.44 to −0.04)	−0.23 (−0.44 to −0.02)	−0.45 (−0.74 to −0.17)	−0.33 (−0.60 to −0.05)

[a]TAU: treatment as usual.

[b]F/U: 3-month follow-up.

[c]Positive effect sizes denote a decrease in scores, whereas negative effect sizes denote an increase.

Significance of Clinical Improvements

The results for the significance of the clinical changes on measures of overall depression and anxiety (BDI-II and BAI, respectively) are summarized below.

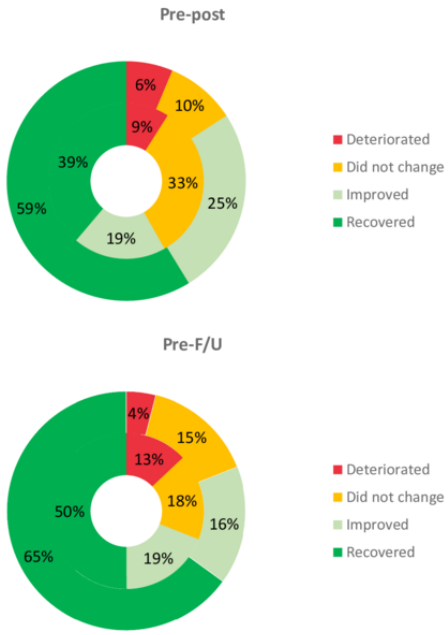
Changes in Depression

Baseline to Posttreatment Changes

In the EmotionRegulation group, 84% (85/99) of the patients achieved a functional change in their depression scores, whereas only 58.4% (59/101) did so in the TAU group, and these differences were significant (X^2_1=10.5; *P*=.001). In the EmotionRegulation group, 59% (58/99) of participants recovered, 25% (25/99) improved, 10% (10/99) did not change, and 6% (6/99) deteriorated; whereas in the TAU group, 38.6% (39/101) of participants recovered, 18.8% (19/101) improved, 32.7% (33/101) did not change, and 8.9% (9/101) deteriorated (Figure 2). Differences betweeen groups were significant (X^2_3=11.7; *P*=.009).

XSL•FO
RenderX

Figure 2. Percentages of participants recovered, improved, did not change, and deteriorated on depression scores (Beck depression inventory-II) in EmotionRegulation (outer circle) and treatment as usual (inner circle). F/U: follow-up.



Baseline to Follow-Up Changes

In the EmotionRegulation group, 80% (79/99) of the patients achieved a functional change in their depression scores, whereas only 70.3% (71/101) of the patients did so in the TAU group, and these differences were not significant (X^2_1=1.6; *P*=.20). In the EmotionRegulation group, 65% (64/99) of participants recovered, 16% (16/99) improved, 16% (16/99) did not change, and 4% (4/99) deteriorated; whereas in the TAU group, 50.5% (51/101) of participants recovered, 19.8% (20/101) improved, 17.8% (18/101) did not change, and 12.9% (13/101) had deteriorated (Figure 2). No significant differences betweeen groups were found (X^2_3=3.7; *P*=.30).

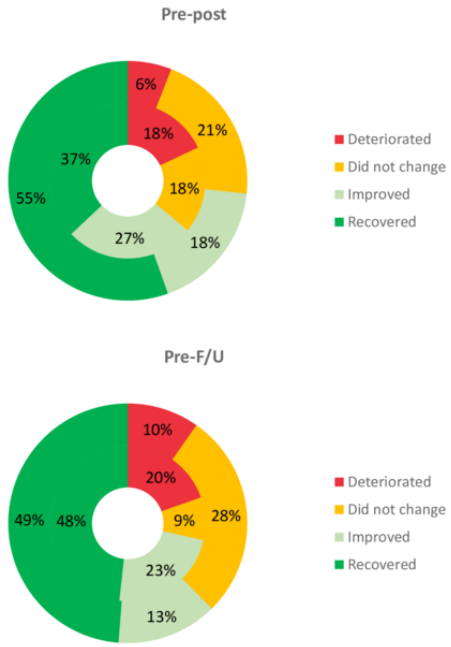
Changes in Anxiety

Baseline to Posttreatment Changes

There were no significant differences in the proportion of patients who achieved a functional change in their anxiety scores between the 2 conditions (X^2_1=1.2; *P*=.28). However, 73% (72/99) of participants in the EmotionRegulation group achieved a functional change in anxiety and 64.4% (65/101) did so in the TAU group. . In the EmotionRegulation group, 56% (55/99) of participants recovered, 18% (18/99) improved, 21% (21/99) did not change, and 6% (6/99) deteriorated. In contrast, in the TAU group, 36.6% (37/101) of participants were recovered, 26.7% (27/101) were improved, 17.8% (18/101) did not change, and 17.8% (18/101) had deteriorated (Figure 3). Differences betweeen groups were marginally significant (X^2_3=7.3; *P*=.06).

XSL•FO
RenderX

Figure 3. Percentages of participants recovered, improved, did not change, and deteriorated on anxiety scores (Beck anxiety inventory) in EmotionRegulation (outer circle) and treatment as usual (inner circle). F/U: follow-up.



Baseline to Follow-Up Changes

There were no significant differences in the proportion of patients who achieved a functional change in their anxiety scores between the 2 conditions (X^2_1=0.9; *P*=.34). In the EmotionRegulation group, 49% (49/99) of participants recovered, 14% (14/99) improved, 28% (28/99) did not change, and 10% (10/99) deteriorated. In contrast, in the TAU group, 47.5% (48/101) of participants were recovered, 22.8% (23/101) were improved, 8.9% (9/101) did not change, and 19.8% (20/101) had deteriorated (Figure 3). Differences betweeen groups were significant (X^2_3=8.2; *P*=.04).

Treatment Acceptability

Table 5 provides the means and SDs for expectations and opinions about treatment in the completer sample of the EmotionRegulation condition (n=63). As hypothesized, the results indicate that participants reported high scores on all the items measuring treatment expectations (scores between 7.56 and 7.81): logic of the treatment, satisfaction with the treatment, recommending the treatment to other people with similar problems, usefulness of the treatment for other psychological problems, and usefulness of the treatment for one's specific problem. After receiving the intervention, scores for treatment opinions were generally higher than scores for treatment expectations (scores between 7.67 and 8.24).

A 1-way repeated measures ANOVA showed that, compared with treatment expectations, the opinion of treatment was significantly better for item 1 (logic of the treatment; $F_{1,62}$=7.81; *P*=.007; η^2 partial=0.11), item 3 (recommending the treatment to other people with similar problems; $F_{1,62}$=4.80; *P*=.03; η^2 partial=0.07), and item 4 (usefulness of the treatment for other psychological problems; $F_{1,62}$=4.92; *P*=.30; η^2 partial=0.07). No significant differences were found for item 2 (satisfaction with the treatment; $F_{1,62}$=3.05; *P*=.09; η^2 partial=0.05) or item 5 (usefulness of the treatment for one's specific problem; $F_{1,62}$=.21; *P*=.65; η^2partial=0.003).

Table 5. Means and SDs for expectations and opinions of treatment (n=63).

Item	Expectations, mean (SD)	Opinion, mean (SD)
Treatment is logical	7.65 (1.88)	8.19 (1.62)
Satisfaction with the treatment	7.56 (1.81)	7.90 (1.71)
Recommend to others	7.81 (1.91)	8.24 (1.85)
Usefulness for other psychological problems	7.64 (1.86)	8.05 (1.65)
Usefulness for one's specific problems	7.76 (1.83)	7.67 (2.13)

XSL•FO
RenderX

Discussion

Principal Findings

The objective of this RCT was to explore whether a transdiagnostic internet-delivered protocol (EmotionRegulation) could be effective in treating a wide range of anxiety and depressive disorders, compared with TAU as provided in Spanish public specialized mental health care services. The effectiveness of EmotionRegulation was evaluated on measures of overall anxiety and depression, temperament (ie, BI and BA), and health-related QoL. Finally, expectations and opinions were evaluated to examine the acceptability of EmotionRegulation for the patients. To our knowledge, this is the first RCT to report data on a transdiagnostic internet-delivered protocol for ED in public specialized mental health care.

Regarding our first hypothesis, the ANCOVAs revealed that participants in the intervention group (EmotionRegulation) improved their depression and anxiety symptoms at posttreatment to a greater degree than participants in the TAU group. Regarding the magnitude of the changes at posttreatment, the analyses showed small but significant between-group effect sizes for both depression and anxiety. With regard to measures of temperament, the findings were mixed. Overall, although patients in the EmotionRegulation group showed better scores on the 2 subscales of the BIS and BAS than the TAU group, only the differences on the BAS subscale were significant, favoring the EmotionRegulation group. These results suggest that the intervention tested in this study was able to modify temperament, in line with a mechanistically transdiagnostic approach that assumes the existence of underlying mechanisms that account for the occurrence of specific symptoms. To our knowledge, only 1 study has previously investigated the effect of a mechanistically transdiagnostic treatment on the dimensions of BI and BA [42]. Similar to the results obtained in our study, in the study by Carl et al [42], both BI and BA improved following treatment with the UP compared with a waitlist control group, with small between-group effect sizes found for both the BIS and BAS temperament dimensions. Moreover, these authors showed associations between the gains in temperament dimensions and symptoms of anxiety and depression, particularly for BI, because lower BI scores were associated with greater improvements in anxiety and depressive symptoms. However, a limitation of this study was that the sample size was small, and so the findings were preliminary. Thus, future RCTs of mechanistically transdiagnostic treatments should analyze the extent to which these interventions are able to modify BI and BA and other related temperament dimensions such as negative and positive affect to shed more light on this question [104,105]. Finally, as hypothesized, EmotionRegulation was superior to TAU in improving health-related QoL, with a small but significant between-group effect size observed for the EQ-5D-3L scale. In this regard, it is known that pharmacological treatment could be linked to some variables associated with health-related QoL, such as fatigue and sexual functioning [106]. Therefore, decreases in medication may have indirectly influenced the QoL in some patients. However, these decreases were not measured, and hence, no analyses could be performed to analyze this aspect.

Furthermore, to evaluate the significance of the clinical gains, Jacobson and Truax's RCI was obtained for the principal outcome measures, that is, BDI-II and BAI.

Consistent with the aforementioned results, significant differences were observed in the proportion of patients who achieved a functional change in depression scores, with a significantly higher number of patients within the functional range in the EmotionRegulation condition (85/99, 84%) than in the TAU group (59/101, 58.4%). Regarding anxiety scores, although no significant differences were observed between the 2 groups in the proportion of patients reaching a functional change, a higher proportion of patients in the EmotionRegulation group (72/99, 73%) than those in the TAU group (65/101, 64.4%) reached a functional change. Furthermore, significant differences were found in the proportion of patients recovered, improved, did not change, and deteriorated, with better general results for the intervention group than for the TAU group. At follow-up, there were no differences in the proportion of patients reaching a functional change in either depression or anxiety. Therefore, it can be concluded that these differences tended to diminish between groups at follow-up.

Regarding the acceptability of EmotionRegulation, expectations about the treatment were high (scores ≥7). After receiving the treatment, compared with expectations, the participants rated the program as significantly more logical, more recommendable for other people with similar problems and more useful for the treatment of other psychological problems. The study of acceptability is important because expectations about treatment have been shown to affect treatment outcomes [107]. Furthermore, because most transdiagnostic internet-delivered protocols have been conducted in community samples [108], it is necessary to continue to explore the acceptability of these interventions in specialized care.

Overall, the findings showed that the transdiagnostic internet-delivered protocol tested in this RCT was more effective than TAU for the treatment of anxiety and depressive disorders in public specialized mental health care. On the one hand, the results show that EmotionRegulation led to greater improvements at posttreatment, and these gains were maintained at follow-up. Regarding TAU, the results reveal that patients undergoing TAU also experienced improvement over time, but it was slower and less pronounced than in patients receiving EmotionRegulation. The generally lower intensity of TAU (ie, lower frequency of therapy sessions) and the fact that for most patients TAU was limited to pharmacotherapy (with no access to psychological treatment) might partly account for these results. On the other hand, as anticipated, and in line with previous studies conducted by our research group using the same treatment platform [94,109], scores on expectations and opinions demonstrated EmotionRegulation's acceptability for participants.

The results obtained in this RCT have implications for both research and clinical practice, especially in the context of public specialized mental health care. First, the findings obtained in this study support the effectiveness of a mechanistically transdiagnostic internet-delivered protocol for the treatment of EDs, and they contribute to the literature in this particular field

XSL•FO
RenderX

[23,27,36,110-112]. Specifically, a combination of the components of the UP and DBT regulation skills was found to be more effective than TAU in treating ED. Second, as far as we know, this is the first study to explore a transdiagnostic internet-delivered treatment in public specialized mental health care. As mentioned earlier, most research on transdiagnostic web-based treatments has been conducted in community settings, with a few of these studies carried out in primary care [108]. The results showed that the intervention was found to be more effective than TAU on the measures of generic depression, anxiety, and QoL. These results are consistent with the literature showing the superiority of CBT over TAU. For instance, a meta-analysis showed that CBT outweighed TAU, with effect sizes in the medium range on measures of generic anxiety (Hedges g=0.70) and depression (Hedges g=0.69) [70]. As surprising as these data might seem, the truth is that current public mental health services still have to deal with a number of barriers that hinder appropriate care delivery, such as excessive waiting times to access mental health care [60], low frequency of sessions [113], or inadequate follow-up care [56]. Moreover, the lack of training in evidence-based treatments among professionals further adds to this problem [114]. Finally, in Spain, several RCTs have been conducted using the internet to provide evidence-based treatments, showing that they are effective for the treatment of ED and, in particular, depression, in community samples [94] and primary care [115,116], and others are underway [117]. This study demonstrated that an internet-delivered protocol for ED was effective in public specialized care, a setting with a high demand, but much less explored, thus adding to the literature on these treatments for ED. Furthermore, the use of TAU as the control condition may help to answer the question of "whether a new treatment or an evidence-based psychotherapy really surpasses in outcome effects what is ordinarily done at a given clinic" [118], thus helping to make clinicians, researchers, and policy makers aware of the limitations and aspects that should be improved in this specific setting.

Limitations

Although the results of this RCT are promising, they should be interpreted in light of the following limitations. First, although several measures were taken to minimize attrition (eg, guidance was provided to all the patients participating in EmotionRegulation), the number of patients who dropped out of the study was high (around 35%). However, this proportion was close to what is typically observed in the literature on internet-delivered psychological treatments (ie, approximately 30%-35%) [71]. Moreover, attrition in the TAU condition was similar (34/108, 31.5%). Second, this RCT was not powered to detect differences in disorder-specific measures (ie, for GAD, PD or AG, SAD, and OCD). Therefore, future studies that meet the minimum levels of statistical power to detect differences in these measures are warranted. Third, although the frequency of the session in TAU was low and most patients in this condition were receiving only pharmacotherapy (as observed by the researchers), these data were not monitored during the trial. Finally, the results for acceptability (ie, expectations and opinions) might not be entirely representative because data from patients who dropped out of the intervention were not included in these analyses.

Conclusions and Future Directions

The effectiveness of a mechanistically transdiagnostic internet-delivered protocol for ED was compared with TAU in public specialized mental health care. Although the results are promising, more research in specialized care should be conducted to extend the findings obtained in this study. First, research on predictors and moderators of treatment outcomes and dropout in this specific setting can help to delineate the profiles of participants who are more likely to benefit from these treatments and help to answer the classic question, "what treatment, by whom, is most effective for this individual with that specific problem, under which set of circumstances?" [119]. Moreover, we believe that the integration of principles from different evidence-based protocols (eg, components of the UP and DBT) can be a powerful strategy that should guide future research on evidence-based psychotherapy. This strategy is in line with process-based CBT and the new foundational question proposed by Hofmann and Hayes [49] (ie, "What core biopsychosocial processes should be targeted with this client given this goal in this situation, and how can they most efficiently and effectively be changed?") . Second, despite the huge advances experienced by the field of internet-delivered treatments in the past two decades, high dropout rates remain a major challenge in the field. To continue to improve current and future internet-delivered interventions, future studies should strive to include dropouts in the analysis of acceptability using both qualitative and quantitative approaches. Moreover, to ensure the integrity, quality, and replicability of these studies, adherence to existing research guidelines is of paramount importance in this endeavor. Third, although we did not assess the acceptability of clinicians involved in the RCT (ie, psychiatrists, clinical psychologists, and nurses), it is worth mentioning that some of them refused to participate in the recruitment process, which might reflect negative attitudes toward internet-delivered interventions among these professionals. In this scenario, research efforts should be made to inform clinicians and state holders about the benefits of internet-delivered treatments, especially because they are seen as authority figures and, therefore, their attitudes can have a major impact on patients' perceptions. Finally, although the need for efficacy studies is out of doubt, we believe that it is of paramount importance to conduct more implementation research [120]. These studies may provide a much deeper understanding of implementation variables that can either facilitate or hamper the effective uptake of evidence-based protocols in real clinical practice, such as the attitudes of clinicians and other professionals toward internet-delivered treatments, or economical and logistic aspects that are difficult to implement.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Consorcio Hospitalario Provincial de Castellón, Hospital Universitario de la Ribera, and Hospital Universitario Vall d'Hebrón for their invaluable contribution to this study. This research was conducted with the financial support of grants P1-1B2014-43 (Universitat Jaume I, Castellón), a PhD grant from the Ministry of Education, Culture and Sports (FPU13/00576), and CIBER Fisiopatología de la Obesidad y Nutrición-ISCIII CB06/03/0052. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics and effect sizes for disorder-specific measures.
[DOCX File , 14 KB - jmir_v22i7e18220_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 1256 KB - jmir_v22i7e18220_app2.pdf]

References

1. Bullis J, Boettcher H, Sauer‐Zavala S, Farchione T, Barlow D. What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention. Clin Psychol Sci Pract 2019 Mar 14;26(2):e12278 [FREE Full text] [doi: 10.1111/cpsp.12278]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
4. Andlin-Sobocki P, Wittchen H. Cost of anxiety disorders in Europe. Eur J Neurol 2005 Jun;12(Suppl 1):39-44. [doi: 10.1111/j.1468-1331.2005.01196.x] [Medline: 15877777]
5. Cuijpers P, Beekman AT, Reynolds CF. Preventing depression: a global priority. J Am Med Assoc 2012 Mar 14;307(10):1033-1034 [FREE Full text] [doi: 10.1001/jama.2012.271] [Medline: 22416097]
6. Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. The global burden of anxiety disorders in 2010. Psychol Med 2014 Aug;44(11):2363-2374. [doi: 10.1017/S0033291713003243] [Medline: 24451993]
7. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
8. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]
9. Hollon SD, Ponniah K. A review of empirically supported psychological therapies for mood disorders in adults. Depress Anxiety 2010 Oct;27(10):891-932 [FREE Full text] [doi: 10.1002/da.20741] [Medline: 20830696]
10. Borkovec T, Ruscio A. Psychotherapy for generalized anxiety disorder. J Clin Psychiatry 2001;62(Suppl 11):37-42; discussion 43 [FREE Full text] [Medline: 11414549]
11. Covin R, Ouimet AJ, Seeds PM, Dozois DJ. A meta-analysis of CBT for pathological worry among clients with GAD. J Anxiety Disord 2008;22(1):108-116. [doi: 10.1016/j.janxdis.2007.01.002] [Medline: 17321717]
12. Mitte K. A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. J Affect Disord 2005 Sep;88(1):27-45. [doi: 10.1016/j.jad.2005.05.003] [Medline: 16005982]
13. Mayo-Wilson E, Dias S, Mavranezouli I, Kew K, Clark DM, Ades AE, et al. Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. Lancet Psychiatry 2014 Oct;1(5):368-376 [FREE Full text] [doi: 10.1016/S2215-0366(14)70329-3] [Medline: 26361000]
14. Öst L, Havnen A, Hansen B, Kvale G. Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. Clin Psychol Rev 2015 Aug;40:156-169. [doi: 10.1016/j.cpr.2015.06.003] [Medline: 26117062]
15. Brown TA, Campbell LA, Lehman CL, Grisham JR, Mancill RB. Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. J Abnorm Psychol 2001 Nov;110(4):585-599. [doi: 10.1037//0021-843x.110.4.585] [Medline: 11727948]

XSL•FO
RenderX

16. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
17. Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders - republished article. Behav Ther 2016 Nov;47(6):838-853. [doi: 10.1016/j.beth.2016.11.005] [Medline: 27993336]
18. McManus F, Shafran R, Cooper Z. What does a transdiagnostic approach have to offer the treatment of anxiety disorders? Br J Clin Psychol 2010 Nov;49(Pt 4):491-505. [doi: 10.1348/014466509X476567] [Medline: 19878622]
19. Hofmeijer-Sevink MK, Batelaan NM, van Megen HJ, Penninx BW, Cath DC, van den Hout MA, et al. Clinical relevance of comorbidity in anxiety disorders: a report from the Netherlands study of depression and anxiety (NESDA). J Affect Disord 2012 Mar;137(1-3):106-112 [FREE Full text] [doi: 10.1016/j.jad.2011.12.008] [Medline: 22240085]
20. Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. Am J Psychiatry 2005 Jun;162(6):1171-1178. [doi: 10.1176/appi.ajp.162.6.1171] [Medline: 15930066]
21. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
22. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
23. Farchione TJ, Fairholme CP, Ellard KK, Boisseau CL, Thompson-Hollands J, Carl JR, et al. Unified protocol for transdiagnostic treatment of emotional disorders: a randomized controlled trial. Behav Ther 2012 Sep;43(3):666-678 [FREE Full text] [doi: 10.1016/j.beth.2012.01.001] [Medline: 22697453]
24. Norton PJ, Hope DA. Preliminary evaluation of a broad-spectrum cognitive-behavioral group therapy for anxiety. J Behav Ther Exp Psychiatry 2005 Jun;36(2):79-97. [doi: 10.1016/j.jbtep.2004.07.002] [Medline: 15814078]
25. Forman EM, Herbert JD, Moitra E, Yeomans PD, Geller PA. A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. Behav Modif 2007 Nov;31(6):772-799. [doi: 10.1177/0145445507302202] [Medline: 17932235]
26. Johansson R, Björklund M, Hornborg C, Karlsson S, Hesser H, Ljótsson B, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the internet: a randomized controlled trial. PeerJ 2013;1:e102 [FREE Full text] [doi: 10.7717/peerj.102] [Medline: 23862104]
27. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
28. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
29. Newby JM, Twomey C, Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
30. Păsărelu CR, Andersson G, Nordgren LB, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
31. Reinholt N, Krogh J. Efficacy of transdiagnostic cognitive behaviour therapy for anxiety disorders: a systematic review and meta-analysis of published outcome studies. Cogn Behav Ther 2014;43(3):171-184. [doi: 10.1080/16506073.2014.897367] [Medline: 24646219]
32. Pearl SB, Norton PJ. Transdiagnostic versus diagnosis specific cognitive behavioural therapies for anxiety: a meta-analysis. J Anxiety Disord 2017 Mar;46:11-24. [doi: 10.1016/j.janxdis.2016.07.004] [Medline: 27466074]
33. García-Escalera J, Chorot P, Valiente R. Efficacy of transdiagnostic cognitive-behavioral therapy for anxiety and depression in adults, children and adolescents: a meta-analysis. Rev Psicopatol Psicol Clin 2016;21(3):147-175 [FREE Full text] [doi: 10.5944/rppc.vol.21.num.3.2016.17811]
34. Sauer-Zavala S, Gutner CA, Farchione TJ, Boettcher HT, Bullis JR, Barlow DH. Current definitions of 'transdiagnostic' in treatment development: a search for consensus. Behav Ther 2017 Jan;48(1):128-138. [doi: 10.1016/j.beth.2016.09.004] [Medline: 28077216]
35. Barlow D, Ellard K, Fairholme C, Farchione TJ, Boisseau CL. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Workbook. New York, USA: Oxford University Press; 2011.
36. Barlow D, Farchione T, Fairholme C, Ellard KK, Boisseau CL. Treatments that work. In: Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide. New York, USA: Oxford University Press; 2011.
37. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a

XSL•FO
RenderX

randomized clinical trial. JAMA Psychiatry 2017 Sep 1;74(9):875-884 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
38. Steele SJ, Farchione TJ, Cassiello-Robbins C, Ametaj A, Sbi S, Sauer-Zavala S, et al. Efficacy of the Unified Protocol for transdiagnostic treatment of comorbid psychopathology accompanying emotional disorders compared to treatments targeting single disorders. J Psychiatr Res 2018 Sep;104:211-216 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.08.005] [Medline: 30103069]
39. Brown TA, Barlow DH. A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: implications for assessment and treatment. Psychol Assess 2009 Sep;21(3):256-271 [FREE Full text] [doi: 10.1037/a0016608] [Medline: 19719339]
40. Barlow DH, Sauer-Zavala S, Carl JR, Bullis JR, Ellard KK. The nature, diagnosis, and treatment of neuroticism. Clin Psychol Sci 2013 Oct 14;2(3):344-365. [doi: 10.1177/2167702613505532]
41. Bullis JR, Fortune MR, Farchione TJ, Barlow DH. A preliminary investigation of the long-term outcome of the unified protocol for transdiagnostic treatment of emotional disorders. Compr Psychiatry 2014 Nov;55(8):1920-1927 [FREE Full text] [doi: 10.1016/j.comppsych.2014.07.016] [Medline: 25113056]
42. Carl JR, Gallagher MW, Sauer-Zavala SE, Bentley KH, Barlow DH. A preliminary investigation of the effects of the unified protocol on temperament. Compr Psychiatry 2014 Aug;55(6):1426-1434 [FREE Full text] [doi: 10.1016/j.comppsych.2014.04.015] [Medline: 24933653]
43. Gray J. The Psychology of Fear and Stress. Cambridge, UK: Cambridge University; 1987.
44. Neacsiu A, Bohus M, Linehan M. Dialectical behavior therapy: an intervention for emotion dysregulation. In: Gross JJ, editor. Handbook of Emotion Regulation. New York, USA: Guilford Press; 2014:491-508.
45. Neacsiu AD, Herr NR, Fang CM, Rodriguez MA, Rosenthal MZ. Identity disturbance and problems with emotion regulation are related constructs across diagnoses. J Clin Psychol 2015 Apr;71(4):346-361. [doi: 10.1002/jclp.22141] [Medline: 25534425]
46. Linehan M. Skills Training Manual for Treating Borderline Personality Disorder. New York, USA: Guilford Press; 1993.
47. Ritschel LA, Lim NE, Stewart LM. Transdiagnostic applications of DBT for adolescents and adults. Am J Psychother 2015;69(2):111-128. [doi: 10.1176/appi.psychotherapy.2015.69.2.111] [Medline: 26160618]
48. Bohus M, Dyer AS, Priebe K, Krüger A, Kleindienst N, Schmahl C, et al. Dialectical behaviour therapy for post-traumatic stress disorder after childhood sexual abuse in patients with and without borderline personality disorder: a randomised controlled trial. Psychother Psychosom 2013;82(4):221-233. [doi: 10.1159/000348451] [Medline: 23712109]
49. Hofmann SG, Hayes SC. The future of intervention science: process-based therapy. Clin Psychol Sci 2019 Jan;7(1):37-50 [FREE Full text] [doi: 10.1177/2167702618772296] [Medline: 30713811]
50. Andersson G. Internet interventions: past, present and future. Internet Interv 2018 Jun;12:181-188 [FREE Full text] [doi: 10.1016/j.invent.2018.03.008] [Medline: 30135782]
51. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
52. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 7;192(S11):S4-11. [Medline: 20528707]
53. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
54. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
55. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
56. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850 [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
57. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
58. Montilla JF, González C, Retolaza A, Dueñas C, Alameda J. Uso de servicios ambulatorios de salud mental en España: consumo de recursos en el primer año de asistencia a pacientes nuevos. Rev Asoc Esp Neuropsiq 2002 Dec(84):25-47. [doi: 10.4321/s0211-57352002000400003]
59. Duro Martínez JC. La psicología clínica en la sanidad pública de la comunidad de Madrid: la insuficiente consolidación de una profesión sanitaria. Clínica y Salud 2016 Jul 1;27(2):45-50 [FREE Full text] [doi: 10.1016/j.clysa.2016.05.001]
60. Martín-Jurado A, de la Gándara Martín JJ, Castro Carbajo S, Moreira Hernández A, Sánchez-Hernández J. [Concordance analysis of referrals from primary care to mental health]. Semergen 2012 Sep;38(6):354-359. [doi: 10.1016/j.semerg.2011.12.005] [Medline: 22935831]

XSL•FO
RenderX

61. McHugh RK, Murray HW, Barlow DH. Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: the promise of transdiagnostic interventions. Behav Res Ther 2009 Nov;47(11):946-953 [FREE Full text] [doi: 10.1016/j.brat.2009.07.005] [Medline: 19643395]
62. Spurgeon JA, Wright JH. Computer-assisted cognitive-behavioral therapy. Curr Psychiatry Rep 2010 Dec;12(6):547-552. [doi: 10.1007/s11920-010-0152-4] [Medline: 20872100]
63. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
64. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
65. Moher D, Schulz KF, Altman DG, CONSORT. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 2001;1:2 [FREE Full text] [doi: 10.1186/1471-2288-1-2] [Medline: 11336663]
66. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
67. González-Robles A, García-Palacios A, Baños R, Riera A, Llorca G, Traver F, et al. Effectiveness of a transdiagnostic internet-based protocol for the treatment of emotional disorders versus treatment as usual in specialized care: study protocol for a randomized controlled trial. Trials 2015 Oct 31;16:488 [FREE Full text] [doi: 10.1186/s13063-015-1024-3] [Medline: 26519046]
68. Labpsitec. 2011. Psicología y Tecnología URL: https://psicologiaytecnologia.labpsitec.es/ [accessed 2015-09-10]
69. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
70. Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. J Affect Disord 2015 Apr 1;175:152-167. [doi: 10.1016/j.jad.2014.12.025] [Medline: 25618002]
71. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
72. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
73. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
74. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. Washington, DC: American Psychiatric Association; 2000.
75. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The mini-international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34 [FREE Full text] [Medline: 9881538]
76. Ferrando L, Bobes J, Gibert J, Lecubrier Y. Mini international neuropsychiatric interview. In: Bobes J, Bousoño M, González MP, editors. Manejo De Los Trastornos Mentales Y Del Comportamiento en Atención Primaria. Oviedo: Gofer; 1997.
77. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1990.
78. Sanz J, García-Vera MP. Adaptación española del Inventario para la Depresión de Beck II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. Clin Salud 2005;16(2):121-142 [FREE Full text] [doi: 10.6018/analesps.29.1.130532]
79. Beck A, Steer R. Beck Anxiety Inventory Manual. San Antonio, TX: Psychological Corporation; 1993.
80. Magán I, Sanz J, García-Vera MP. Psychometric properties of a Spanish version of the beck anxiety inventory (BAI) in general population. Span J Psychol 2008 Nov;11(2):626-640. [Medline: 18988448]
81. Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol 1994;67(2):319-333. [doi: 10.1037/0022-3514.67.2.319]
82. Caseras X, Àvila C, Torrubia R. The measurement of individual differences in behavioural inhibition and behavioural activation systems: a comparison of personality scales. Pers Individ Dif 2003 Apr;34(6):999-1013. [doi: 10.1016/s0191-8869(02)00084-3]
83. Badía X. EuroQol; un instrumento para valorar la salud EQ-5D guía del usuario, versión española. Med Clin 1999;114:6-14.
84. Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European quality of life scale]. Med Clin (Barc) 1999;112(Suppl 1):79-85. [Medline: 10618804]
85. Meyer T, Miller M, Metzger R, Borkovec TD. Development and validation of the Penn state worry questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]

XSL•FO
RenderX

86. Sandín B, Chorot P, Valiente R, Lostao L. Validación española del cuestionario de preocupación PSWQ: estructura factorial y propiedades psicométricas. RPPC 2009 Aug 1;14(2):- [FREE Full text] [doi: 10.5944/rppc.vol.14.num.2.2009.4070]
87. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety 2002;15(4):183-185. [doi: 10.1002/da.10049] [Medline: 12112724]
88. Santacana M, Fullana MA, Bonillo A, Morales M, Montoro M, Rosado S, et al. Psychometric properties of the Spanish self-report version of the panic disorder severity scale. Compr Psychiatry 2014 Aug;55(6):1467-1472. [doi: 10.1016/j.comppsych.2014.04.007] [Medline: 24850072]
89. Mattick RP, Clarke J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]
90. Zubeidat I, Salinas JM, Sierra JC, Fernández-Parra A. Psychometric properties of the social interaction anxiety scale and separation criterion between Spanish youths with and without subtypes of social anxiety. J Anxiety Disord 2007;21(5):603-624. [doi: 10.1016/j.janxdis.2006.09.010] [Medline: 17084589]
91. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The obsessive-compulsive inventory: development and validation of a short version. Psychol Assess 2002 Dec;14(4):485-496. [Medline: 12501574]
92. Fullana M, Tortella-Feliu M, Caseras X, Andión O, Torrubia R, Mataix-Cols D. Psychometric properties of the Spanish version of the obsessive-compulsive inventory--revised in a non-clinical sample. J Anxiety Disord 2005;19(8):893-903. [doi: 10.1016/j.janxdis.2004.10.004] [Medline: 16243637]
93. Borkovec TD, Nau SD. Credibility of analogue therapy rationales. J Behav Ther Exp Psychiatry 1972 Dec;3(4):257-260. [doi: 10.1016/0005-7916(72)90045-6]
94. Mira A, Bretón-López J, García-Palacios A, Quero S, Baños RM, Botella C. An internet-based program for depressive symptoms using human and automated support: a randomized controlled trial. Neuropsychiatr Dis Treat 2017;13:987-1006 [FREE Full text] [doi: 10.2147/NDT.S130994] [Medline: 28408833]
95. Campos D, Bretón-López J, Botella C, Mira A, Castilla D, Mor S, et al. Efficacy of an internet-based exposure treatment for flying phobia (NO-FEAR Airlines) with and without therapist guidance: a randomized controlled trial. BMC Psychiatry 2019 Mar 6;19(1):86 [FREE Full text] [doi: 10.1186/s12888-019-2060-4] [Medline: 30841930]
96. Commify Italia SRL. Trendoo URL: https://www.trendoo.es/ [accessed 2015-09-15]
97. Newman DA. Missing data. Organ Res Methods 2014 Sep 26;17(4):372-411. [doi: 10.1177/1094428114548590]
98. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: With Readings. Seventh Edition. River, NJ: Pearson-Prentice Hall International Edition; 2014.
99. Arias R, Chacon J, Castellanos M. Analisis De Datos en Psicologia Y Ciencias De La Salud. Vol 1. Madrid, Spain: EOS Instituto de Orientación Psicológica; 2015.
100. van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
101. Weinfurt K. Repeated measures analysis: ANOVA, MANOVA, and HLM. In: Grimm LG, Yarnold PR, editors. Reading and Understanding More Multivariate Statistics. New York, USA: American Psychological Association; 2000:317-361.
102. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Erlbaum; Dec 1988.
103. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
104. Dunn B, German R, Khazanov G, Xu C, Hollon SD, DeRubeis RJ. Changes in positive and negative affect during pharmacological treatment and cognitive therapy for major depressive disorder: a secondary analysis of two randomized controlled trials. Clin Psychol Sci 2019 Nov 7;8(1):36-51 [FREE Full text] [doi: 10.1177/2167702619863427]
105. González-Robles A, García-Palacios A, Baños R, Quero S, Botella C. Upregulating positive affectivity in the transdiagnostic treatment of emotional disorders: a randomized pilot study. Behav Modif 2019 Jan;43(1):26-55. [doi: 10.1177/0145445517735631] [Medline: 29047288]
106. Bet PM, Hugtenburg JG, Penninx BW, Hoogendijk WJ. Side effects of antidepressants during long-term use in a naturalistic setting. Eur Neuropsychopharmacol 2013 Nov;23(11):1443-1451. [doi: 10.1016/j.euroneuro.2013.05.001] [Medline: 23726508]
107. Greenberg RP, Constantino MJ, Bruce N. Are patient expectations still relevant for psychotherapy process and outcome? Clin Psychol Rev 2006 Oct;26(6):657-678. [doi: 10.1016/j.cpr.2005.03.002] [Medline: 15908088]
108. González-Robles A, Díaz-García A, Miguel C, García-Palacios A, Botella C. Comorbidity and diagnosis distribution in transdiagnostic treatments for emotional disorders: a systematic review of randomized controlled trials. PLoS One 2018;13(11):e0207396. [doi: 10.1371/journal.pone.0207396] [Medline: 30440020]
109. Campos D, Mira A, Bretón-López J, Castilla D, Botella C, Baños RM, et al. The acceptability of an internet-based exposure treatment for flying phobia with and without therapist guidance: patients' expectations, satisfaction, treatment preferences, and usability. Neuropsychiatr Dis Treat 2018;14:879-892 [FREE Full text] [doi: 10.2147/NDT.S153041] [Medline: 29636613]
110. Norton PJ. A randomized clinical trial of transdiagnostic cognitve-behavioral treatments for anxiety disorder by comparison to relaxation training. Behav Ther 2012 Sep;43(3):506-517 [FREE Full text] [doi: 10.1016/j.beth.2010.08.011] [Medline: 22697440]

XSL•FO
RenderX

111. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
112. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
113. González-González S, González M, García M. Psychological treatments features and outcomes in Spanish public mental health centres. Intern Jour Psych Psycho Therapy 2014;14(1):17-32 [FREE Full text]
114. Shafran R, Clark DM, Fairburn CG, Arntz A, Barlow DH, Ehlers A, et al. Mind the gap: improving the dissemination of CBT. Behav Res Ther 2009 Nov;47(11):902-909. [doi: 10.1016/j.brat.2009.07.003] [Medline: 19664756]
115. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
116. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, Campayo JG, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: Results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
117. Sandín B, Valiente RM, García-Escalera J, Pineda D, Espinosa V, Magaz A, et al. Protocolo unificado para el tratamiento transdiagnóstico de los trastornos emocionales en adolescentes a través de internet (iUP-A): aplicación web y protocolo de un ensayo controlado aleatorizado. Rev Psicopatol Psicol Clin 2020 Jan 29;24(3):197. [doi: 10.5944/rppc.26460]
118. Kazdin AE. Treatment as usual and routine care in research and clinical practice. Clin Psychol Rev 2015 Dec;42:168-178. [doi: 10.1016/j.cpr.2015.08.006] [Medline: 26431668]
119. Paul GL. Strategy of outcome research in psychotherapy. J Consult Psychol 1967 Apr;31(2):109-118. [doi: 10.1037/h0024436] [Medline: 5342732]
120. Theobald S, Brandes N, Gyapong M, El-Saharty S, Proctor E, Diaz T, et al. Implementation research: new imperatives and opportunities in global health. Lancet 2018 Nov 17;392(10160):2214-2228. [doi: 10.1016/S0140-6736(18)32205-0] [Medline: 30314860]

Abbreviations

AG: agoraphobia
ANCOVA: analysis of covariance
ANOVA: analysis of variance
BA: behavioral activation
BAI: Beck Anxiety Inventory
BAS: behavioral activation scale
BDI-II: Beck depression inventory, 2nd edition
BI: behavioral inhibition
BIS: behavioral inhibition scale
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DBT: dialectical behavioral therapy
DD: dysthymic disorder
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
ED: emotional disorder
EQ-5D-3L: EuroQoL-5D-3L questionnaire
GAD: generalized anxiety disorder
ITT: intention-to-treat
MCAR: missing completely at random
MDD: major depressive disorder
MINI: mini-international neuropsychiatric interview
ML: maximum likelihood
NOS: not otherwise specified
OCD: obsessive compulsive disorder
PD: panic disorder
QoL: quality of life
RCI: reliable change index
RCT: randomized controlled trial
SAD: social anxiety disorder

XSL•FO
RenderX

SI: societal index
TAU: treatment as usual
UP: Unified Protocol
VAS: visual analog scale

Edited by G Eysenbach; submitted 14.02.20; peer-reviewed by B Sandín, L Bücker; comments to author 24.03.20; revised version received 08.04.20; accepted 08.04.20; published 07.07.20.

Please cite as:
González-Robles A, Díaz-García A, García-Palacios A, Roca P, Ramos-Quiroga JA, Botella C
Effectiveness of a Transdiagnostic Guided Internet-Delivered Protocol for Emotional Disorders Versus Treatment as Usual in Specialized Care: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e18220
URL: https://www.jmir.org/2020/7/e18220
doi:10.2196/18220
PMID:32673226

©Alberto González-Robles, Amanda Díaz-García, Azucena García-Palacios, Pablo Roca, Josep Antoni Ramos-Quiroga, Cristina Botella. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Habits and Attitudes of Video Gaming and Information Technology Use in People with Schizophrenia: Cross-Sectional Survey

William TH Choi[1], RN, MSc; Dan KS Yu[2], RSW, MSW; Terry Wong[3], RSW, MSW; Tella Lantta[4], RN, MSc, PhD; Min Yang[5,6], MPH; Maritta Välimäki[1,4], PhD

[1]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong
[2]The Mental Health Association of Hong Kong, Hong Kong, Hong Kong
[3]New Life Psychiatric Rehabilitation Association, Hong Kong, Hong Kong
[4]Department of Nursing Science, Faculty of Medicine, University of Turku, Turku, Finland
[5]West China School of Public Health, Sichuan University Huaxi Medical Center, Sichuan, China
[6]Faculty of Health, Design and Art, Swinburne University of Technology, Melbourne, Australia

Corresponding Author:
Maritta Välimäki, PhD
Department of Nursing Science, Faculty of Medicine
University of Turku
ICT-City, A-wing, 5th floor
Joukahaisenkatu 3–5
Turku, 20520
Finland
Phone: 358 29 450 236
Email: maritta.valimaki@polyu.edu.hk

Abstract

Background: Information technology and video gaming have potential advantages in the treatment of schizophrenia. However, information regarding the habits and attitudes related to internet use and video gaming in people with schizophrenia is limited.

Objective: The aim of this study was to explore the habits and attitudes regarding video gaming and information technology usage and their associated factors in people with schizophrenia in Hong Kong.

Methods: In this cross-sectional survey, service users with schizophrenia were recruited from 6 halfway hostels and 7 integrated centers for mental wellness in Hong Kong. A 79-item self-report questionnaire was utilized to explore the habits of internet use and video gaming in these people with schizophrenia. The attitude toward video gaming was assessed using the Gaming Attitudes, Motivations, and Experiences Scales. Of the 148 individuals in a convenience sample who were invited to participate in this study, 110 willingly participated (a response rate of 74.3%). The data were analyzed using descriptive statistics, a two-tailed independent t test, Pearson correlation, and principal analysis with 3 methods of rotation (varimax, equimax, and promax).

Results: Most participants (100/110, 90.9%) had access to the internet and half of them (54/110, 49.1%) used the internet daily mostly to watch videos (66/110, 60.0%) or read news or books, etc (42/110, 38.2%). One-third of the participants (36/110, 32.7%) used the internet to play web-based games, and most of them (88/110, 80.0%) had played a video game in the past year. The most favorable gaming platforms were cellular phones (43/88, 49%) followed by computers (19/88, 22%) and arcade cabinets (6/88, 7%). The most favorable game genre was action games (34/145, 23.4%). Those who had a bachelor's degree or higher scored lower in social interaction than those with a lower education level (P=.03). Those who played video games daily scored higher in the category of story than those who did not play daily (t_{86}=2.03, P=.05). The most popular gaming category was autonomy and the least popular categories were violent catharsis and violent reward. Two motives, "social playing" and "evasive playing," were formed to describe the characteristics of playing video games.

Conclusions: Our data showed a high internet utilization rate among people with schizophrenia in Hong Kong. Only a few of them used the internet to search for health-related information. Our study also exemplified the unique habits of gaming among the participants. Health care professionals could utilize video games to engage people with schizophrenia and promote coping with stress and provide social skills training to such people with schizophrenia. Identification of the gaming attitudes can contribute to the development of serious games for the schizophrenic population. Further investigation is vital for the promotion of mental health through web-based platforms.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e14865) doi:10.2196/14865

KEYWORDS

video gaming; internet; information technology; schizophrenia

Introduction

Schizophrenia is a serious mental illness that is associated with significant cognition and functioning disabilities [1,2]. According to the World Health Organization [3], there are more than 21 million people diagnosed with schizophrenia worldwide. Many of them live in isolation [4,5] and are lonely without social interactions [3,6,7]. A recent systematic review investigated the situation of schizophrenic patients in 24 countries and revealed that the annual cost for the population of patients with schizophrenia ranged from US $94 billion to US $102 billion per country, which consumes up to 1.65% of a country's gross domestic product [8]. These expenses cover antipsychotic treatments and long-term psychological interventions [9], which are the recommended treatments for schizophrenia [10], and these treatments have demonstrated optimal control over positive symptoms. However, multiple factors hinder the efficacy of this treatment, including the patients' limited insight into mental illness [11], noncompliance [12], and treatment resistance [13,14]. The management of negative symptoms also remains a clinical challenge [15]. Therefore, new types of treatment methods are needed to promote recovery and optimize rehabilitation [15].

In recent years, different types of video games, including computer games [16,17], mobile games [18], serious games [19-23], and virtual reality [24-26], have been created as adjunctive treatments for various illnesses and they have resulted in desirable outcomes. Scholars suggest that video gaming is a novel intervention for social reconstruction and skills training in people with mental illness [16,19,21,27]. The use of video gaming in psychiatric treatment is promising because video gaming is a rapidly expanding market. In 2018, there were 2.3 billion gamers making up the total revenue of US $137.9 billion globally [28]. The Asia-Pacific region was the largest contributing region, which shared 52% of the worldwide revenue [28]. Mobile gaming has increased by 25.5% per year from 2017 and it now holds 51% of the market. In contrast, a downtrend has been noted for computer gaming [28]. Although the use of information and communication technologies has been rapidly expanding in the world, the accessibility, frequency, and purpose of using information and communication technologies vary among geographical and demographical areas [6,29-32]. In the United States, 97% of the citizens possess a personal computer and 81% have a smartphone, which they use every day [6].

In the United States, 89% of the people with schizophrenia possess a computer and 54% have a smartphone [29]. The majority of these people spend more than 1 hour per day on a computer and a mobile phone, and the main purpose of use includes surfing the internet, social networking, and playing web-based games. One study in Finland revealed that only 55% of the people with schizophrenia had a computer and even fewer had access to internet connection [30]. A recent pilot study investigated the usage of mobile and computer devices among people with severe mental illness [31]. Despite up to 45% of them showing interest in using the devices to support recovery, the majority only used them for general purposes such as listening to music (60%), accessing the internet (59%), and for making calls (59%) [31]. Specifically, young people were more likely to possess a computer and have internet connection, and the main purpose of using the internet included seeking information, studying, web-based chatting, shopping, and gaming [6,30].

People with schizophrenia have positive attitudes toward modern information technology use [33], including gaming [34,35]. Gaming attitudes specifically refers to gamers' motives and preferences regarding video games [32] and these attitudes vary among gamers with different backgrounds [36]. Ryan et al [36] explored gaming attitudes through the application of the self-determination theory and suggested that if a gamer's preferences are satisfied, they experience enjoyment when playing the game. Researchers have also identified certain game features to describe the gaming attitudes of gamers, such as storytelling, individual games, and social games [32,36]. Hilgard et al [32] identified 9 factors to measure the motives and preferences of gamers: story, violent catharsis, violent reward, social interaction, escapism, loss aversion, customization, grinding, and autonomy. These factors vary among different players depending on their characteristics such as their social motives, favorite games, and gaming platforms [32]. In general, gaming serves as a primary source of entertainment while simultaneously aiding in constructing and maintaining interpersonal relationships [16]. Beyond enjoyment, players can learn and develop new skills through gaming [16,29]. A study in the United States found that video games are most frequently utilized as a stress reliever; gamers enjoy playing video games as it takes them away from reality. Most video gamers enjoy playing with friends, while male gamers prefer competitions and feel proud when they achieve awards in a game [37].

Recently, gaming has been seen as a possible method of rehabilitation for people with schizophrenia [34]. However, there is still a lack of knowledge about the habits and attitudes of video gaming among people with schizophrenia [15,21,22,24,27,38,39]. Additionally, cultural differences in the attitudes of gaming [40] among people with schizophrenia might exist and have a potential impact on the treatment focus. By exploring the knowledge gap, we might contribute to the development of new interventions for the management of schizophrenia in the future [15]. Therefore, we conducted a survey to explore the habits and attitudes of video gaming and information technology use among people with schizophrenia in Hong Kong and the factors that are associated with their gaming attitudes.

XSL•FO
RenderX

Methods

Study Design

A cross-sectional survey design was adopted in this study because it allowed us to explore the attitudes of the respondents by using a structured survey format [41-44].

Ethical Approval

This study was designed to fulfill the ethical assessments mentioned in the Declaration of Helsinki [45]. This study was assessed by the Human Subjects Ethics subcommittee at the Hong Kong Polytechnic University (HSEARS20180313005) and approved by study setting organizations. An information sheet with the explanation of the study and the rights of participants was provided to each participant. Informed consent was obtained from each participant and witnessed and countersigned by a researcher. The study was guided by the ethical principles of autonomy, confidentiality, anonymity, and nonmalfeasance [45]. It emphasized voluntary participation and participants having the right to withdraw from the study at any time [41,45]. The personal identity of the participants was protected such that their identities were not reported in any publication [45]. Collected data were kept in a locked cabinet and password-protected files, which were only accessible to the research personnel [41]. If there was any concern of emotional burden, participants were free to contact the research team for support through the contact information printed on the information sheet [46].

Setting

This study was performed in 2 settings, which are both in Hong Kong: integrated community centers for mental wellness and halfway hostels. In Hong Kong, rehabilitation services for people with mental illness in the community are operated by nongovernmental organizations, which operate various services to support the rehabilitation of people with mental illness living in the community such as the settings used in this study. There are 24 integrated community centers for mental wellness providing community support to individuals and their families on a district basis [47] and 36 halfway hostels providing residential support in Hong Kong [48]. Altogether, these types of services cover more than 26,000 citizens annually [49]. An invitation letter was sent to all the nongovernmental organizations that operated integrated community centers for mental wellness and halfway hostels in Hong Kong. Out of the 13 possible nongovernmental organizations, 2 agreed to join the study. Finally, the study was performed at 7 integrated centers for mental wellness and 6 halfway hostels.

Target Population

The target population consisted of people with schizophrenia who were living in the community [1,3,5]. At the time of the study, there were approximately 48,000 citizens who had been diagnosed with schizophrenia in Hong Kong, and most of them resided in the community [49]. Unlike those who were receiving inpatient psychiatric treatments, people living in the community were free to access video games and information technologies [50]. Therefore, it was more appropriate to explore the habits and gaming attitudes of the people with schizophrenia who resided in the community [27,32,38].

Recruitment and Sampling

This study adopted convenience sampling in the participant recruitment. Poster promotion and data collection sessions were first arranged at the respective study organizations. Second, inclusion and exclusion criteria for the study participants were screened by representatives of the nongovernmental organizations. This was done to protect the privacy of the participants [45]. The inclusion criteria were as follows: (1) primary diagnosis of schizophrenia (schizophrenia spectrum and other psychotic disorders) under the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [1]; (2) aged 18 to 64 years and capable of giving informed consent without a legal guardian; (3) able to read and understand traditional Chinese, which was the prevailing language and understandable by the majority of the citizens in Hong Kong; (4) willing and able to give informed consent [45]; and (5) a resident in the community. The exclusion criteria were as follows: (1) aged <18 or >64 years; (2) meeting diagnostic criteria for a current major depressive, manic, or hypomanic episode (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition), intellectually disabled, severe visual impairment [1]; (3) active gamer (ie, playing >5 hours/day), for screening purposes only [51]; (4) meeting the diagnostic criteria for substance use disorder (other than tobacco use) [1], head injury, hemiplegia, or other neurological disorders.

Instruments

Background Information of the Participants

Demographic characteristics were collected, including age, gender, education qualification, (no formal education, primary and middle school, high school, college or vocational training, bachelor's degree, master's degree, doctoral degree), employment status (unemployed, social welfare benefit, student, disability pension, employed [including sick leave], other), marital status (single, partnership or married, separated or divorced, widowed), residential condition (own household [with partner or family], own household [alone], flat share, parents' household, supported housing), and the age of the first contact with psychiatric services.

Internet Use

The use of the internet was explored using a set of questions extracted from Choi and DiNitto's study [52]. The questionnaire consisted of 6 items, that is, internet connection (yes/no), possession of email address (yes/no), frequency of using the internet (at least once a day, every few days, once a week, a few times a month, once a month or less often than that), purpose of using the internet (seeking information, communication, shopping, gaming, peer support, others), accessibility to web information (5-point Likert scale, 1=always easy, 5=very difficult), and issues that posed problems in using the internet (problems with limbs, hands, concentration, long-term sitting, eyes, others).

XSL•FO
RenderX

Habits of Video Gaming

Habits of video gaming were explored using 7 items, namely, video gaming in the past year (yes/no), seriousness of video gaming (5-point Likert scale, 1=very intense, 5=very casual), frequency of video gaming (daily, 2-3 times per week, weekly, 2-3 times per month, monthly, less than monthly), proportion of spare time for video gaming (5-point Likert scale, 1=almost all of the spare time, 5=almost none of the spare time), favorite and current video games (respondents could list 3 of each), and favorite gaming platform. Subjects were not required to continue answering questions if they indicated that they had not a played a video game within the past year [6,47].

Gaming Attitudes

Gaming attitudes were assessed using the Gaming Attitudes, Motivations, and Experiences Scales (GAMES) [32]. The GAMES has 59 items to measure the gaming attitudes, habits, and preferences of the gamers. The items are classified into 9 categories: story (12 items), violent catharsis (7 items), violent reward (6 items), social interaction (6 items), escapism (6 items), loss aversion (7 items), customization (4 items), grinding (6 items), and autonomy (5 items, Table 1) [32]. Each item was rated on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The summation of the items formed the subscores of the 9 categories; the higher the score reflected the higher preference for a specific category. Factor analysis procedures were conducted with split-halves exploratory factor analysis and confirmatory factor analysis [32]. This instrument possesses good internal consistency in all categories with the Cronbach α=.79-.92 [32].

Table 1. GAMES[a] categories, score range, and their definitions.

Category	Score range	Definition
Story	12-60	Game stories are important, engaging, and emotionally compelling.
Violent catharsis	7-35	Game violence is perceived to harmlessly help release negative moods or aggression.
Violent reward	6-30	Game violence provides positive or thrilling emotions such as satisfaction or power.
Social interaction	6-30	Playing games with a group; developing personal relationships with other players.
Escapism	6-30	Using games to regulate dysphoric moods or to escape the frustrations of daily life.
Loss aversion	7-35	Tendency of a loss to frustrate or to “spoil the fun.” Likely subsumes a search for a challenge.
Customization	4-20	Interest in in-game creative pursuits such as personalizing an in-game avatar or building a house.
Grinding	6-30	Attitudes toward performing repetitive actions or paying real-life money to earn in-game rewards; interest in performing every possible action in a game or collecting every in-game item.
Autonomy	5-25	Enthusiasm for games with many choices, options, multiple solutions to puzzles, and open areas to explore.

[a]GAMES: Gaming Attitudes, Motivations, and Experiences Scales.

Translations of the Instruments and a Pilot Study

The original version of GAMES was written in English; thus, a translation and back translation were performed to ensure that the validity was preserved [41]. First, the original instrument was translated into traditional Chinese by the researchers. Then, a language expert translated it back into English. Third, the original and the translated English versions were compared to ensure that the validity was retained after the translation [41]. In cases of any discrepancy between the original and the translated versions, the translation was reassessed by a researcher, translator, and the original author of GAMES.

A pilot study was conducted to examine the data collection procedure and reliability of the instrument [53]. Meanwhile, the feasibility of the data collection was tested [54]. We recruited 20 people with schizophrenia who were living in their community in Hong Kong. Two identical 2-way translated questionnaires were distributed to participants with an interval of 2 weeks [55]. The intraclass correlation coefficient was assessed to identify the test-retest reliability, and the Cronbach α was assessed to identify the internal consistency of the instrument [53,54]. The test-retest reliability was moderate to good with an intraclass correlation coefficient of 0.58-0.95. The internal consistency was good with a Cronbach α of .71-.94. The reliability of the translated version was comparative to that of the original GAMES.

Data Collection

The data collection period lasted for 16 weeks—from October 2017 to January 2018. A researcher promoted the study with posters and arranged venues for briefing and data collection. Identical sessions were run by the same researcher at different time points at each data collection site [41]. All eligible participants were invited to attend an information session organized together with the nongovernmental organization. Potential participants were informed about the purpose of the study, the rights of the participants, risks of participation, and possible benefits, both orally and in written format. They were also informed about their right to refuse and withdraw from the study at any time without any consequences [45]. After informed consent forms were signed, the questionnaire was distributed to each participant. After completion, questionnaires were collected using sealed envelopes [41]. Each participant received a supermarket redemption coupon worth HKD $50 (US $6.45) as compensation for their participation in the study. Data collection sessions lasted for 30-60 minutes [32]. A total of 18 data collection sessions were held at the 13 data collection sites.

XSL•FO
RenderX

Finally, 148 people were contacted and invited to participate in the study, and 110 (74.3%) eligible people were willing to participate in the study and they completed the questionnaire.

Data Analysis

First, collected data were processed through data input, data cleansing, and data reversing [41]. Second, frequencies, percentages, maximum score, minimum score, mode, median, mean, and standard deviation were used to describe the demographic characteristics and the participants' habits and attitudes regarding internet use and video gaming [52]. To facilitate the analysis of the current and favorite video games played, the entries were recategorized into different game genres. We utilized an extensive gaming archive [56] to sort the entries into 10 mutually exclusive game genres. Lastly, a two-tailed independent *t* test was used to describe the possible differences in the normally distributed variables of the gaming attitudes. The Pearson's correlation coefficient was used to quantify the linear relationship between the categories of the gaming attitudes. Statistical significance was established at P=.05. For the dimensions of the categories of the gaming attitudes that were the most intercorrelated, principal analysis with 3 methods of rotation (varimax, equamax, and promax) was used to confirm the convergence of the 9 categories of gaming attitudes into fewer principle factors for easy interpretation. The SPSS software for Windows 22.0 (IBM Corp) was used for the data analysis.

Results

Background Information of the Participants

The mean age of the 110 participants was 39.0 (SD 11.2) years. About two-thirds of the participants (70/110, 63.6%) were males. A clear majority (91/110, 82.7%) was single and had completed high school or higher education (73/110, 66.4%). Totally 26.4% (29/110) of the participants were receiving social welfare benefits, 18.2% (20/110) were receiving disability pension, while 25.4% (28/110) were employed and 23.6% (26/110) were unemployed. Of the 110 participants, 78 (70.9%) lived in supported housing (Table 2). The mean reported age at the first contact with psychiatric services was 24.7 (SD 8.5) years, and the average duration of illness was 14.3 (SD 10.3) years.

Internet Use

The majority of the participants had access to an internet connection (100/110, 90.9%). The most common purposes of using the internet were to watch videos (66/110, 60.0%), read the news, papers, magazines, and books on the internet (42/110, 38.2%), research information about other topics or issues of interest to them (41/110, 37.3%), and play web-based games (36/110, 32.7%). Participants found using the internet to be easy (69/110, 62.7%). The major causes that lessened the use of the internet were when eyes were tired easily (40/110, 36.4%) and when one had difficulty concentrating for long periods of time (26/110, 23.6%) (Table 3).

XSL•FO
RenderX

Table 2. Background information of the participants (N=110).

Variables	Values, n (%)
Gender	
Male	70 (63.6)
Female	40 (36.4)
Marital status	
Single	91 (82.7)
Partnership/Married	9 (8.2)
Separated/Divorced	10 (9.1)
Level of education	
Primary and middle school	37 (33.6)
High school	51 (46.4)
College or vocational training	10 (9.1)
Bachelor's degree	12 (10.9)
Employment status	
Unemployed	26 (23.6)
Social welfare benefit	29 (26.4)
Student	2 (1.8)
Disability pension	20 (18.2)
Employed (including sick leave)	28 (25.5)
Other	5 (4.5)
Living situation	
Own household (with partner/family)	8 (7.3)
Flat share	17 (15.4)
Parents' household	7 (6.4)
Supported housing	78 (70.9)

XSL•FO
RenderX

Table 3. Patterns of internet use among the participants (N=110).

Patterns of internet use	Values, n (%)
Access location of the internet	
Home	77 (70.0)
Apartment complex	5 (4.5)
Family/friend's home	9 (8.2)
Library	4 (3.6)
Other	5 (4.5)
Not using the internet at all	10 (9.1)
Possession of an email address	
Yes	63 (57.3)
No	47 (42.7)
Frequency of using internet	
At least once a day	54 (49.1)
Every few days	14 (12.7)
Once a week	6 (5.5)
A few times a month	3 (2.7)
Once a month or less	33 (30.0)
Purpose of internet use	
Research health-related information	18 (16.4)
Communicate with health professionals about health-related issues	4 (3.6)
Communicate with other users about health-related issues	8 (7.3)
Research information about other topics or issues of interest to me	41 (37.3)
Send/receive email	24 (21.8)
Online shopping	15 (13.6)
Perform online banking/pay bills	13 (11.8)
Read web-based news, papers, magazines, and books	42 (38.2)
Play web-based games	36 (32.7)
Watch videos	66 (60.0)
Use social networking websites and/or dating sites	24 (21.8)
Other	9 (8.2)
Easiness to access the intended information on the internet	
Always easy	22 (20.0)
Sometimes easy	47 (42.7)
Not so easy	17 (15.5)
Difficult	11 (10.0)
Very difficult	13 (11.8)
Problems that lessen internet use	
Pain in the limbs	7 (6.4)
Unsteady hands	6 (5.5)
Difficulty concentrating for long periods of time	26 (23.6)
Difficulty sitting for long periods of time	18 (16.4)
Eyes that tire easily	40 (36.4)
Other problems	6 (5.5)

XSL•FO
RenderX

Habits of Video Gaming

The majority of the participants (88/110, 80.0%) reported that they had played a video game within the past year. The habits of video gaming of the 88 participants were explored (Table 4). Almost half of the participants (41/88, 47%) regarded themselves as playing games seriously or very seriously. Half of our participants played video games on a daily (22/88, 25%) to weekly basis (26/88, 30%). The most favorable gaming platforms were cellular phones (43/88, 49%), computers (19/88, 22%), and arcade cabinets (6/88, 7%) (Table 4). Referring to the video games that participants were currently playing and their favorite video games, there were 118 and 145 entries, which included 81 and 84 games, respectively (Table 5). The favorite game genre among the participants was that of action games.

Table 4. Habits of video gaming of the participants (n=88).

Habits of video gaming	Values, n (%)
Seriousness of video gaming	
Very serious	14 (16)
Serious	27 (31)
Neither serious nor casual	17 (19)
Casual	25 (28)
Very casual	5 (6)
Frequency of video gaming	
Daily	22 (25)
2-3 times per week	15 (17)
Weekly	11 (13)
2-3 times per month	11 (13)
Monthly	8 (9)
Less than monthly	21 (24)
Proportion of spare time spent on video gaming	
Almost all of the spare time	4 (5)
Most of the spare time	6 (7)
Some of the spare time	27 (31)
Less of the spare time	30 (34)
Almost none of the spare time	21 (24)
The most typical media platform used to play games	
Cellular phone	43 (49)
Computer	19 (22)
Arcade cabinets	6 (7)
Facebook	3 (3)
Board or card games (tabletop games)	3 (3)
Sony PlayStation	2 (2)
Sony PlayStation Portable	2 (2)
Microsoft Xbox	2 (2)
Real-life sports	2 (2)
Nintendo Wii	1 (1)
Nintendo DS	1 (1)
Not specified	4 (5)

XSL•FO
RenderX

Table 5. Video games played by the participants.

Game genres	Currently played game genres (n=118), n (%)	Favorite game genres (n=145), n (%)
Action	23 (19.5)	34 (23.4)
Strategy	23 (19.5)	26 (17.9)
Miscellaneous (board/card games)	20 (17.0)	27 (18.6)
Role-playing	15 (12.7)	17 (11.7)
Simulation	14 (11.9)	13 (9.0)
Puzzle	10 (8.5)	11 (7.6)
Sports	9 (7.6)	13 (9.0)
Action adventure	3 (2.5)	3 (2.1)
Adventure	1 (0.8)	N/A[a]
Rhythm	N/A	1 (0.7)

[a]Not applicable.

Gaming Attitudes

The gaming attitudes of the participants were explored, and the results are summarized in Table 6. Of the 88 participants, the distribution of the scores in the category of story ranged from 21 to 50 with a mean score of 37.81 (SD 6.33) (Table 6).

To compare the scores among the 9 attitude measures, we standardized the raw score of each participant as a relative score for each attitude measure. The mean of the standard score in the last column of Table 6 suggests that the top 3 favorable video gaming categories for participants were autonomy, escapism, and story, and the last two were violent catharsis and violent reward. Furthermore, the gaming attitudes of the participants were mostly correlated with each other (Table 7). A principal analysis with 3 methods of rotation (varimax, equamax, and promax) indicated that the categories of story, social interaction, customization, and autonomy formed one factor, which are related to social playing and explains 13.1% of the variation, and the rest of the categories formed another factor, which are related to evasive playing, thereby explaining 45.7% of the variation.

Table 6. A summary of the gaming attitudes of the participants (n=88).

Categories	Minimum score	Median score	Mode	Maximum score	Mean (SD)	Mean of standard score in %[a] (rank)
Social playing						
Story	21	38	37	50	37.81 (6.33)	54 (3)
Social interaction	6	19	20	30	18.55 (4.78)	52 (5)
Customization	4	12	12	20	12.05 (3.89)	50 (6)
Autonomy	6	17	15	25	16.67 (4.01)	58 (1)
Evasive playing						
Violent catharsis	7	21	14, 21	35	19.92 (5.91)	46 (8)
Violent reward	6	16	18	30	15.19 (5.44)	38 (9)
Escapism	6	20	18	30	19.16 (4.92)	55 (2)
Loss aversion	8	21	22	33	20.74 (4.60)	49 (7)
Grinding	8	18	18	30	18.88 (4.09)	54 (4)

[a]Standard score is defined as Absolute (actual score – minimum value)/theoretical range.

XSL•FO
RenderX

Table 7. Pearson's correlation coefficients and *P* values among the video gaming attitude measures (n=88).

Variable	Violent catharsis	Violent reward	Social interaction	Escapism	Loss aversion	Customization	Grinding	Autonomy
Story								
r	0.23[a]	0.22[a]	0.37[b]	0.28[b]	–0.05	0.33[b]	0.30[b]	0.38[b]
P value	.03	.04	<.001	.008	.67	.002	.004	<.001
Violent catharsis								
r	—[c]	0.44[b]	0.32[b]	0.40[b]	0.46[b]	0.40[b]	0.53[b]	0.39[b]
P value	—	<.001	.003	<.001	<.001	<.001	<.001	<.001
Violent reward								
r	—	—	0.20	0.26[a]	0.24[a]	0.21[a]	0.31[b]	0.17
P value	—	—	.06	.02	.02	.05	.004	.11
Social interaction								
r	—	—	—	0.40[b]	0.20	0.54[b]	0.52[b]	0.38[b]
P value	—	—	—	<.001	.06	<.001	<.001	<.001
Escapism								
r	—	—	—	—	0.53[b]	0.44[b]	0.52[b]	0.59[b]
P value	—	—	—	—	<.001	<.001	<.001	<.001
Loss aversion								
r	—	—	—	—	—	0.30[b]	0.48[b]	0.38[b]
P value	—	—	—	—	—	.005	<.001	<.001
Customization								
r	—	—	—	—	—	—	0.60[b]	0.50[b]
P value	—	—	—	—	—	—	<.001	<.001
Grinding								
r	—	—	—	—	—	—	—	.74[b]
P value	—	—	—	—	—	—	—	<.001

[a]The correlation is significant at a significance level of .05 (two-tailed).

[b]The correlation is significant at a significance level of .01 (two-tailed).

[c]Not applicable.

Association of the Background Information with Gaming Attitudes

Several background characteristics such as marital status, education level, and employment status were found to be significantly associated with several categories of gaming attitudes (Multimedia Appendix 1). As for marital status, participants who were married or who were in a partnership had a higher score than those who were single, separated, or divorced in the violent catharsis category (t_{86}=2.66, *P*=.01). With regard to the education level, participants who held a bachelor's degree or higher had a lower score than those who were educated with college or vocational training or a lower level of education in the violent reward category (t_{86}=–2.57, *P*=.01). Likewise, the participants with a bachelor's degree or higher had a lower score in social interaction than those who did not (t_{86}=–2.25, *P*=.03). Further, the participants who had a bachelor's degree had a lower score in customization than those who did not (t_{86}=–2.08, *P*=.04). Regarding the employment status, participants who were employed or studying scored lower than those who were unemployed or others in the violent reward category (t_{86}=–3.11, *P*=.003). Similarly, participants who were employed or studying scored lower than those who were unemployed or others in the loss aversion category (t_{86}=1.98, *P*=.05).

The gaming habits of the participants were also found to have a significant association with several gaming attitudes (Multimedia Appendix 1). Regarding the seriousness of video gaming, noncasual players had a higher score than casual players in the customization category (t_{86}=2.00, *P*=.05). Participants who played video games on a daily basis scored higher on story than those who played less than daily (t_{86}=2.03, *P*=.05). Participants who reported that they spent more spare time on video gaming (almost all/most of/some of their spare time)

XSL•FO
RenderX

scored higher on violent reward than those who spent less spare time (few/almost none of their spare time) (t_{86}=2.14, *P*=.04). Those who spent more spare time on video gaming also scored higher in escapism than those who spent less spare time on gaming (t_{86}=2.79, *P*=.007). However, no significant correlation was found between age, the age at the first contact with psychiatric services, or duration of illness and any of the categories of the gaming attitudes.

The analysis (Multimedia Appendix 2) based on the factor loading weighted mean score of the 2 summarized factors of the gaming attitudes, "social playing" and "evasive playing" provided further support for the above findings in a summarized manner. This analysis suggested that participants with a bachelor's degree or above and employed or studying seemed to have lower scores in social playing than their counterparts, thereby reflecting a difference in their attitude regarding social interaction. However, the differences were not statistically significant owing to the sample being small. Participants employed or studying had significantly low scores in evasive playing (t_{86}=2.16, *P*=.03), which is a reflection of the significant differences in their attitudes on violent reward and loss aversion. The notable difference between the 2 gaming attitude components was that those in partnership/marriage scored much higher in evasive playing than single/separated/divorced individuals (68.73 vs 61.94, respectively), although they scored similarly in social playing (60.62 vs 61.38, respectively), that is, the difference was not significant in this study. The differences in the levels of seriousness and frequency were similar to the differences between the 2 factors of attitudes, but there was a larger difference in the proportion of spare time spent on video gaming; more spare time was spent on evasive playing than on social playing, with a borderline significance of t_{86}=1.94, *P*=.06.

Discussion

Principal Findings

This study is the first to describe the habits and attitudes of internet use and video gaming among people with schizophrenia in an Asian city. Similar to the global trends, the majority of the participants (100/110, 90.9%) utilized the internet with adequate skills and participated in video gaming [6]. Their favorite gaming platform and game genres were cellular phones and action games, respectively. However, only few of the participants had utilized the internet for seeking health information, and the efficacy of health education through web-based platforms to this population was limited. The gaming attitudes were identified and found to be associated with several characteristics. The participants valued autonomy in gaming and disfavored violence in video games. Through our analysis, we established the motives of the gamers. "Social playing" included social interaction, story, customization, and autonomy, while "evasive playing" included grinding, loss aversion, escapism, violent catharsis, and violent reward.

Use of Information Technology

The utilization of information technology was found to extend to almost the entire population of Hong Kong, which is a modern city [57]. We found that the frequency of internet use by most of our participants was higher than those reported in the United States [29] and Finland [30]. The high accessibility to the internet could be explained by the affordable and extensive coverage of the internet network [58]. As reported in previous studies [6,30], our participants mainly utilized the internet to watch videos and obtain information by reading news, papers, magazines, and books. Over half of the participants (69/110, 62.7%) found it easy to obtain the intended information from the internet. Our study also demonstrates that most of the participants were able and willing to obtain information from the internet.

Despite the availability of web-based resources and the ability to use them, only 16.4% (18/110) of the study participants reported utilizing the internet for researching health-related information. This may imply that the efficacy of mental health promotion through web-based platforms is still limited. McCloud et al [59] suggest that the low motivation of internet users to search for health-related information might be due to users lacking the skills needed to gather and identify trustworthy sources of information. Athanasopoulou et al [60] and Schrank et al [33] also demonstrated that the quality of most of the easily accessed information sources about schizophrenia on the internet is poor. This can hinder internet users from receiving useful and nonstigmatized information about managing schizophrenia. It would be preferable for health care authorities to develop and use a trustworthy internet platform through which health-related information for health promotion could be easily distributed. Furthermore, our results reflect that there are physical constraints that pose barriers to the use of the internet. The major barriers were found to be tired eyes and concentration difficulties. The adverse effects of antipsychotic drugs, such as drowsiness, sedation, and anticholinergic effects, might be a cause of the barriers [61]. Thus, it is important for health care staff to address safe methods in promoting health-related issues through trusted web-based platforms on a daily basis.

Habits of Video Gaming

Video gaming is popular in Hong Kong, and approximately 94% of the adolescents comprise the gaming population [62]. Likewise, a majority of the participants had played video games within the past year (88/110, 80%). As reported globally [63], the most popular gaming platform among the participants was cellular phones. There has been a global shift toward playing games on cellular phones rather than sitting in front of a television or a desktop computer. With the high mobility and convenience of using phones, gamers can enjoy playing without geographical constrictions. Indeed, game developers have launched numerous games on mobile platforms in recent years [6,63]. However, since players use a limited-sized screen and gaming times are generally short in mobile gaming, designing games for mobile phones with long and complex stories is not preferred [62].

The participants also played video games frequently; more than half of them (48/88, 55%) played at least once a week and (22/88, 25%) played on a daily basis. About half of the gamers emphasized on video gaming, as they played seriously. As reflected in the best-selling video games of 2017, action games

XSL•FO
RenderX

were the favorite game genre among the participants. Action games challenge gamers through physical means and players usually control a character to complete tasks that require a certain level of hand-eye coordination and skills [6,64]. Action games attract players by rewarding them with excitement [62]. Contrary to the global best-selling trends, our results indicated that card and board games and strategy games were the next favorite types of games, while the best-selling games worldwide were role-playing and sports games [6]. This finding might be accounted for by the cultural and local context. Local game developers in Hong Kong have converted some culturally traditional board games such as Mahjong, Chinese Chess, and some card games into digital versions, which are easy to use. In addition, some traditional Chinese stories such as "Romance of the three kingdoms" has been converted to a series of strategy games. However, the characteristics of the players in Hong Kong do not seem to differ significantly from those of the rest of the world with regard to popular mobile games [63]. Thus, one possible explanation for our finding of people with schizophrenia favoring traditional games could be related to the negative symptoms of the disease, such as cognitive deficits and a lack of confidence about one's own abilities [64]. These factors could lead to favoring familiar game genres.

Gaming Attitudes

Our study identified that autonomy was the most common motive for video gaming among the participants. This is seen in the preference of the participants to play games in which they could manipulate settings and make choices. Video games provide a normalized platform wherein people can exercise self-sufficiency and thus develop self-competence [65]. The next most common motive was escapism, which implies that participants played games to regulate dysphoria and escape from the hassles of daily life. Indeed, video gaming has been considered a culturally acceptable way to relieve the stress of reality [23,25]. On the other hand, some scholars argue that escapism is associated with gaming addiction and adverse emotional experiences [66-69]. A recent study examined the effects of video gaming on a group of military veterans with mental problems and found that video gaming contributed to their recovery by promoting individuals' adaptive coping, eudaimonic well-being, and socializing [70]. As the evidence related to harm caused by gaming among people with schizophrenia is not yet conclusive, staff working in clinical practice need to be sensitive to and be aware of any potential harmful effects of the interventions and avoid any such threats.

Our participants regarded violent reward and violent catharsis as the least favorable motives for video gaming. Numerous studies suggest that violent games might contribute to aggression [70-72] and moral desensitization [73,74]. On the contrary, scholars also suggest that violent games support emotional coping after stressful events [75,76]. Interestingly, our results indicate that people with schizophrenia do not favor violence in video gaming. Moreover, several characteristics were associated with a preference of violent video gaming. People who were married or in a partnership tended to channelize their negative emotions by playing violent games to avoid hurting others. In contrast, the unemployed and those with lower education levels enjoyed the satisfaction they received from playing violent games. These results implied that people with different sociodemographic characteristics achieve emotional coping through different ways. However, more studies are needed with larger sample sizes to confirm valid links between sociodemographic information and gaming motives.

Studies have shown that persons with schizophrenia are lonely and isolated and less motivated to socially engage with others [3-7]. Our analysis regarding gamers' motives identified a factor, "social playing," in several categories, namely, social interaction, story, customization, and autonomy. This factor refers to people who emphasize social processes in video gaming. These players enjoy games that enable them to customize and manipulate numerous settings. The story and settings of the video games strongly influence the gaming experience for the players. Gamers value the social interaction with friends and other players during the gaming process. Likewise, several background characteristics were related to social playing motives. In the past decade, there has been an expansion of multiplayer web-based role-playing games in the market. These have provided a virtual platform for players to interact with people worldwide [6,55,62,77]. Scholars have proven that the formation of social bonds in virtual communities through web-based video games facilitates the development of interpersonal skills in the real world [78,79]. This shows the potential of utilizing video games, regardless of the players' education levels, to facilitate interpersonal skills among people with mental illness [80,81] and to increase feelings of relatedness [36]. Frequent game players in our data were more concerned with the game story. Likewise, noncasual players and people with a lower education level preferred games with customizable settings. With the rapid expansion of the gaming market, gamers have a higher expectation of the quality of games, including that of the story, settings, ability to personalize, and the gaming experience [82]. Indeed, gamers can exercise creativity and control their game by tailor-making their own game settings. Studies have shown that the ability to personalize games motivates people to play games [82-84].

On the contrary, another factor, "evasive playing," was formed by the remaining categories, namely, grinding, loss aversion, escapism, violent catharsis, and violent reward. Evasive playing refers to when gamers emphasize the rewards from video gaming and spend time on achieving all gaming missions and collecting all possible props in a game. In this style of playing, the violence in games is utilized to channel negative emotions and attain enjoyment. Our results indicate that unemployed people have a high loss-aversion score, which could imply that gaming is significantly important to them. People who spent more spare time on gaming were more likely to utilize gaming to escape from the hassles of daily life. However, scholars suggest that these motives are linked to problematic use [15,85] and pathological gaming [32], which should be seriously considered when discussing gaming among people with mental illness. In the future, more studies should therefore be conducted in schizophrenic patient populations with increased sample sizes and a longitudinal design and follow-up to explore problematic uses of video games, both in clinical and community settings.

XSL•FO
RenderX

Strengths and Limitations

This study fills a knowledge gap by depicting a pattern of internet use and video gaming among people with schizophrenia. Despite the research aims being met, there are some limitations to address. First, participant recruitment was conducted using a convenience sample. Thus, the sample size was small, which limits the generalization of the results to the wide population of video game players in Hong Kong or in other parts of the world. Second, our participants were recruited from 2 nongovernmental organizations (supported housing); therefore, the results may be biased toward individuals who live in communities with better general levels of mental health. Third, the study may have attracted people who were interested and skillful in video gaming. Further, the life span and the popularity of video games are highly influenced by the social trends that might change rapidly.

Implications

The utilization of the internet as a platform for health promotion and education has been considered to be a cost-effective measure [86,87]. Most people with schizophrenia in our study were found to be able to utilize information technology to obtain information on the internet. This finding is promising. Based on a systematic review by Suenderhauf et al [15], regular video game playing could enable home monitoring of patients. For example, staff could use gaming as a tool for patient monitoring in order to initiate supportive measures to avoid relapses. Decreased login times could indicate diminishing compliance or worsening of disease. Likewise, alarming habits in computer use such as the lack of logins in the regular gaming schedule or playing beyond suggested gaming hours could be an indicator for health care professionals of changes in the patients' mental status. However, this type of monitoring would require specified technological devices and access for patients who play games remotely, which may increase potential concerns in patient privacy and security issues. It is therefore important to delineate the roles and responsibilities of health professionals in patient monitoring in any exceptional situation (eg, suicide attempt). Professionals should also support the use of digital health interventions; their attitudes are crucial for successful implementation of health technology in patient daily care [88]. Our previous studies have shown that staff members may still be less willing to adopt new technology-based interventions into mental health care [89,90] as they tend to prefer traditional face-to-face methods in mental health care services [91].

We found that the internet was less often used to seek health-related information. Scholars and health care professionals might need to investigate the cause of the low utilization rate of the internet with regard to searching for health-related information. To promote internet use in a population, health care authorities should consider developing an internet platform with high-quality health information for health promotion. Service organizations should provide equipment, training, and ongoing information technology support to the population in order to expand their social networks and maximize self-help abilities. Policy makers should promote the use of information technology in a schizophrenic population and offer support to the nongovernmental organizations for the abovementioned measures.

Over the last decade, numerous studies have shown video gaming as an intervention for various medical conditions. From the mental health aspect, health care professionals could utilize video games to promote patient engagement, stress coping, and social skills training. Our findings show that game developers could design video games to fit the needs of the population. Game developers and health care professionals could refer to the identified gaming attitudes to tailor-make attractive serious games for populations with mental health problems accordingly. Therefore, in the future, more studies should be conducted with more diverse samples to identify the potential benefits of video gaming for patients with mental problems. The use of health-related games should also be implemented and tested as part of the mental health recovery. Despite the huge potential that information technology has in the treatment of patients with mental disorders, we still may have a long way to go before research on computer-based interventions can realize its full potential in psychiatric services [92].

Conclusion

This study demonstrates that the internet utilization rate of people with schizophrenia was as high as that of the general population in Hong Kong. Relatively few of them used the internet to search for health-related information. Further investigation into the underlying reasons would be vital to mental health promotion through web-based platforms. Our study also exemplifies the unique habits of gaming among the study participants. Health care professionals could utilize video games to engage patients, promote stress coping, and provide social skills training. The identified gaming attitudes can be used to contribute to the development of serious games for the population in the future.

Acknowledgments

We would like to thank the Hong Kong Polytechnic University (code YBYB, grant holder MV) and the University of Turku, Finland (code 26003424, grant holder MV) who supported this study. We would also like to thank the mental health associations in Hong Kong for their assistance in recruiting participants for the study. In particular, we would like to thank all the participants and the staff from the associations who offered their kind assistance to our study.

XSL•FO
RenderX

Authors' Contributions

MV designed the study. WTHC, DKSY, and TW collected the patient survey data. WTHC conducted statistical analysis and wrote the manuscript. TL and MV wrote the manuscript, and MY conducted additional statistical analysis. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Associations between background characteristics and video gaming attitudes.
[DOCX File , 19 KB - jmir_v22i7e14865_app1.docx]

Multimedia Appendix 2
Associations between background characteristics and principle components of video gaming attitudes.
[DOCX File , 17 KB - jmir_v22i7e14865_app2.docx]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. In: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
2. Brekke JS, Hoe M, Green MF. Neurocognitive change, functional change and service intensity during community-based psychosocial rehabilitation for schizophrenia. Psychol Med 2009 Oct;39(10):1637-1647. [doi: 10.1017/S003329170900539X] [Medline: 19243648]
3. World Health Organization. Schizophrenia. 2017. URL: https://www.who.int/mental_health/management/schizophrenia/en/ [accessed 2019-02-25]
4. Chen EY, Hui CL, Dunn EL, Miao MY, Yeung W, Wong C, et al. A prospective 3-year longitudinal study of cognitive predictors of relapse in first-episode schizophrenic patients. Schizophr Res 2005 Sep 01;77(1):99-104. [doi: 10.1016/j.schres.2005.02.020] [Medline: 16005389]
5. Grove TB, Tso IF, Chun J, Mueller SA, Taylor SF, Ellingrod VL, et al. Negative affect predicts social functioning across schizophrenia and bipolar disorder: Findings from an integrated data analysis. Psychiatry Res 2016 Sep 30;243:198-206 [FREE Full text] [doi: 10.1016/j.psychres.2016.06.031] [Medline: 27416540]
6. Pew Research Center. Mobile fact sheet. 2019. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-06-05]
7. Hedges JH, Adolph KE, Amso D, Bavelier D, Fiez JA, Krubitzer L, et al. Play, attention, and learning: how do play and timing shape the development of attention and influence classroom learning? Ann N Y Acad Sci 2013 Jul;1292:1-20 [FREE Full text] [doi: 10.1111/nyas.12154] [Medline: 23763338]
8. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou C, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr Dis Treat 2016;12:357-373 [FREE Full text] [doi: 10.2147/NDT.S96649] [Medline: 26937191]
9. Knapp M. Costs of schizophrenia. Br J Psychiatry 1997 Dec;171:509-518. [doi: 10.1192/bjp.171.6.509] [Medline: 9519088]
10. National Institute for HealthCare Excellence. Psychosis and schizophrenia in adults: prevention and management (CG178). 2014. URL: https://www.nice.org.uk/guidance/cg178 [accessed 2019-02-25]
11. Mohamed S, Rosenheck R, He H, Yuping N. Insight and attitudes towards medication among inpatients with chronic schizophrenia in the US and China. Soc Psychiatry Psychiatr Epidemiol 2014 Jul;49(7):1063-1070. [doi: 10.1007/s00127-014-0824-1] [Medline: 24549835]
12. Saboo AV, Deshmukh PS, Deshmukh SB. Study on non-compliance in patients of schizophrenia. Journal of Evolution of Medical and Dental Sciences 2015 Jun 09;4(47):8112-8118. [doi: 10.14260/jemds/2015/1176]
13. Demjaha A, Egerton A, Murray R, Kapur S, Howes O, Stone JM, et al. Antipsychotic treatment resistance in schizophrenia associated with elevated glutamate levels but normal dopamine function. Biol Psychiatry 2014 Mar 01;75(5):e11-e13. [doi: 10.1016/j.biopsych.2013.06.011] [Medline: 23890739]
14. Demjaha A, Lappin J, Stahl D, Patel M, MacCabe J, Howes O, et al. Antipsychotic treatment resistance in first-episode psychosis: prevalence, subtypes and predictors. Psychol Med 2017 Apr 11;47(11):1981-1989. [doi: 10.1017/s0033291717000435] [Medline: 28395674]
15. Suenderhauf C, Walter A, Lenz C, Lang U, Borgwardt S. Counter striking psychosis: Commercial video games as potential treatment in schizophrenia? A systematic review of neuroimaging studies. Neurosci Biobehav Rev 2016 Sep;68:20-36. [doi: 10.1016/j.neubiorev.2016.03.018] [Medline: 27090742]
16. Desseilles M. Perspectives on Games, Computers, and Mental Health: Questions about Paradoxes, Evidences, and Challenges. Front Psychiatry 2016;7:122 [FREE Full text] [doi: 10.3389/fpsyt.2016.00122] [Medline: 27458390]
17. Allan S. Early Intervention in First Episode Psychosis: A Service User's Experience. Schizophr Bull 2017 Mar;43(2):234-235 [FREE Full text] [doi: 10.1093/schbul/sbv227] [Medline: 26757755]

XSL•FO
RenderX

18. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A Mobile Gaming Intervention to Increase Adherence to Antiretroviral Treatment for Youth Living With HIV: Development Guided by the Information, Motivation, and Behavioral Skills Model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e96 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
19. Thompson D. Designing serious video games for health behavior change: current status and future directions. J Diabetes Sci Technol 2012 Jul 01;6(4):807-811 [FREE Full text] [doi: 10.1177/193229681200600411] [Medline: 22920806]
20. Amado I, Brénugat-Herné L, Orriols E, Desombre C, Dos Santos M, Prost Z, et al. A Serious Game to Improve Cognitive Functions in Schizophrenia: A Pilot Study. Front Psychiatry 2016;7:64 [FREE Full text] [doi: 10.3389/fpsyt.2016.00064] [Medline: 27148093]
21. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious Games and Gamification for Mental Health: Current Status and Promising Directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
22. Lau HM, Smit JH, Fleming TM, Riper H. Serious Games for Mental Health: Are They Accessible, Feasible, and Effective? A Systematic Review and Meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
23. Vugts MA, Joosen MC, Mert A, Zedlitz AM, Vrijhoef HJ. Effectiveness of Serious Gaming During the Multidisciplinary Rehabilitation of Patients With Complex Chronic Pain or Fatigue: Natural Quasi-Experiment. J Med Internet Res 2018 Aug 15;20(8):e250 [FREE Full text] [doi: 10.2196/jmir.9739] [Medline: 30111527]
24. Chan CLF, Ngai EKY, Leung PKH, Wong S. Effect of the adapted Virtual Reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. Int J Geriatr Psychiatry 2010 Jun;25(6):643-649. [doi: 10.1002/gps.2403] [Medline: 19806599]
25. Oing T, Prescott J. Implementations of Virtual Reality for Anxiety-Related Disorders: Systematic Review. JMIR Serious Games 2018 Nov 07;6(4):e10965 [FREE Full text] [doi: 10.2196/10965] [Medline: 30404770]
26. Tashjian VC, Mosadeghi S, Howard AR, Lopez M, Dupuy T, Reid M, et al. Virtual Reality for Management of Pain in Hospitalized Patients: Results of a Controlled Trial. JMIR Ment Health 2017 Mar 29;4(1):e9 [FREE Full text] [doi: 10.2196/mental.7387] [Medline: 28356241]
27. Fernández-Aranda F, Jiménez-Murcia S, Santamaría JJ, Gunnard K, Soto A, Kalapanidas E, et al. Video games as a complementary therapy tool in mental disorders: PlayMancer, a European multicentre study. J Ment Health 2012 Aug;21(4):364-374 [FREE Full text] [doi: 10.3109/09638237.2012.664302] [Medline: 22548300]
28. Newzoo. Newzoo's 2018 Report: Insights Into the $137.9 Billion Global Games Market. 2018. URL: https://newzoo.com/insights/articles/newzoos-2018-report-insights-into-the-137-9-billion-global-games-market/ [accessed 2019-03-11]
29. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital Technology Use Among Individuals with Schizophrenia: Results of an Online Survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
30. Välimäki M, Kuosmanen L, Hätönen H, Koivunen M, Pitkänen A, Athanasopoulou C, et al. Connectivity to computers and the Internet among patients with schizophrenia spectrum disorders: a cross-sectional study. Neuropsychiatr Dis Treat 2017;13:1201-1209. [doi: 10.2147/NDT.S130818] [Medline: 28490882]
31. Noel VA, Acquilano SC, Carpenter-Song E, Drake RE. Use of Mobile and Computer Devices to Support Recovery in People With Serious Mental Illness: Survey Study. JMIR Ment Health 2019 Feb 20;6(2):e12255 [FREE Full text] [doi: 10.2196/12255] [Medline: 30785401]
32. Hilgard J, Engelhardt CR, Bartholow BD. Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (GAMES). Front Psychol 2013;4:608 [FREE Full text] [doi: 10.3389/fpsyg.2013.00608] [Medline: 24058355]
33. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
34. Shimizu N, Umemura T, Matsunaga M, Hirai T. An interactive sports video game as an intervention for rehabilitation of community-living patients with schizophrenia: A controlled, single-blind, crossover study. PLoS One 2017;12(11):e0187480 [FREE Full text] [doi: 10.1371/journal.pone.0187480] [Medline: 29131826]
35. Han DH, Renshaw PF, Sim ME, Kim JI, Arenella LS, Lyoo IK. The effect of internet video game play on clinical and extrapyramidal symptoms in patients with schizophrenia. Schizophr Res 2008 Aug;103(1-3):338-340. [doi: 10.1016/j.schres.2008.01.026] [Medline: 18304782]
36. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
37. Sherry JL, Lucas K, Greenberg BS, Lachlan K. Video game uses and gratifications as predictors of use and game preference. In: Vorderer P, Bryant J, editors. Playing video games: Motives, responses, and consequences. New York: Routledge; 2006:213-224.
38. Bacon N, Farnworth L, Boyd R. The Use of the Wii Fit in Forensic Mental Health: Exercise for People at Risk of Obesity. British Journal of Occupational Therapy 2012 Feb;75(2):61-68. [doi: 10.4276/030802212X13286281650992]

XSL•FO
RenderX

39. Palaus M, Marron EM, Viejo-Sobera R, Redolar-Ripoll D. Neural Basis of Video Gaming: A Systematic Review. Front Hum Neurosci 2017;11:248 [FREE Full text] [doi: 10.3389/fnhum.2017.00248] [Medline: 28588464]
40. Tang CS, Koh YW, Gan Y. Addiction to Internet Use, Online Gaming, and Online Social Networking Among Young Adults in China, Singapore, and the United States. Asia Pac J Public Health 2017 Nov;29(8):673-682. [doi: 10.1177/1010539517739558] [Medline: 29191049]
41. Polit DF, Beck CT. Essentials of nursing research: appraising evidence for nursing practice. 8th ed. Philadelphia: Lippincott Williams & Wilkins; 2014.
42. Leach MJ, Hofmeyer A, Bobridge A. The impact of research education on student nurse attitude, skill and uptake of evidence-based practice: a descriptive longitudinal survey. J Clin Nurs 2016 Jan;25(1-2):194-203. [doi: 10.1111/jocn.13103] [Medline: 26769207]
43. Zhang L, Gong R, Han Q, Shi Y, Jia Q, Xu S, et al. Survey of knowledge, attitude, and practice regarding reproductive health among urban men in China: a descriptive study. Asian J Androl 2015;17(2):309-314 [FREE Full text] [doi: 10.4103/1008-682X.142139] [Medline: 25532571]
44. Poreddi V, Gandhi S, Chandra R, Wilson A, Math SB. Smoking cessation support: Indian nursing students' practices, attitudes and perceived barriers. Br J Nurs 2015 Dec 10;24(22):1120-1124. [doi: 10.12968/bjon.2015.24.22.1120] [Medline: 26653512]
45. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
46. Draucker CB, Martsolf DS, Poole C. Developing distress protocols for research on sensitive topics. Arch Psychiatr Nurs 2009 Oct;23(5):343-350. [doi: 10.1016/j.apnu.2008.10.008] [Medline: 19766925]
47. Social Welfare Department. Integrated community centre for mental wellness (ICCMW). 2017. URL: https://www.swd.gov.hk/en/index/site_pubsvc/page_rehab/sub_listofserv/id_supportcom/id_iccmw/ [accessed 2019-02-25]
48. Social Welfare Department WD. Halfway house. 2017. URL: https://www.swd.gov.hk/en/index/site_pubsvc/page_rehab/sub_listofserv/id_residcare/id_halfwayhou/ [accessed 2019-02-25]
49. The Government of the Hong Kong Special Administrative Region. LCQ20: community support services for patients with mental illness. URL: https://www.info.gov.hk/gia/general/201612/14/P2016121400645.htm [accessed 2019-02-25]
50. Tse S, Cheung E, Kan A, Ng R, Yau S. Recovery in Hong Kong: service user participation in mental health services. Int Rev Psychiatry 2012 Feb;24(1):40-47. [doi: 10.3109/09540261.2011.646244] [Medline: 22385425]
51. Bavelier D, Achtman RL, Mani M, Föcker J. Neural bases of selective attention in action video game players. Vision Res 2012 May 15;61:132-143 [FREE Full text] [doi: 10.1016/j.visres.2011.08.007] [Medline: 21864560]
52. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
53. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005 Feb;19(1):231-240. [doi: 10.1519/15184.1] [Medline: 15705040]
54. Holcomb ZC. Fundamentals of descriptive statistics. Oxford: Taylor and Francis; 1997.
55. Harkness J, Pennell B, Schoua-Glusberg A. Survey questionnaire translation and assessment. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al, editors. Methods for Testing and Evaluating Survey Questionnaires. New Jersey: John Wiley & Sons, Inc; 2004:453-473.
56. CBS Interactive Inc. GameFAQs. 2018. URL: https://gamefaqs.gamespot.com/ [accessed 2019-02-25]
57. Chang LYC, Poon R. Internet Vigilantism: Attitudes and Experiences of University Students Toward Cyber Crowdsourcing in Hong Kong. Int J Offender Ther Comp Criminol 2017 Dec;61(16):1912-1932. [doi: 10.1177/0306624X16639037] [Medline: 26992831]
58. Zhu JJH, He Z. Diffusion, Use and Impact of the Internet in Hong Kong: A Chain Process Model. J Comput Mediat Commun 2002;7(2):724. [doi: 10.1111/j.1083-6101.2002.tb00139.x]
59. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
60. Athanasopoulou C, Välimäki M, Koutra K, Löyttyniemi E, Bertsias A, Basta M, et al. Internet use, eHealth literacy and attitudes toward computer/internet among people with schizophrenia spectrum disorders: a cross-sectional study in two distant European regions. BMC Med Inform Decis Mak 2017 Sep 20;17(1):136 [FREE Full text] [doi: 10.1186/s12911-017-0531-4] [Medline: 28931385]
61. Taylor D, Paton C, Kapur S. The Maudsley prescribing guidelines in psychiatry. 12th edition. New Jersey: Wiley-Blackwell; 2015.
62. Wang C, Chan CLW, Mak K, Ho S, Wong PWC, Ho RTH. Prevalence and correlates of video and internet gaming addiction among Hong Kong adolescents: a pilot study. Scientific World Journal 2014;2014:874648 [FREE Full text] [doi: 10.1155/2014/874648] [Medline: 25032242]
63. Allcorrect Games. Hong Kong Mobile Game Market. 2019. URL: https://allcorrectgames.com/insights/mobile-game-market-index/hong-kong/ [accessed 2019-05-14]

XSL•FO
RenderX

64. Serper M, Payne E, Dill C, Portillo C, Taliercio J. Allocating effort and anticipating pleasure in schizophrenia: Relationship with real world functioning. Eur Psychiatry 2017 Oct;46:57-64. [doi: 10.1016/j.eurpsy.2017.07.008] [Medline: 29031122]
65. Bejjanki VR, Zhang R, Li R, Pouget A, Green CS, Lu Z, et al. Action video game play facilitates the development of better perceptual templates. Proc Natl Acad Sci U S A 2014 Nov 25;111(47):16961-16966 [FREE Full text] [doi: 10.1073/pnas.1417056111] [Medline: 25385590]
66. Howard-Jones PA, Jay T. Reward, learning and games. Current Opinion in Behavioral Sciences 2016 Aug;10:65-72. [doi: 10.1016/j.cobeha.2016.04.015]
67. Gooch P, Living R. The Therapeutic Use of Videogames within Secure Forensic Settings: A Review of the Literature and Application to Practice. British Journal of Occupational Therapy 2016 Nov 05;67(8):332-341. [doi: 10.1177/030802260406700802]
68. Hagström D, Kaldo V. Escapism among players of MMORPGs--conceptual clarification, its relation to mental health factors, and development of a new measure. Cyberpsychol Behav Soc Netw 2014 Jan;17(1):19-25. [doi: 10.1089/cyber.2012.0222] [Medline: 24003967]
69. Blasi MD, Giardina A, Giordano C, Coco GL, Tosto C, Billieux J, et al. Problematic video game use as an emotional coping strategy: Evidence from a sample of MMORPG gamers. J Behav Addict 2019 Mar 01;8(1):25-34 [FREE Full text] [doi: 10.1556/2006.8.2019.02] [Medline: 30739460]
70. Colder Carras M, Kalbarczyk A, Wells K, Banks J, Kowert R, Gillespie C, et al. Connection, meaning, and distraction: A qualitative study of video game play and mental health recovery in veterans treated for mental and/or behavioral health problems. Soc Sci Med 2018 Nov;216:124-132 [FREE Full text] [doi: 10.1016/j.socscimed.2018.08.044] [Medline: 30257787]
71. Scharrer E, Kamau G, Warren S, Zhang C. Violent video games do contribute to aggression. In: Ferguson C, editor. Video game influences on aggression, cognition, and attention. Switzerland: Springer International Publishing; 2018:5-21.
72. Turel O. Videogames and guns in adolescents: Tests of a bipartite theory. Comput Human Behav 2020 Aug;109:106355 [FREE Full text] [doi: 10.1016/j.chb.2020.106355] [Medline: 32372846]
73. Prescott AT, Sargent JD, Hull JG. Metaanalysis of the relationship between violent video game play and physical aggression over time. Proc Natl Acad Sci U S A 2018 Oct 02;115(40):9882-9888 [FREE Full text] [doi: 10.1073/pnas.1611617114] [Medline: 30275306]
74. Teng Z, Nie Q, Guo C, Zhang Q, Liu Y, Bushman BJ. A longitudinal study of link between exposure to violent video games and aggression in Chinese adolescents: The mediating role of moral disengagement. Dev Psychol 2019 Jan;55(1):184-195. [doi: 10.1037/dev0000624] [Medline: 30335433]
75. Hodge SE, Taylor J, McAlaney J. It's Double Edged: The Positive and Negative Relationships Between the Development of Moral Reasoning and Video Game Play Among Adolescents. Front Psychol 2019;10:28 [FREE Full text] [doi: 10.3389/fpsyg.2019.00028] [Medline: 30740072]
76. Ferguson CJ, Rueda SM. The Hitman Study. European Psychologist 2010 Jan;15(2):99-108. [doi: 10.1027/1016-9040/a000010]
77. Olson CK, Kutner LA, Warner DE. The Role of Violent Video Game Content in Adolescent Development. Journal of Adolescent Research 2008 Jan;23(1):55-75. [doi: 10.1177/0743558407310713]
78. Martončik M, Lokša J. Do World of Warcraft (MMORPG) players experience less loneliness and social anxiety in online world (virtual environment) than in real world (offline)? Computers in Human Behavior 2016 Mar;56:127-134. [doi: 10.1016/j.chb.2015.11.035]
79. Molyneux L, Vasudevan K, Gil de Zúñiga H. Gaming Social Capital: Exploring Civic Value in Multiplayer Video Games. J Comput-Mediat Comm 2015 May 09;20(4):381-399. [doi: 10.1111/jcc4.12123]
80. Trepte S, Reinecke L, Juechems K. The social side of gaming: How playing online computer games creates online and offline social support. Computers in Human Behavior 2012 May;28(3):832-839. [doi: 10.1016/j.chb.2011.12.003]
81. DeRosier M, Thomas J. Video Games and Their Impact on Teens Mental Health. In: Moreno M, Radovic A, editors. Technology and Adolescent Mental Health. Swizerland: Springer; 2018:237-253.
82. Bennis L, Benhlima S. J Theor Appl Inform Tech. A new approach to design an attractive game based learning in various domains. 2016. URL: http://www.jatit.org/volumes/Vol85No3/10Vol85No3.pdf [accessed 2020-04-05]
83. Schadenberg B, Neerincx M, Cnossen F, Looije R. Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. Cognitive Systems Research 2017 Jun;43:222-231. [doi: 10.1016/j.cogsys.2016.08.003]
84. Harteveld C, Sutherland S. Personalized gaming for motivating social and behavioral science participation. In: HUMANIZE '17: Proceedings of the 2017 ACM Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces. New York: Association for Computing Machinery; 2017 Presented at: IUI'17: 22nd International Conference on Intelligent User Interfaces; March 2017; Limassol Cyprus p. 3-38. [doi: 10.1145/3039677.3039681]
85. Király O, Tóth D, Urbán R, Demetrovics Z, Maraz A. Intense video gaming is not essentially problematic. Psychol Addict Behav 2017 Nov;31(7):807-817. [doi: 10.1037/adb0000316] [Medline: 28956935]
86. Kumar V, Sattar Y, Bseiso A, Khan S, Rutkofsky IH. The Effectiveness of Internet-Based Cognitive Behavioral Therapy in Treatment of Psychiatric Disorders. Cureus 2017 Aug 29;9(8):e1626 [FREE Full text] [doi: 10.7759/cureus.1626] [Medline: 29098136]

XSL•FO
RenderX

87. Newman K, Wang AH, Wang AZY, Hanna D. The role of internet-based digital tools in reducing social isolation and addressing support needs among informal caregivers: a scoping review. BMC Public Health 2019 Nov 09;19(1):1495 [FREE Full text] [doi: 10.1186/s12889-019-7837-3] [Medline: 31706294]
88. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
89. Anttila M, Välimäki M, Koivunen M, Luukkaala T, Kaila M, Pitkänen A, et al. Adoption of an Internet-based patient education programme in psychiatric hospitals. J Psychiatr Ment Health Nurs 2011 Dec;18(10):914-923. [doi: 10.1111/j.1365-2850.2011.01765.x] [Medline: 22070578]
90. Kurki M, Anttila M, Koivunen M, Marttunen M, Välimäki M. Nurses' experiences of the use of an Internet-based support system for adolescents with depressive disorders. Inform Health Soc Care 2018 Sep;43(3):234-247. [doi: 10.1080/17538157.2016.1269110] [Medline: 28139155]
91. Kurki M, Koivunen M, Anttila M, Hätönen H, Välimäki M. Usefulness of internet in adolescent mental health outpatient care. J Psychiatr Ment Health Nurs 2011 Apr;18(3):265-273. [doi: 10.1111/j.1365-2850.2010.01661.x] [Medline: 21395918]
92. Kendrick KM, Yao D. Can Computer-Based Cognitive Therapy Become a Front-Line Option for Prevention and Treatment of Mental Disorders? Am J Psychiatry 2017 Apr 01;174(4):303-304. [doi: 10.1176/appi.ajp.2017.16121439] [Medline: 28366093]

Abbreviations

GAMES: Gaming Attitudes, Motivations, and Experiences Scales

Edited by G Eysenbach, W Curioso; submitted 29.05.19; peer-reviewed by T Campellone, F Fernandez-Aranda; comments to author 03.03.20; revised version received 25.03.20; accepted 14.05.20; published 22.07.20.

Please cite as:
Choi WTH, Yu DKS, Wong T, Lantta T, Yang M, Välimäki M
Habits and Attitudes of Video Gaming and Information Technology Use in People with Schizophrenia: Cross-Sectional Survey
J Med Internet Res 2020;22(7):e14865
URL: http://www.jmir.org/2020/7/e14865/
doi:10.2196/14865
PMID:32459646

©William TH Choi, Dan KS Yu, Terry Wong, Tella Lantta, Min Yang, Maritta Välimäki. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mobile Mental Health Apps in China: Systematic App Store Search

Huifang Yin[1,2], PhD; Klaas J Wardenaar[2], PhD; Yuhao Wang[1], MSc; Nan Wang[1], MSc; Wenjin Chen[1], MSc; Yan Zhang[1], MSc; Guangming Xu[1], MD; Robert A Schoevers[2], MD, PhD

[1]Tianjin Anding Hospital, Tianjin, China

[2]Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Guangming Xu, MD
Tianjin Anding Hospital
No 13, Liulin Road, Hexi district
Tianjin
China
Phone: 86 15122997716
Email: xugm@ymail.com

Abstract

Background: Smartphones have become ubiquitous in China, offering a promising way to deliver mental health interventions; however, little is known about the current use and characteristics of smartphone apps for mental health.

Objective: The purpose of this study was to gain insight into mobile mental health apps available in China as of December 2018.

Methods: A systematic search was conducted to identify and evaluate the most downloaded apps from iOS and Android platforms. Apps were categorized according to their main purpose and downloaded to evaluate their content. Each app's affiliation, cost, target users, information security, and evidence-based nature were evaluated.

Results: Of the 172 unique apps that were identified, there were 37 apps (21.5%) for psychological counseling, 50 apps (29.1%) for assessment, 12 apps (7.0%) to relieve stress, 24 apps (14.0%) for psychoeducation, and 49 (28.4%) multipurpose apps (ie, a combination of counseling and assessment). Most apps were developed for adults in the general population (166/172, 96.5%), rather than for psychiatric patients. App-based counseling was mostly provided by psychologists, and of the assessed apps, only 40% (70/172) used evidence-based scales to assess mental health problems such as anxiety or depressed mood. Guided meditation was used as the main technique in stress-relieving apps.

Conclusions: Many apps contained useful and evidence-based elements, such as good quality information, validated measurements, and useful meditation methods; however, for mobile apps to contribute significantly to mental health care in China, considerable challenges remain, including the need for more patient-focused apps that can actually take on the role of a health care provider. In addition, efficacy studies are needed.

(J Med Internet Res 2020;22(7):e14915) doi:10.2196/14915

KEYWORDS

smartphone; app; mobile mental health; mental illness; mental health problem; China

Introduction

Mental health problems cause significant distress and have a negative impact on social relationships, school performance, occupational attainment, and physical health [1]. Mental health disorders have been found to affect 12.0% to 47.4% of populations worldwide [2]. In China, the lifetime prevalence of mental health disorders has been found range from 16.6% [3] to 23.6% [4]. Unfortunately, the availability of mental health services in the existing Chinese health care system is currently limited, and many patients cannot obtain appropriate treatment or help for their mental health problems [5,6]. In addition to the lack of available facilities and trained professionals, a low perceived need, poor knowledge of mental health disorders, and mental illness stigma have been identified as potential barriers to seeking help for mental illness [5].

Mobile health (mHealth) technologies could help to partly overcome the above-mentioned obstacles to seeking help, as they could extend the reach of mental health care beyond available health care providers and clinics. The smartphone was

XSL•FO
RenderX

one of the most rapidly adopted technological innovations in history [7]; it now provides ubiquitous internet access and the possibility to access, download, and run externally created software apps. As such, smartphone technology provides a unique opportunity to deliver cost-effective and evidence-based mental health care services to large groups of people [8]. Indeed, studies have shown that mental health care apps can play an important role in the assessment, prediction, and monitoring of mental health, as well as psychoeducation delivery, self-management strategies, recovery support, prevention, and promotion. In addition, apps can be used for the training of mental health providers [9-16].

The large number of smartphone users in China—estimated to have reached 748.3 million in 2019 [17]—as well as the large mobile health market in China—which grew by 74.5% between 2016 and 2017 [18]—indicates a large potential for mental health care apps in China; however, very few studies have been done to gain insight into the number and types of available apps and their potential effectiveness in China.

On the one hand, available literature [19-23] seems to suggest limited availability and use of mental health–related apps in China. For instance, a review study [19] on health apps in China found that there were many apps targeting both nonprofessional and professional users, but the study described only 1 app related to mental health. Another review [24] of mHealth technology for the treatment or prevention of mental health disorders in low-and middle-income countries found 6 studies about mHealth technology in China, among which only 1 applied to mental health. On the other hand, an examination of the Chinese iOS and Android app stores shows that, despite a scarcity of research, there is an abundance of publicly available mental health apps. Unfortunately, consumers currently have no access to information about the quality of the available apps beyond the star rating system and user reviews on the app stores. It is unlikely that these indicators of popularity also reflect the quality, effectiveness, or evidence base of an app. Previous studies [20-23] outside of China have found that most mental health apps in commercial marketplaces do not provide evidence-based therapies, do not follow clinical guidelines, and do not respect privacy regulations with respect to the handling of personal information. Consequently, it is a challenge for both patients and clinicians to find useful apps when needed. A recent search of 44 mental health apps (available as of October 2017) in China showed some common features, such as commercial purpose, and services that included counseling, education, and assessment [25]; however, since this review, the market for mental health apps has grown considerably larger (eg, more than 100 available apps were found based on the search term representing mental in the Baidu app store) [26]. In addition, the important characteristics of the apps that are available have remained unclear, such as the target user population, the evidence base for their content, and consideration of data-safety issues. To determine if and how apps can play a role in Chinese mental health care delivery, these aspects should be thoroughly explored, and it is for this reason that this systematic search was undertaken.

The purpose of this study was to (1) characterize the purpose and content of the mental health smartphone apps available for use and most downloaded by the general Chinese public, (2) to evaluate whether app content is evidence-based, and (3) to gain insight into the app costs and into the quality and comprehensiveness of data safety in the apps.

Methods

mHealth App Market Search

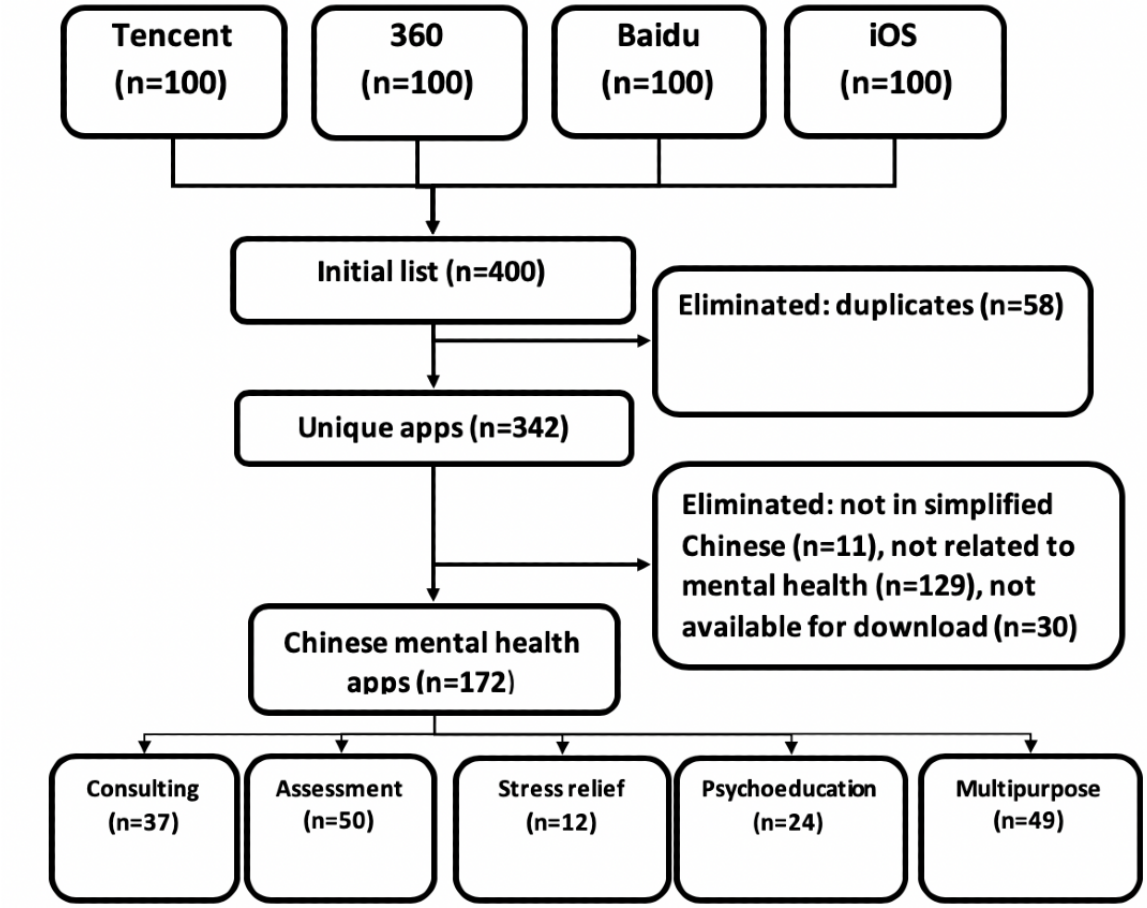
On December 26, 2018, we conducted a search in both Android (Google LLC) and iOS (Apple Inc) app stores in China. The 3 largest Android app stores in China are Tencent (Tencent Holding Ltd), 360 (Qihoo 360 Technology Co Ltd), and Baidu (Baidu Inc), and these were selected for the study because they constitute more than 50% of Chinese Android market share [27]. In addition, we used the Chinese iOS App store to gather a list of apps for the Chinese iOS market.

We restricted our search to apps that used the Chinese characters for 1 or more of the following keywords in the title or store description of the app: *(mental health)* OR *(mental)* OR *(psychology)* OR *(psychiatry)* OR *(psychological intervention)* OR *(psychological prevention)*. The top 100 apps were collected for each app store.

Android apps are free to download, and iOS apps are either free or paid. We listed free and paid apps together, mainly because there were few paid apps. The eventual selection of apps for this study (Figure 1) was done according to the following inclusion criteria: The app should (1) be related to mental health, (2) be available for download, (3) use simplified Chinese so that any literate Chinese individual can easily understand the content of the app, and (4) be usable in mainland China.

XSL•FO
RenderX

Figure 1. Chinese mental health app inclusion and exclusion flowchart.



App Assessment

The apps that met the inclusion criteria were downloaded onto either a Huawei mate 9 (Android version 8.0.0), a Xiaomi 8 (Android version 8.0.0), or an iPhone 5s (iOS version 7.1) device for a complete assessment. The reason for downloading each app for inspection was that the app description may not be sufficient to allow for a full judgment of the nature and quality of its content. For instance, in a previous study [23], when the authors downloaded the apps, they found that some apps provided patients with poor or potentially harmful advice (such as a recommendation to people with a manic episode to drink alcohol before bedtime to assist with sleeping). Therefore, for each app, we evaluated if the information provided was evidence-based. For apps that were listed in multiple app stores, the consistency of the descriptions across stores was evaluated, and identical duplicates were eliminated. Some apps had a patient version and a clinician version. We assessed these versions separately because they might provide different services. If an app was available for both iOS and Android, the Android version was evaluated.

For each app, the following general information about the app was extracted (Table 1): (1) main purpose, (2) affiliation (ie, commercial, medical center, or university), (3) cost, (4) target user, (5) whether or not information security was emphasized, and (6) whether the therapeutic contents were based on an evidence-based psychological theory, such as cognitive behavior therapy [28], acceptance and commitment therapy [29], interpersonal psychotherapy [30], or problem-solving therapy [31], or if the content was based on advice by professional certified staff including psychologists, psychotherapists, or psychiatrists. Based on the app description, the first author categorized the apps into 1 of 5 categories indicating their main purpose: psychological counseling, assessment, stress relief, psychoeducation, and multipurpose (incorporating components of assessment, education, psychological counseling, and stress relief). Each category of apps was then further investigated by 1 of the coauthors (HY, YW, NW, YZ, and WC; each author was randomly assigned to a category). If anyone found that the initially selected category did not fit an app, the 5 authors discussed until consensus was reached. For each category, specific additional information was extracted (detailed in Table 1). For mental health education apps, core aspects of mental health information (Table 2) were investigated. These aspects have previously been proposed by the National Health Commission of China [32] and are used to guide mental health education work in the whole country.

XSL•FO
RenderX

Table 1. The information collected for each of the selected apps.

Information extracted and content	Values
General information collected for each app	
Affiliation	Commercial, university, or medical center.
Cost	Pay for download, completely free, or partly free.
Target user	Children or adolescents, general public adults, specific for women, special for patients with mental health problems, or specific for professional staff.
Data safety information	Whether an app report relevant information security measures or not? (yes or no).
Professional background	Whether the contents of the app were based on a psychological theory or were based on advice by professional certified staff including psychologists, psychotherapists or psychiatrists? (yes or no).
Specific information for psychological counseling apps	
Counseling mode	Online (text typing or audio) or offline (eg, an app can help to make an appointment).
Counseling provider	Professional (including psychologists, psychotherapists, or psychiatrists) or trained lay health supporter.
Target users	Individuals with diagnoses of a mental health disorder or general public.
Specific information for assessment apps	
Assessment aspect	Personality, symptoms, intelligence, cognitive functions, or assessment just for entertainment (eg, constellations, fortune, or other).
Evidence	Are evidence-based measurements or scales used? (yes or no)
Specific information for stress relief apps	
Methods	Music, mindfulness, relaxation therapy, etc.
Evidence	Is there a reference in the app to work that shows the therapy is effective?
Specific information for education apps	
Apps for training mental health providers	Books, lectures, audio recordings, or other method.
Apps for mental health knowledge to public	According to 8 core information aspects (see Table 2).

Table 2. Eight aspects of mental health information.

Items	Content
1	Mental health is an integral part of health. No mental illness does not necessarily mean mental health. Everyone needs not only physical health, but mental health as well.
2	Mental health and mental illness, such as physical health and physical illness, are determined by multiple interacting biological, psychological, and social factors.
3	Given that anyone could encounter mental health problems in their life, it is necessary to pay attention to and maintain mental health.
4	Schizophrenia, depression, behavior disorders in children and adolescents, and senile dementia are China's current priorities for mental illness prevention and treatment.
5	A person suffering from psychological or behavior problems or mental illness should seek help from a medical institution as soon as possible to receive advice and formal diagnosis and treatment.
6	Mental illness can be prevented and treated.
7	Care for and do not discriminate against mentally ill patients, and help them return to their families, communities, and society.
8	Mental health is related to social harmony and development. Promotion of mental health and prevention and treatment of mental illness is the responsibility of society as a whole.

Results

General

Of the 400 apps that were screened, 172 unique apps met the inclusion criteria (Figure 1). Of these, 105 (61.0%) were available in the Android app stores, 28 (16.3%) were available in the iOS app store, and 39 (37.1%) were available in both the iOS and Android app stores. Based on the descriptions of the main purpose of the apps, 21.5% (37/172) were for psychological counseling, 29.1% (50/172) were for assessment, 7.0% (12/172) were for stress relief, 14.0% (24/172) were for psychoeducation, and 28.4% (49/172) were multipurpose apps. The number of downloads per app ranged from 16 to 107 million

XSL•FO
RenderX

in the Android store. Download numbers could not be obtained in the iOS app store. Detailed information about the different groups of apps is shown in Table 3.

Counseling Apps

There were 37 counseling apps, and they were all developed by commercial companies. Of the counseling apps, 7 (19%) apps were free to use, 29 (78%) apps could be downloaded for free but charged counseling fees to consumers wishing to use the app, and 1 (3%) app charged a download fee. In terms of the target population, 8 (22%) apps were aimed at adults and adolescents, 1 (3%) was aimed specifically at children and adolescents, and 27 (76%) were only aimed at adults. Of the 27 apps for adults, 1 (3%) app was only aimed at women and 3 (8%) apps were aimed specifically at professionals. Of the 37 counseling apps, 20 (54%) apps made explicit claims about information security measures. All of the counseling apps provided some kind of online consultation, which meant that staff was available online to give users feedback and consultation. Counseling by text-based messages was used in 34 (92%) apps. Users could voice-chat with staff in 26 (70%) apps. Of the 37 counseling apps, 20 (54%) provided services to make an appointment with a counselor and 4 (11%) provided services to register appointments at hospitals; 36 (97%) had professionals acting as consultation providers (eg, psychologists, psychotherapists, or psychiatrists), and only 1 (3%) app provided consultation by a trained lay health supporter.

Assessment Apps

Of the 50 assessment apps, 30 (60%) apps provided information about the affiliations of their developers and all were developed by companies; 44 (88%) were free to use, and the other 6 (12%) were free to download but charged a fee for some of the assessments that were offered. All apps were aimed at adults, and 3 (6%) apps were specifically for women. Only 5 (10%) apps made a claim about information security. Evidence-based scales were used in 17 (34%) apps (some had more than 1 type—12 assessed personality, 8 assessed emotional symptoms, and 1 assessed cognitive ability). The most popular validated questionnaires were the Eysenck personality questionnaire [33,34] used in 8 apps and the Sixteen Personality Factor Questionnaire [35,36] used in 4 apps. The Self-Rating Anxiety Scale [37,38] and Self-Rating Depression Scale [39,40] were used to assess anxious and depressive symptoms in 10 apps. The other 33 (66%) apps were mainly aimed at providing some form of entertainment; they assessed fortunes, mental age, marriage and love, and occupational abilities using assessment items of unknown source.

Stress Relief Apps

All 12 stress relief apps were developed by companies; 3 (25%) apps were free, and 9 (75%) apps were free to download but charged a fee for some of the content. Of the 12 stress relief apps, only 1 (8%) app made a claim about information security. Of the 12 stress relief apps, 9 (75%) provided evidence showing that meditation is an effective method to relieve stress. Audio was the main medium used in stress relief apps; of the 12 stress relief apps, 1 (8%) app played audio recordings of methods to maintain mental stability from the *Huangdi Neijing* (an ancient Chinese medical text that has been treated as the fundamental doctrinal source for Chinese medicine for more than two millennia), 1 (8%) provided audio of mood control courses, 2 (17%) provided nature sound recordings, and 7 (58%) apps provided audio recordings of guided meditations. Of the 12 stress relief apps, 2 apps (17%) were specifically aimed at children and adolescents and focused on improving their concentration and memory and on relieving examination anxiety, and only 1 (8%) provided an online community function, allowing users to communicate with each other to relieve stress. Meditation apps used different techniques such as respiration exercises (5/7, 71%), gradual relaxation exercises (7/7, 100%), and biofeedback therapy (1/7, 14%).

Psychoeducation Apps

All 24 psychoeducation apps were developed by commercial companies. Of the 24 psychoeducation apps, 23 (96%) were for adults of which 4 (17%) were targeted at mental health professionals and provided training courses or e-books about mental health, and the other 19 (79%) were aimed at the general population to provide education and knowledge about psychology and mental health. All of these apps were free to download; however, 3 of the apps for professionals charged a fee for some of the courses in the apps. None of the psychoeducation apps made a claim about data security. Of the apps aimed at the general population, 11% (2/19) were e-books about behavioral psychology and general psychology, 5% (1/19) provided a list of psychological website links, 5% (1/19) provided audio recordings about mental health, 5% (1/19) provided case descriptions showing users the how mental health disorders can manifest, 5% (1/19) provided information about the diagnostic process of mental health disorders, and 74% (14/19) apps included psychological popular science articles, in which common mental health problems in parenting, in the workplace, and in marriage and love were covered, and coping methods were described. We scanned the articles in the 14 apps to check whether they referred to any of the eight aspects of mental health information and found that 10 out of 14 referred at least one aspect, and 9 out of 14 apps covered all of the eight aspects.

XSL•FO
RenderX

Table 3. The characteristics of apps in different categories.

Content characteristics	Counseling (n=37), n (%)	Assessment (n=50), n (%)	Stress relief (n=12), n (%)	Psychoeducation (n=24), n (%)	Multipurpose (n=49), n (%)
Affiliation					
Commercial company	33 (89)	30 (60)	11 (92)	14 (58)	43 (88)
University	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)
No information	4 (11)	20 (40)	1 (8)	10 (42)	5 (10)
Cost					
Pay for download	1 (3)	0 (0)	0 (0)	0 (0)	0 (0)
Fully free	7 (19)	44 (88)	3 (25)	20 (83)	10 (20)
Partly free	29 (78)	6 (12)	9 (75)	4 (17)	39 (80)
Target user					
Adults and children	8 (22)	0 (0)	2 (17)	0 (0)	7 (14)
Only children or adolescents	1 (3)	0 (0)	0 (0)	1 (4)	4 (8)
Only adults					
General adults	24 (65)	47 (94)	10 (83)	19 (79)	37 (76)
Women	1 (3)	3 (6)	0 (0)	0 (0)	0 (0)
Patients with mental health problems	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Professional staff	3 (8)	0 (0)	0 (0)	4 (17)	1 (2)
Data safety information					
Yes	20 (54)	5 (10)	1 (8)	0 (0)	20 (41)
No	17 (46)	45 (90)	11 (92)	24 (100)	29 (59)
Professional background or evidence-based					
Yes	36 (97)	17 (34)	9 (75)	24 (100)	45 (92)
No	1 (3)	33 (66)	3 (25)	0 (0)	4 (8)
Counseling mode					
Only online	13 (35)	—	—	—	24 (49)
Only offline	0 (0)	—	—	—	0 (0)
Both	24 (65)	—	—	—	15 (31)
None	0 (0)	—	—	—	10 (20)
Counseling provider					
Professionals	36 (97)	—	—	—	37 (76)
Trained lay health supporter	1 (3)	—	—	—	2 (4)
Assessment aspect[a]					
Personality	—	12 (24)	—	—	21 (43)
Emotional symptoms	—	10 (20)	—	—	33 (67)
Intelligence	—	0 (0)	—	—	0 (0)
Cognitive function	—	1 (2)	—	—	0 (0)
Entertainment	—	36 (72)	—	—	22 (45)
Stress relief methods					
Music	—	—	2 (17)	—	4 (8)
Meditation	—	—	7 (58)	—	6 (12)
Course	—	—	1 (8)	—	1 (2)
Audio books	—	—	1 (8)	—	1 (2)

XSL•FO
RenderX

Content characteristics	Counseling (n=37), n (%)	Assessment (n=50), n (%)	Stress relief (n=12), n (%)	Psychoeducation (n=24), n (%)	Multipurpose (n=49), n (%)
Online community	—	—	1 (8)	—	2 (4)
Psychoeducation content (eight aspects)[b]					
Mental health is an integral part of health	—	—	—	10 (42)	35 (71)
Multiple interacting factors determinants	—	—	—	9 (38)	26 (53)
Anyone could encounter a mental health problems in their life	—	—	—	10 (42)	32 (65)
China's current priority mental illnesses	—	—	—	10 (42)	26 (53)
Go to the medical institutions for treatment	—	—	—	10 (42)	33 (67)
Mental illness can be prevented and treated	—	—	—	10 (42)	32 (65)
Care for and do not discriminate against mentally ill patients	—	—	—	10 (42)	30 (61)
Promotion of mental health is the joint responsibility of society	—	—	—	10 (42)	30 (61)

[a]These subheading values do not add to n or 100% because it was possible for an app to assess more than one aspect.

[b]These subheading values do not add to n or 100% because it was possible for an app to address more than one item.

Multipurpose Apps

Of the 49 multipurpose apps, 44 apps (90%) provided affiliation information; 1 (2%) was developed by the Institute of Psychology of Chinese Academy of Sciences, and the other 43 (88%) were developed by companies. Of the 49 apps, 38 (78%) were mainly aimed at adults, 7 (14%) were for individuals of any age, and 4 (8%) were for children and adolescents. Of the 38 apps for adults, there was an app (*Guan Ai ++*) for people in the workplace, an app (*Guan Ai Ji Ceng Gan Bu*) for community-level officials, and an app (*Hao Xin Qing*) that had a version for clients and a version for clinicians. Of the 49 multipurpose app, data safety claims were found for 20 (41%) apps. All the multipurpose apps could be downloaded for free, but users were charged an additional fee for some content in 39 apps (80%). Most multipurpose apps (45/49, 92%) involved professional staff, evidence-based therapies, or references to evidence-based materials.

Consulting, assessment, psychoeducation, and stress relief were combined in 5 (10%) multipurpose apps; consulting, assessment, and stress relief were combined in 2 apps (4%); consulting, assessment, and psychoeducation were combined in 19 (39%) apps; none combined counseling and assessment; counseling and psychoeducation were combined in 5 (10%), assessment and stress relief were combined in 2 (4%), assessment and stress relief were combined in 2 (4%), and stress relief and psychoeducation were combined in 3 (6%).

Of the 49 multipurpose apps, 39 apps (80%) provided some kind of online counseling; with the exception of 2 apps, counseling services were provided by mental health professionals (31 apps involved a psychologist or psychological therapist; 6 apps involved a psychiatrist). In several apps, users could choose which professional to have counseling with, after looking through the profiles of listed professionals. Depending on the app, users could contact the professional by telephone, online video or voice chat, text-based messages, or they could make an appointment with the professional; users had to pay a fee, depending on the price of the listed professional. Two apps offered online consults with a professional but also included a system to make an appointment with a professional at a hospital. One app (*Hao Xin Qin doctor version*) aimed to help professionals to manage their clients online.

Of the 49 multipurpose apps, 41 apps (84%) had some sort of assessment functionality; 80% (33/41) used evidence-based self-report questionnaires to evaluate personality (2 used the Eysenck personality questionnaire, and 1 used the Sixteen Personality Factor Questionnaire), anxiety (14 apps used the Self-Rating Anxiety Scale), and depression (26 apps use the Self-Rating Depression Scale, 1 app uses the Hamilton depression rating scale [41], and 1 used the Symptom Checklist-90 [42]), or social phobia (items from the Liebowitz social anxiety scale)[43]. Some entertainment assessments such as fortune, love, and marriage were also included in 23 apps. None of the apps aimed to assess intelligence or cognitive abilities.

Of the 49 multipurpose apps, 14 (29%) of the multipurpose apps had stress relieving sections and provided audio recordings of relaxing music or meditation guidance, and 36 apps provided either psychoeducation articles or audio recordings related to the eight key aspects of mental health knowledge.

There were 4 (8%) multipurpose apps specifically for children and adolescents; 1 app (*jie you nuan xin mao*) focused on assessing everyday mood, providing counseling services, and had the option for discussion with fellow users; 1 app (*xin li mei*) was developed specifically for students in a specific region's primary and secondary schools, providing multiple services including assessment, stress relief, counseling, and making appointments with a teacher-counselor; 1 app (*Q xin li*) provided counseling, questions and answers, microlectures, and psychoeducational articles for users; and 1 app (*Gao kao jian ya bao*) was developed to help students to reduce stress

XSL•FO
RenderX

induced by college entrance examinations by providing counseling, stress-reducing music, and assessment.

Discussion

Principal Findings

To our knowledge, this is the first study to search and evaluate the contents of the full range of available and most widely used mental health smartphone apps in China. Evaluation of each of the apps showed different primary aims, with most focusing on assessment, counseling, and a combination of multiple purposes (ie, assessment, counseling, stress relief, and psychoeducation). Most apps were developed for profit and focused on the adult population. Only 6 apps were specifically aimed at children or adolescents. A majority of the apps provided counseling or assessment, often enlisting online psychological counseling services from professionals such as psychologists, psychotherapists, or psychiatrists; however, even though psychometrically valid assessments of personality, anxiety, and depression were included in some assessment apps, assessments in most apps were more for entertainment than for assessment of psychiatric symptoms. Audio-recordings to guide or accompany meditation were the main materials provided by the apps that focused on stress reduction. Most psychoeducation apps provided information that aligned with national guidelines on mental health information.

Evaluation of the target populations for the reviewed apps provided clear insight into the perceived market opportunities by app developers. This study showed that developers of mental health apps prefer to target the general population rather than patients with mental health disorders: No reviewed app focused specifically on patients with mental health disorders. In addition, although there was a small number of mental health apps for specific adult populations (eg, women, workplace, staff) or for mental health professionals, most apps were aimed at a broad audience. This is understandable given that most developers will want to acquire as many users for their app as possible, but this does leave room for apps that focus on assisting patients with mental health disorders. Interestingly, only a few apps were developed for children and adolescents. This does not reflect the current situation, where the prevalence of mental health disorders in children and adolescents in China lies close to the high worldwide prevalence of 20% [44]. In fact, the mental health of children and adolescents is a point of attention for both policy maker and researchers [45]. The fact that app developers have largely ignored this population could be due to various reasons. For instance, app developers may be put off by worries in the public about negative effects induced by mobile phone use [46]. In addition, children and adolescents may be a less attractive market because many parents and schools impose time restrictions on smartphone use by children and adolescents [47].

Few apps provided special interventions for patients with mental health disorders. This may partly be motivated by market considerations (ie, smaller target user population), but could also reflect the current lack of consensus about evidence-based guidelines for mHealth therapy for mental health patients. There is a large body of international literature about the use of mHealth interventions for people with mental health disorders [11,48-50] and similar work has been carried out in China [51]; however, the development of smartphone-based therapy is still in quite an early phase, and it may indeed be too early to offer it outside an experimental or academic mental health care setting. At this point, there is still a long way to go, both in terms of app development and efficacy evaluation, before evidence-based mHealth therapy can be routinely offered to patients in China.

With the exception of one, all apps could be downloaded for free; however, the actual cost of the services offered depended strongly on what was provided in the apps. In this study, more than 70% of the counseling apps, stress relief apps, and multipurpose apps charged a fee for some services, whereas a large part of the assessment apps and psychoeducation apps were free to use. This might be because counseling apps, stress relief apps, and multipurpose apps involved more user-specific services, offered either by paid professionals or as part of evidence-based stress relief methods that require a larger initial investment from the developer than, for instance, the inclusion of articles about mental health or the use of questionnaires that have developed by others. The cost of certain services in apps might present an obstacle for users, but this issue was not within the scope of this study.

Assessment was the most popular function across all apps; there were 49 assessment apps and 41 multipurpose apps that had an assessment section; however, most assessment apps (or sections) focused on assessment as a form of entertainment rather than on the assessment of real psychological problems or characteristics. Most multipurpose apps that also included counseling and stress relief sections, used assessments with psychological scales, whereas entertainment assessments were included in most single purpose assessment apps. Assessments of personality, anxiety, and depression were most common in apps that used psychological assessment scales, which makes sense given the high prevalence of anxiety and depression in China [3,4] and aligns with previous observations that people with anxiety and depression often do not seek help from professional institutes or physicians [52]. It could be that this group is relatively more likely to use a freely available mobile app for self-assessment.

Psychological counseling was the second most popular service and was provided in all 37 counseling apps and in 39 multipurpose apps. Psychologists were the main providers of counseling. Only 16 apps involved psychiatrists as providers of counseling. This indicates that psychologists in China play an important role in delivery of mental health care to the general population. This could be due to their larger number, which makes it easier for users to find a psychologist for mental health services. Indeed, in 2017, it was estimated that approximately 40,000 people in China held a national second- or third-level certificate in psychological counseling and provided psychological counseling on either a full- or part-time basis [53]. In contrast, the number of certified psychiatrists in China (in 2015) was 27,733. Another explanation for the relatively large number of psychologists involved in the reviewed apps could be that regulation in China leads psychologists and psychiatrists working in different settings. Psychologists work

XSL•FO
RenderX

in a broad range of settings, whereas all psychiatrists in China need to be registered to a mental health hospital or psychiatric ward in a general hospital that provides inpatient service [5]. According to the Mental Health Law of the People's Republic of China [54] psychologists should only perform counseling and are not authorized to perform psychotherapy, nor engage in the diagnosis or treatment of mental health disorders. If psychologists detect that a person receiving counseling may suffer from mental health disorders, he or she should recommend that the person seek services at a medical institute. As such, online counseling is a good way for psychologists to perform initial counseling in population-dwelling persons; however, it is currently hard to estimate the effectiveness of online psychological counseling by psychologists in China because of serious problems with the level of professionalism of psychological counseling (low performance levels, a lack of norms, and uneven levels of training and expertise across consulting staff) [55]. Unfortunately, we did not have any access to information about the kinds of problems users seek treatment for via the apps, nor did we have insight into the number of users that should be and the number that were, in fact, referred to medical care by consulting psychologists.

Another important finding was that most mental health apps simply served as a means to connect help seekers to health care providers, but did not use digital technologies (such as interactive computer-guided treatment algorithms) to deliver mHealth interventions that could (partly) replace a health care provider. This is not a problem per se, but it means that existing apps are unlikely to solve the current problem of insufficient mental health care providers for the demand in China. To achieve that, apps that can (partially) provide services that would otherwise be provided by health care professionals should be developed. More and more apps are being developed that offer a completely digital interactive therapy environment. For instance, cognitive behavior therapy–based mHealth apps have been developed for a large range of mental health problems, and the efficacy of such interventions has been confirmed in several studies [56-60]. To the best of our knowledge, such apps have not yet been developed for the Chinese population, nor have they been tested.

This study showed that meditation was the main intervention offered by apps that aim to help users relieve stress. Previous studies [61,62] have, indeed, shown that meditation supported by online tools can have a significant beneficial effect on depression, anxiety, and well-being and a large effect on stress. This seemed to be at least partially supported by the evidence-based nature of these apps; however, no research exists on the efficacy of mobile app–assisted meditation, specifically in China.

Most psychoeducation apps covered one or more of the eight core aspects of mental health information that have been proposed as part of national policy. As a result, the apps were usually quite complete and accurate in their information coverage. The fact that some psychoeducation apps were downloaded more than 100,000 times (eg, the app *geilixinli* had been downloaded 780,000 times, and the app *xinlizixunyidianling* had been downloaded 580,000 times), indicates that mHealth psychoeducation can play an important role in disseminating mental health knowledge, which in turn could help reduce levels of depression and psychological distress [63,64].

Limitations

This study has some limitations. First, we could not obtain the number of downloads for all apps because the iOS app store does not provide this information. In addition, even for the apps for which the download numbers were known, we could not be sure about the number of persons who actually used the app. Second, we based the evaluation of the apps on quite a superficial screening the app appearance and the information delivered by the developers. We did not check and cannot be sure that all services (eg, contact with certified professional) are provided as claimed and that all apps function correctly. Third, we evaluated if apps made any statements or claims about data safety, but could not evaluate if they, in fact, handled data securely. Fourth, we chose a set of search terms to make sure that we included many relevant apps; however, it is still possible that we missed some potentially relevant apps. Fifth, we did not carry out a formal evaluation of each app's quality as this was outside the scope of this study. This study looked at the landscape of services and functionalities that are provided by the apps and evaluated their potential to contribute to mental health care development; however, app quality is also an important factor in determining an app's usefulness. An interesting possibility for future studies could be to use a standardized measure to assess and compare the quality of individual apps (eg, the mobile app rating scale [65]); indeed, the mobile app rating scale has been used to evaluate the quality of mobile apps in other health-related fields in China [66]. Proper translation and validation of scales such as the mobile app rating scale in Chinese would be extremely useful for the systematic evaluation of mental health care apps. Sixth, all available apps were evaluated as of December 2018. Apps that have been newly added since then have not been included in this study; however, a new search (May 23, 2020) for the top 100 apps in the three Android app stores revealed that only 9 apps had been added since the initial search, making it unlikely that the current results and conclusions would have been significantly affected by their inclusion.

Conclusions

This study uncovered a large number of mental health apps available in China. The results show that there is a large variety in the aims and evidence-base among the apps. A considerable number of apps contained useful and evidence-based elements, such as good quality information about mental health, validated questionnaires, useful meditation methods, and access to counselors, which makes it likely that they can contribute positively to mental health care in China; however, for mental health care apps to substantially contribute, considerable challenges remain, including the need for more patient-focused and child- and adolescent-focused apps, the development of apps that can take over some of the roles of a health care provider, and efficacy studies.

XSL•FO
RenderX

Acknowledgments

This study was supported by a grant from the Tianjin Health and Family Planning Commission (the name has now been changed into Tianjin Health Commission) to the Tianjin Key Programs for Science and Technology Development in Health Industry (No. 13KG119).

Authors' Contributions

HY, GX, and RAS designed the study. HY, YW, NW, WC, and YZ downloaded the apps and collected data. HY analyzed the data. HY and KJW wrote the initial draft of the paper. All authors read, contributed to, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Leucht S, Heres S, Kissling W, Davis JM. Evidence-based pharmacotherapy of schizophrenia. Int J Neuropsychopharmacol 2011 Mar;14(2):269-284. [doi: 10.1017/S1461145710001380] [Medline: 21208500]
2. Kessler RC, Angermeyer M, Anthony JC, Demyttenaere K, Gasquet I, Gluzman S, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
3. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
4. Yin H, Xu G, Tian H, Yang G, Wardenaar KJ, Schoevers RA. The prevalence, age-of-onset and the correlates of DSM-IV psychiatric disorders in the Tianjin Mental Health Survey (TJMHS). Psychol Med 2018 Dec;48(3):473-487. [doi: 10.1017/S0033291717001878] [Medline: 28714421]
5. Patel V, Xiao S, Chen H, Hanna F, Jotheeswaran AT, Luo D, et al. The magnitude of and health system responses to the mental health treatment gap in adults in India and China. Lancet 2016 Dec 17;388(10063):3074-3084. [doi: 10.1016/S0140-6736(16)00160-4] [Medline: 27209149]
6. Liang D, Mays VM, Hwang W. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018 Jan 01;33(1):107-122 [FREE Full text] [doi: 10.1093/heapol/czx137] [Medline: 29040516]
7. Loo JMY, Raylu N, Oei TPS. Gambling among the Chinese: a comprehensive review. Clin Psychol Rev 2008 Oct;28(7):1152-1166. [doi: 10.1016/j.cpr.2008.04.001] [Medline: 18486290]
8. Aboujaoude E, Salame W, Naim L. Telemental health: a status update. World Psychiatry 2015 Jun;14(2):223-230 [FREE Full text] [doi: 10.1002/wps.20218] [Medline: 26043340]
9. Hilty DM, Chan S, Hwang T, Wong A, Bauer AM. Advances in mobile mental health: opportunities and implications for the spectrum of e-mental health services. Mhealth 2017;3:34 [FREE Full text] [doi: 10.21037/mhealth.2017.06.02] [Medline: 28894744]
10. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
11. Van AM, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
12. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
13. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
14. Guo Y, Yang F, Hu F, Li W, Ruggiano N, Lee HY. Existing mobile phone apps for self-care management of people with Alzheimer disease and related dementias: systematic analysis. JMIR Aging 2020 Jan 24;3(1):e15290 [FREE Full text] [doi: 10.2196/15290] [Medline: 32012045]
15. Guo Y, Yang F, Hu F, Li W, Ruggiano N, Lee HY. Existing mobile phone apps for self-care management of people with alzheimer disease and related dementias: systematic analysis. JMIR Aging 2020 Jan 24;3(1):e15290 [FREE Full text] [doi: 10.2196/15290] [Medline: 32012045]
16. Castillo-Sánchez G, Camargo-Henríquez I, Muñoz-Sánchez JL, Franco-Martín M, de la Torre-Díez I. Suicide prevention mobile apps: descriptive analysis of apps from the most popular virtual stores. JMIR Mhealth Uhealth 2019 Aug 13;7(8):e13885 [FREE Full text] [doi: 10.2196/13885] [Medline: 31411144]

XSL•FO
RenderX

17. Number of smartphone users in China from 2017 to 2023. Number of smartphone users in China from 2017 to 2023.: Hambg Stat 2018 URL: https//www.statista.com/statistics/467160/forecast-of-smartphone-users-in-china/ [accessed 2016-06-01] [WebCite Cache ID 78np6j5mg]
18. Hamburg:Statista 2016. China's mHealth market value from 2011 to 2017. URL: https://www.statista.com/statistics/328927/mobile-health-market-scale-in-china/ [accessed 2019-06-01] [WebCite Cache ID 78npfl5Iy]
19. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016 Aug 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
20. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
21. Sinha Deb K, Tuli A, Sood M, Chadda R, Verma R, Kumar S, et al. Is India ready for mental health apps (MHApps)? A quantitative-qualitative exploration of caregivers' perspective on smartphone-based solutions for managing severe mental illnesses in low resource settings. PLoS One 2018;13(9):e0203353 [FREE Full text] [doi: 10.1371/journal.pone.0203353] [Medline: 30231056]
22. Abuwalla Z, Clark MD, Burke B, Tannenbaum V, Patel S, Mitacek R, et al. Long-term telemental health prevention interventions for youth: a rapid review. Internet Interv 2018 Mar;11:20-29 [FREE Full text] [doi: 10.1016/j.invent.2017.11.006] [Medline: 30135756]
23. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198. [doi: 10.2196/jmir.4581]
24. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
25. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 7;7(11):e13236. [doi: 10.2196/13236] [Medline: 31697245]
26. Search results for “mental” in the Baidu app store. URL: https://shouji.baidu.com/s?wd=%E5%BF%83%E7%90%86&data_type=app&f=search_sug%40app%40btn_search [accessed 2016-06-01] [WebCite Cache ID 78nnISVDK]
27. Yoo E. The Top Ten Android App Stores in China.: Technode; 2016. URL: http://technode.com/2015/09/22/ten-best-android-app-stores-china/ [accessed 2016-08-08] [WebCite Cache ID 6jbyEdiWG]
28. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
29. Levin ME, Haeger J, Pierce B, Cruz RA. Evaluating an adjunctive mobile app to enhance psychological flexibility in acceptance and commitment therapy. Behav Modif 2017 Nov;41(6):846-867. [doi: 10.1177/0145445517719661] [Medline: 28689449]
30. Swartz HA, Novick DM. Psychotherapy in the digital age: what we can learn from interpersonal psychotherapy. Am J Psychother 2020 Mar 01;73(1):15-21. [doi: 10.1176/appi.psychotherapy.20190040] [Medline: 32050784]
31. Cuijpers P, de Wit L, Kleiboer A, Karyotaki E, Ebert DD. Problem-solving therapy for adult depression: an updated meta-analysis. Eur Psychiatry 2018 Dec;48:27-37. [doi: 10.1016/j.eurpsy.2017.11.006] [Medline: 29331596]
32. URL: http://www.nhc.gov.cn/wjw/xlws/201304/210ec38904944ae9a8c0747a6f4b6747.shtml [accessed 2019-06-01] [WebCite Cache ID 78nxYI7AX]
33. Eysenck HJ, Eysenck SBG. Manual of the Eysenck Personality Questionnaire. London: Hodder and Stoughton; 1975.
34. Qian M, Wu G, Zhu R, Zhang X. Revision of the Chinese version of Eysenck Personality Questionnaire. Acta Psychol Sin 2000;32(3):317-323.
35. Cattell HEP, Mead AD. The sixteen personality factor questionnaire (16PF). In: The SAGE handbook of personality theory and assessment. Personality Measurement and Testing (Volume 2). ISBN. London: Sage Publications; 2008:135-159.
36. Zhu P, Dai H. Norm of the Sixteen Personality Factor Test in Chinese population. Xinli Kexue Tongxun 1988;6:16-20.
37. Liu X, Tang M, Peng X, Chen K, Dai Z. Factor analysis of Self-rating Anxiety Scale. Chinese J Nerv Ment Dis 1995;21(6):359-360. [doi: 10.1037/t04092-000]
38. Zung WW. A rating instrument for anxiety disorders. Psychosomatics 1971;12(6):371-379. [doi: 10.1016/S0033-3182(71)71479-0] [Medline: 5172928]
39. Zhang J, Wong Z. Factor analysis of Self-rating Depression Scale. Shandong Arch Pyschiatry 2004;17(3):131-133.
40. ZUNG WWK. A Self-Rating Depression Scale. Arch Gen Psychiatry 1965 Jan 01;12(1):63. [doi: 10.1001/archpsyc.1965.01720310065008]
41. Tang Y, Zhang M. Hamilton Depression Scale. Shandong Arch Pyschiatry 1984;2:61-64. [doi: 10.1007/springerreference_184614]
42. Wang Z. Symptom Checklist 90. Shanghai Arch Psychiatry 1984;2:68-70. [doi: 10.32388/sg7b5l]
43. Pan J, Zhang J, Ma P, Liang H, Wang H, Wen S, et al. The utility of Liebowits Social Anxiety Scale in the patients with social anxiety disorder in Chinese. Chinese J Nerv Ment Dis 2006;32(3):206-210.

XSL•FO
RenderX

44. Zheng Y, Zheng X. Current state and recent developments of child psychiatry in China. Child Adolesc Psychiatry Ment Health 2015;9:10 [FREE Full text] [doi: 10.1186/s13034-015-0040-0] [Medline: 25972919]
45. Colizzi M, Lasalvia A, Ruggeri M. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? Int J Ment Health Syst 2020;14:23 [FREE Full text] [doi: 10.1186/s13033-020-00356-9] [Medline: 32226481]
46. Shao Y, Wang J, Xie B. The first mental health law of China. Asian J Psychiatr 2015 Feb;13:72-74 [FREE Full text] [doi: 10.1016/j.ajp.2014.11.002] [Medline: 25486869]
47. State Council of the People's Republic of China. Opinions of the State Council on the implementation of healthy China action. Bulletin. 2019. URL: http://english.www.gov.cn/policies/latest_releases/2019/07/15/content_281476765919352.htm [accessed 2020-07-06]
48. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry ? 2017;16(3):298. [Medline: 28941113]
49. Kassianos AP, Georgiou G, Papaconstantinou EP, Detzortzi A, Horne R. Smartphone applications for educating and helping non-motivating patients adhere to medication that treats mental health conditions: aims and functioning. Front Psychol 2017;8:1769 [FREE Full text] [doi: 10.3389/fpsyg.2017.01769] [Medline: 29075216]
50. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? a meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
51. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500. [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
52. Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001-05: an epidemiological survey. Lancet 2009 Jun 13;373(9680):2041-2053. [doi: 10.1016/S0140-6736(09)60660-7] [Medline: 19524780]
53. 2016-06-01. The Mental Healthcare Industry in China. URL: https://www.china-briefing.com/news/mental-healthcare-industry-china/ [accessed 2019-06-01] [WebCite Cache ID 78nq6l1U3]
54. URL: http://www.npc.gov.cn/npc/xinwen/2018-06/12/content_2055876.htm [accessed 2019-06-01] [WebCite Cache ID 78nsnCHxu]
55. Wang X. Issues and recommendation to China's professionalization of psychological counseling. Gansu Gaoshi Xuebao 2010;15(4):114-116. [doi: 10.1057/9781137287670.0014]
56. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
57. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
58. Simblett S, Birch J, Matcham F, Yaguez L, Morris R. A systematic review and meta-analysis of e-mental health interventions to treat symptoms of posttraumatic stress. JMIR Ment Health 2017 May 17;4(2):e14 [FREE Full text] [doi: 10.2196/mental.5558] [Medline: 28526672]
59. Hennemann S, Farnsteiner S, Sander L. Internet- and mobile-based aftercare and relapse prevention in mental disorders: a systematic review and recommendations for future research. Internet Interv 2018 Dec;14:1-17 [FREE Full text] [doi: 10.1016/j.invent.2018.09.001] [Medline: 30510909]
60. Christoforou M, Sáez FJA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. J Med Internet Res 2017 Nov 24;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
61. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
62. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
63. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
64. Van DT, Hermans D, Van AC, Van DBO. Stress reduction through psychoeducation: a meta- analytic review. Health Educ Behav 2012 Aug;39(4):474-485. [doi: 10.1177/1090198111419202] [Medline: 21986242]

XSL•FO
RenderX

65. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
66. Xie B, Su Z, Zhang W, Cai R. Chinese cardiovascular disease mobile apps' information types, information quality, and interactive functions for self-management: systematic review. JMIR Mhealth Uhealth 2017 Dec 14;5(12):e195-e120 [FREE Full text] [doi: 10.2196/mhealth.8549] [Medline: 29242176]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 03.06.19; peer-reviewed by G Wu, W Zhang; comments to author 24.12.19; revised version received 23.03.20; accepted 03.06.20; published 27.07.20.

Please cite as:
Yin H, Wardenaar KJ, Wang Y, Wang N, Chen W, Zhang Y, Xu G, Schoevers RA
Mobile Mental Health Apps in China: Systematic App Store Search
J Med Internet Res 2020;22(7):e14915
URL: https://www.jmir.org/2020/7/e14915
doi:10.2196/14915
PMID:32716301

©Huifang Yin, Klaas J Wardenaar, Yuhao Wang, Nan Wang, Wenjin Chen, Yan Zhang, Guangming Xu, Robert A Schoevers. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Surveilling Influenza Incidence With Centers for Disease Control and Prevention Web Traffic Data: Demonstration Using a Novel Dataset

Wendy K Caldwell[1,2], BS, MA, PhD; Geoffrey Fairchild[3], BS, MS, PhD; Sara Y Del Valle[3], BS, MS, PhD

[1]X Computational Physics Division, Los Alamos National Laboratory, Los Alamos, NM, United States

[2]School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, United States

[3]Analytics, Intelligence, and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, United States

Corresponding Author:
Wendy K Caldwell, BS, MA, PhD
X Computational Physics Division
Los Alamos National Laboratory
P.O. Box 1663
Mail Stop T086
Los Alamos, NM, 87545
United States
Phone: 1 5056678593
Email: wkcaldwell@lanl.gov

Abstract

Background: Influenza epidemics result in a public health and economic burden worldwide. Traditional surveillance techniques, which rely on doctor visits, provide data with a delay of 1 to 2 weeks. A means of obtaining real-time data and forecasting future outbreaks is desirable to provide more timely responses to influenza epidemics.

Objective: This study aimed to present the first implementation of a novel dataset by demonstrating its ability to supplement traditional disease surveillance at multiple spatial resolutions.

Methods: We used internet traffic data from the Centers for Disease Control and Prevention (CDC) website to determine the potential usability of this data source. We tested the traffic generated by 10 influenza-related pages in 8 states and 9 census divisions within the United States and compared it against clinical surveillance data.

Results: Our results yielded an r^2 value of 0.955 in the most successful case, promising results for some cases, and unsuccessful results for other cases. In the interest of scientific transparency to further the understanding of when internet data streams are an appropriate supplemental data source, we also included negative results (ie, unsuccessful models). Models that focused on a single influenza season were more successful than those that attempted to model multiple influenza seasons. Geographic resolution appeared to play a key role, with national and regional models being more successful, overall, than models at the state level.

Conclusions: These results demonstrate that internet data may be able to complement traditional influenza surveillance in some cases but not in others. Specifically, our results show that the CDC website traffic may inform national- and division-level models but not models for each individual state. In addition, our results show better agreement when the data were broken up by seasons instead of aggregated over several years. We anticipate that this work will lead to more complex nowcasting and forecasting models using this data stream.

(J Med Internet Res 2020;22(7):e14337) doi:10.2196/14337

KEYWORDS

influenza; surveillance; infoveillance; infodemiology; projections and predictions; internet; data sources

XSL•FO
RenderX

Introduction

Background and Motivation

Every year, an estimated 5% to 20% of people in the United States become infected with influenza [1]. The typical influenza season begins in October and ends in May, with the peak occurring in the winter months. Annually, 3000 to 50,000 people die from the flu, with another 200,000 people requiring hospitalization [2]. The yearly flu burden is estimated to cost around US $11.2 billion in lost productivity, with some estimates as high as US $87 billion [2,3]. Timely surveillance of influenza can help reduce this burden, allowing health care facilities to more adequately prepare for the influx of patients when flu levels are high [4].

One common surveillance measure is the fraction of patients presenting with influenza-like illness (ILI), consisting of a fever of at least 100°F (37.8°C) and a cough or sore throat with no other known cause [5]. ILI data are collected from about 2900 volunteer health care providers throughout the United States, although each week, only about 1800 of them report their data. These data are then aggregated and made public after a time lag of about 1 to 2 weeks [1,6-10]. As the ILI data are collected from volunteer health care providers, the dataset is incomplete. If policies were enacted to provide incentives for health care providers who report these data or to make reporting compulsory, the result would be a more complete dataset. Other surveillance systems include virological data from the World Health Organization, emergency department visits, electronic health records, crowd-sourced ILI reports, Widely Internet Sourced Distributed Monitoring, Influenzanet, and Flu Near You [11,12].

Internet Data Streams

In the United States, 87% [13] of adults use the internet. Of those internet users, 72% [13] have used the internet to search for health information within the last year. The most common health-related searches are for information regarding a specific disease or condition (66%) and information about a specific treatment or procedure (56%) [13,14].

There are two main types of health-related internet activity. The first is health sharing, in which internet users post about health-related topics (eg, a tweet about being sick). The second is health seeking, in which users use the internet to obtain information about health-related topics [6]. In this paper, we focused on health-seeking behavior. Previous studies have shown that analyzing web-based health-seeking behavior can improve early detection of disease incidence by detecting changes in disease activity [9,15-19]. Similarly, other studies have shown that internet data emerging from search queries can aid detection of outbreaks in areas with large populations of internet users [20] because web-based health-related search queries and epidemics are often strongly correlated [20,21].

Internet data have been used to forecast disease incidence in other models. Polgreen et al [9] developed linear influenza forecasting models with lags of 1 to 10 weeks for each of the 9 US census regions using search queries from Yahoo [9]. The best performing models had lags of 1 to 3 weeks and an average r^2 value of 0.38 (with a high of 0.57 in the East South Central region) [9]. These low r^2 values demonstrate potential problems in relying on search information alone. Ginsberg et al [15] were able to predict influenza epidemics 2 weeks in advance using Google search queries to fit linear models using log-odds of ILI visits and related searches.

Using a Poisson distribution and Lasso regression, McIver and Brownstein [8] obtained an r^2 value of 0.946 using Wikipedia data, although some data were excluded from analyses because of increased media attention and higher than normal influenza activity. Generous et al [7] used Wikipedia data to train a statistical model with linear regression, which demonstrated its potential for forecasting disease incidence worldwide, including influenza in the United States, which had an r^2 value of 0.89. Hickmann et al [1] conducted a similar study of linear regression models, which showed that using Wikipedia to forecast influenza in the United States for the 2013 to 2014 season resulted in an r^2 value greater than 0.9 in some instances.

Integrating both Wikipedia data and Google Flu Trends, Bardak et al [22] obtained r^2 values of 0.94 and 0.91 using ordinary least squares (OLS) and ridge regression, respectively, for forecasting influenza outbreaks. For OLS nowcasting, the r^2 value was 0.98 in the best case. For the best fit, the weekly data were offset by 1 week, so that ILI visits were correlated with internet data from the prior week [22].

As part of the Centers for Disease Control and Prevention (CDC)'s 2013 to 2014 Predict the Influenza Season Challenge, 9 teams used digital data sources to create forecasting models. The digital sources these teams used were Wikipedia, Twitter, Google Flu Trends, and HealthMap. The teams used either mechanistic or statistical models to create their forecasts, with the most successful team using multiple data sources, which may have reduced biases usually associated with internet data streams [23]. Broniatowski et al [24] used Twitter data to detect increasing and decreasing influenza prevalence with 85% accuracy. Zhang et al [25] used Twitter data to inform stochastically, spatially structured mechanistic models of influenza in the United States, Italy, and Spain.

Internet data streams have also been used to supplement traditional surveillance techniques with nowcasting models. Paul et al [26] used Twitter along with ILI data from the CDC to produce nowcasting influenza models as well as nowcasting models using solely ILI data. They concluded that the addition of Twitter data led to more accurate nowcasting models. Santillana et al [27] combined Google Trends data and CDC-reported ILI data to create models for nowcasting and forecasting influenza. Lampos et al [28] used search query data to explore both linear and nonlinear nowcasting models. Yang et al [29] used Google search data to create an influenza tracking model with autoregression.

In contrast, we considered data on page views of the CDC website rather than search data from sites not solely devoted to public health. We used this dataset because we expect it to be inherently less noisy because of its focus on public health issues. We used OLS to nowcast influenza nationally, across the 9 US

XSL•FO
RenderX

census divisions, and across 8 states using access data from 10 influenza-related CDC pages. Our nowcasting models cover influenza seasons from 2013 to 2016, with the 2012 to 2013 season being partially included because the CDC page view dataset begins on January 1, 2013. The inclusion of an incomplete influenza season serves to inform if this dataset can be used given a more restrictive time frame. We included both positive and negative results to advance our knowledge regarding when internet data may or may not work. The negative results are crucial to advancing the field of disease surveillance using internet data, as they demonstrate when these data sources contribute to unreliable surveillance. We focus on answering the following two research questions: (1) Can CDC page visits be used as an additional data source for monitoring disease incidence? and (2) What is the appropriate time shift of the page view data needed to obtain the best data fit?

Methods

Data Sources

We used page view data provided by the CDC. Each data point contains the page name, date and time of access, and the geographic location from where the page was viewed. These data are available at geographic resolutions of national and state levels and include some metropolitan areas (eg, New York City). The data are available at a number of temporal resolutions beginning on January 1, 2013. For these models, we used weekly page view data to coincide with the ILI data temporal resolution. The data are available as raw page view counts and page view counts normalized with respect to all CDC page views, and we considered the latter for this work. We selected pages associated with general influenza information, treatment, and diagnosis. Pages were sometimes renamed, but we were able to follow the evolution of each selected page by using keywords in the page titles as well as the date ranges for available data.

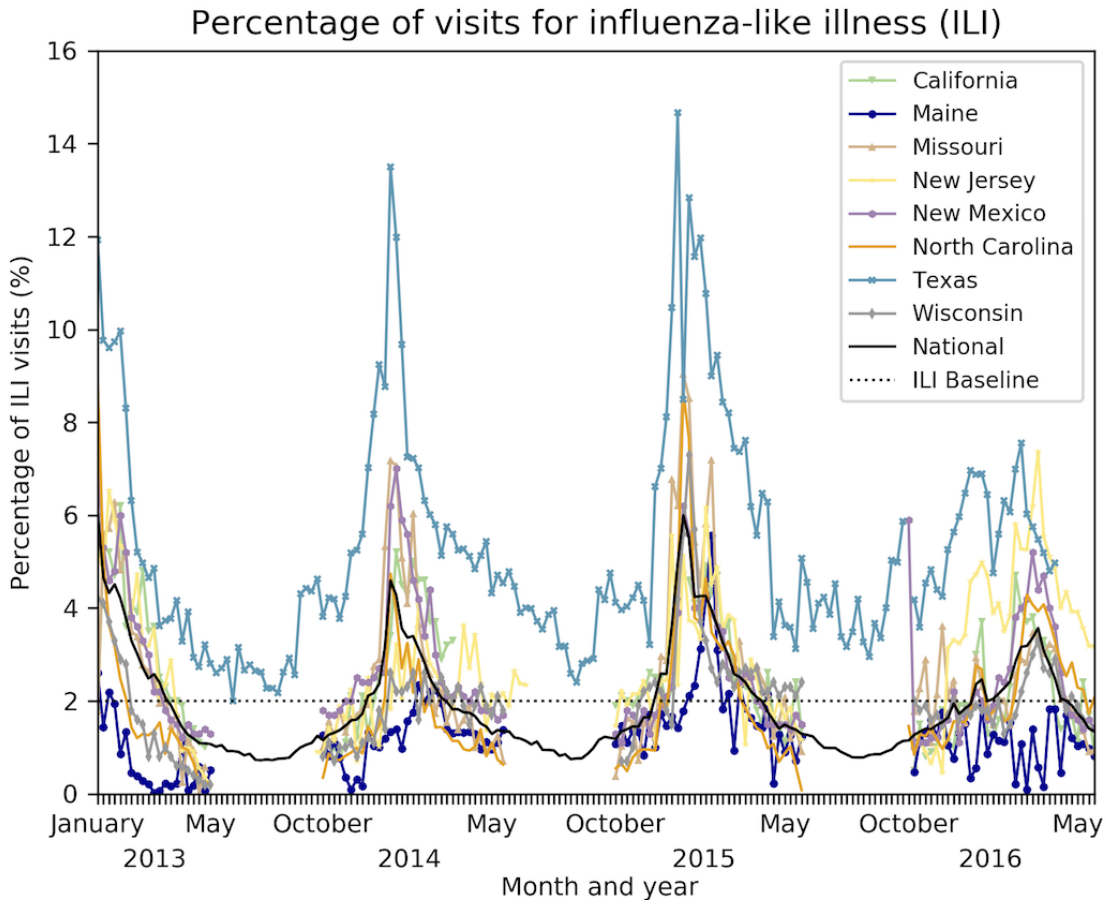
As the majority of health-related internet searches concern specific conditions, treatments, and procedures [14], we selected pages related to those topics. These pages also align with the study by Johnson et al [30], who used pages in the categories of diagnosis and treatment as well as prevention and vaccination for influenza surveillance [30]. Specifically, we used the following pages: antivirals, flu basics, FluView, high risk complications, key facts, prevention, symptoms, treating influenza, treatment, and vaccine. We then aggregated the page views of interest for each of our models. FluView has the potential to be an outlier page, especially when used alone, as this page tracks the severity of the influenza season and could have higher page views as a result of media attention and severe influenza seasons. However, when combined with other pages focused on treatment and prevention, we expected these page view data to be useful for our models. A complete list of pages can be found in Multimedia Appendix 1.

The states we selected were based on the severity of flu (determined from FluView) during the available seasons and the availability of ILI data at the time of the study, which is not standardized and is dependent on each state's reporting mechanism. ILI data for each state include the week ending or starting date and the percentage of ILI for the specified week. Although some states also report additional data, such as school closures and hospitalizations, these data are not made available by every state. Note that ILI reporting and accessibility vary across all states. The states we selected were (1) California, (2) Maine, (3) Missouri, (4) New Jersey, (5) New Mexico, (6) North Carolina, (7) Texas, and (8) Wisconsin. With the exception of Texas, these states did not release ILI data outside of the typical flu season. As the purpose of this study was to demonstrate the viability of nowcasting, we considered only those ILI data available during the study period. Although some states have made their ILI more accessible since the end of the study, we did not consider these data, as they were not available during the study period. The exclusion of additional data not available during the study period helps to preserve the premise of nowcasting by focusing only on data sources available during the study period. Likewise, our state ILI data often came from the state's individual weekly reports during the seasons used in the study. A complete list of the data sources for the state ILI can be found in Multimedia Appendix 2, and the clinical data are available in Multimedia Appendices 3 and 4.

Figure 1 shows the percentage of ILI visits for each state considered in this study and the national percentage of ILI visits. We see distinct spikes that indicate the peaks of the flu seasons. With the exception of Maine, which behaves as an outlier at times, the figure shows spikes indicating there are peak weeks for influenza-related page views. Texas also exhibits outlier behavior with ILI percentages consistently higher than the typical national baseline of 2%, which is used to determine when the flu has reached epidemic status. These 2 outliers are shown in teal (Texas) and dark blue (Maine). The national ILI is shown in black. The remaining states exhibit behavior consistent with the national ILI trend. Figure 2 shows the CDC page view data as a heat map: weeks with more page views are shown darker than weeks with fewer page views.

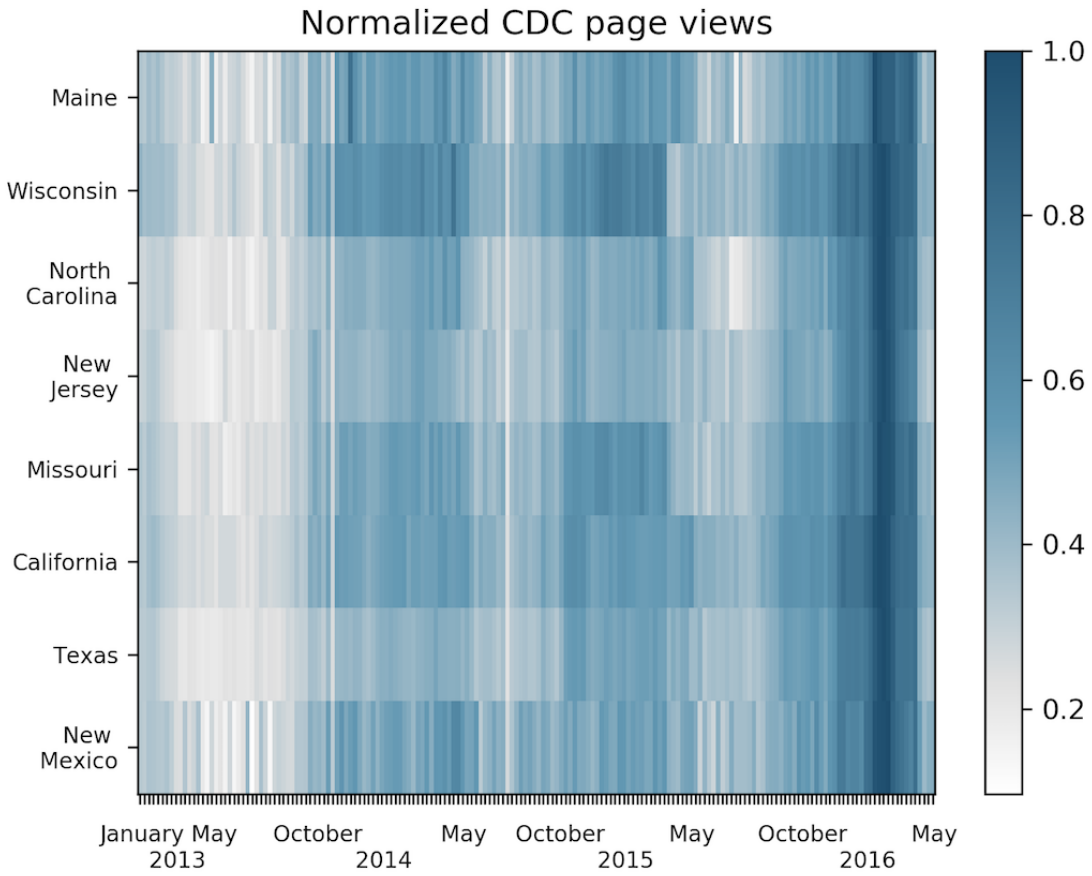
XSL•FO
RenderX

Figure 1. Percentage of ILI visits per state compared with the typical national baseline of 2%. Maine (dark blue) and Texas (teal) exhibit outlier behavior, with Texas having a greater ILI percentage and Maine having a lesser ILI percentage. The remaining states follow the national ILI trend, shown in black. ILI: influenza-like illness.



XSL•FO
RenderX

Figure 2. Normalized CDC web traffic as a heat map. Darker areas indicate more page views and appear to correlate with increases in influenza-like illness. The page views also appear to be more prevalent during the typical influenza season, October to May. CDC: Centers for Disease Control and Prevention.



In addition to selected states, we also considered the 9 US census divisions: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. Multimedia Appendix 5 provides a list of states included in each division. Data for the census divisions were obtained from the CDC and are presented in Multimedia Appendix 3.

Linear Regression

We used statsmodels version 0.9.0 [31], a statistical analysis module for Python, to perform linear regression on our datasets using OLS. This creates a linear model *M*, the summation of regression coefficients multiplied by page view data. Figure 3 shows the mathematical formula of *M*, where [x] are the regression coefficients and X=1, X_1, ... X_n is the vector of CDC page view data, with *n* representing the number of CDC pages used for the model, ranging from 1 to 10. *M* is a value between 0 and 1, representing the fraction of ILI visits. To plot the models and data on the same axes, we normalized *M* for visualization purposes, with *M=1* corresponding to the ILI percentage during the peak week of the influenza season. We correlated ILI and CDC page views for the same week or with a 1-week shift. In the shifted cases, we shifted the ILI data forward by 1 week, so that the model associates the current week's page views with the following week's ILI data. This shift is performed to account for the incubation period of influenza and the time between the onset of symptoms and the first doctor visit. Statsmodels [31] uses the CDC page view and ILI data to determine the appropriate regression coefficients; fits parameters with OLS; and computes the goodness-of-fit, r^2, also referred to as the coefficient of determination. The r^2 value measures how well 2 time series correlate. An r^2 value of 1 indicates a perfect fit, whereas an r^2 value of 0 indicates no correlation. Although r^2 is not necessarily the best metric to judge goodness-of-fit [6], it is nonetheless the most common metric used and still provides one with a decent overall sense of fit quality. In addition, we examined the root mean square error (RMSE) and the normalized root mean square error (NRMSE) using Python scikit-learn libraries. The RMSE and NRMSE metrics measure how the model prediction differs from the actual data, with the NRMSE normalized so that the greatest possible value is 1. For these metrics, lower numbers indicate a better fit.

XSL•FO
RenderX

Figure 3. Mathematical formula of the linear ILI models created in this study. The model M represents the fraction of ILI visits, where <inline-graphic xlink:href="jmir_v22i6e14337_fig7.png" mimetype="image" xlink:type="simple"/> are the regression coefficients and X=1, X_1, ... X_n is the vector of CDC page view data, with n representing the number of CDC pages used for the model, ranging from 1 to 10. ILI: influenza-like illness; CDC: Centers for Disease Control and Prevention.

$$M = \sum_{i=0}^{n} \alpha_i X_i$$

Results

Format of Results

We analyzed the data at the national, division, and state levels and computed r^2 for each geographic resolution. In this section, we discuss the results of our experiments, both successes and failures. We include figures of models at the national, census division, and state levels. Owing to the varying scales between page views and ILI percent, we chose to normalize the data and our models to plot them on the same axes. Raw data were used to create the models and then each model was normalized with respect to its maximum. We also normalized the ILI data and CDC web traffic data with respect to their maximums for the given period so that all 3 curves may appear in the same plot. Additional model successes and failures not discussed here can be found in Multimedia Appendix 6.

National Results

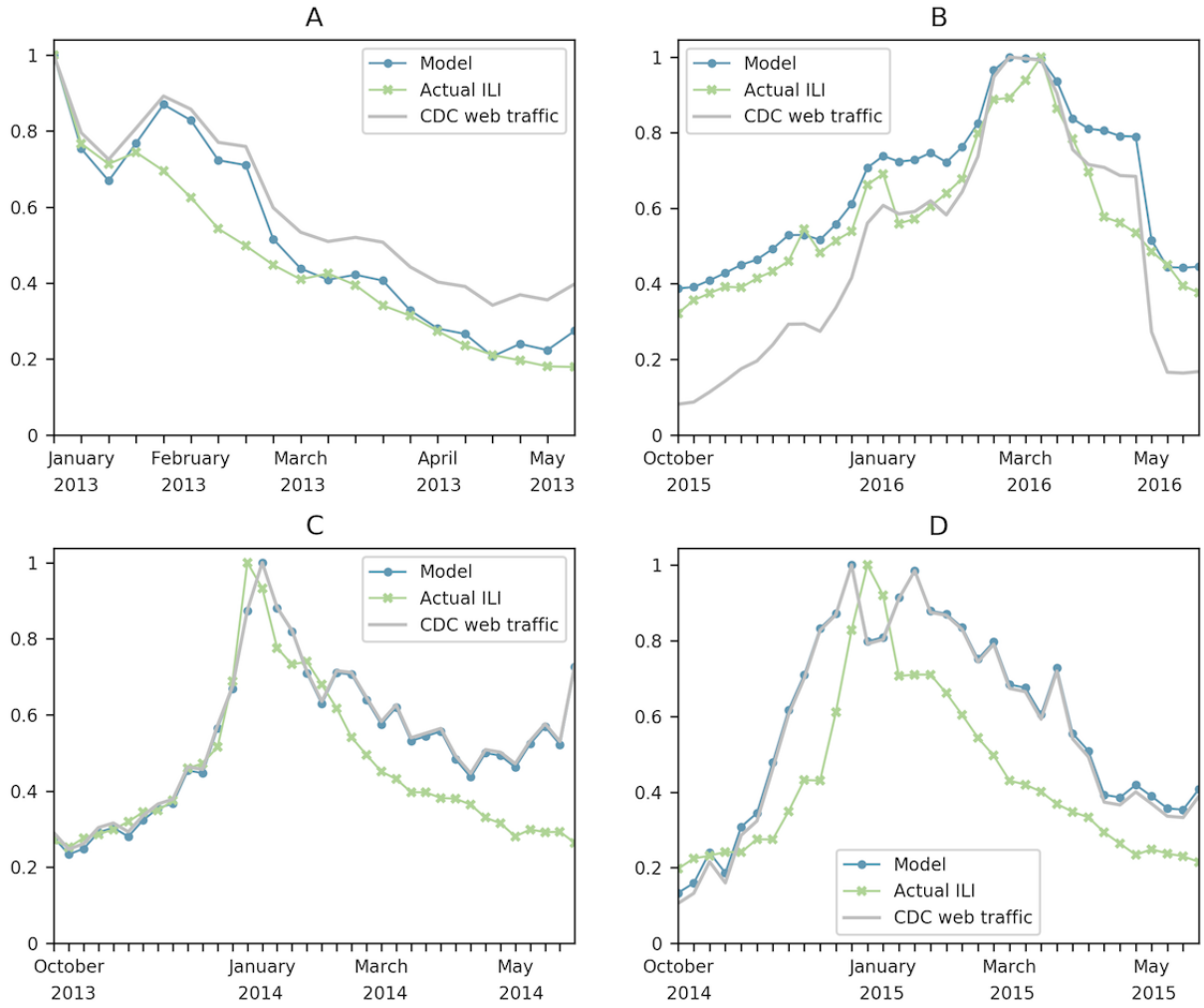
We selected pages that corresponded to the topics most often searched during web-based health-seeking activities. Aggregating all 10 pages in a single model, we were able to achieve an r^2 value of 0.889 for the national 2012 to 2013 influenza season after implementing a 1-week shift. We also succeeded in modeling the national 2015 to 2016 influenza season with no shift, achieving an r^2 value of 0.834. We obtained better results when limiting the pages to FluView, Symptoms, and Treatment, which we attribute to the information on these pages aligning with topics most commonly used for internet health seeking. For these pages, the most successful models did not have a shift. For the 2012 to 2013 influenza season, we achieved an r^2 value of 0.906. The model for the 2015 to 2016 season had an r^2 value of 0.891. Table 1 shows the most successful model for each influenza season included in this study. Figure 4 shows these models, with each figure caption indicating which page(s) comprise CDC web traffic, which appears in each figure and are the data used in the model.

Table 1. Pages and shifts for the most successful models for each influenza season at the national level.

Pages used in model	Season	Shift	r^2	Root mean square error	Normalized root mean square error
FluView, Symptoms, and Treatment	2012-2013	None	0.912	0.423	0.070
Symptoms	2015-2016	None	0.892	0.213	0.060
FluView	2013-2014	None	0.802	0.510	0.111
Antivirals and Prevention	2014-2015	None	0.778	0.615	0.103

XSL•FO
RenderX

Figure 4. These plots show national models and the associated pages and influenza seasons. (A) FluView, Symptoms, and Treatment, 2012 to 2013. (B) Symptoms, 2015 to 2016. (C) FluView, 2013 to 2014. (D) Antivirals and Prevention, 2014 to 2015. CDC: Centers for Disease Control and Prevention; ILI: influenza-like illness.

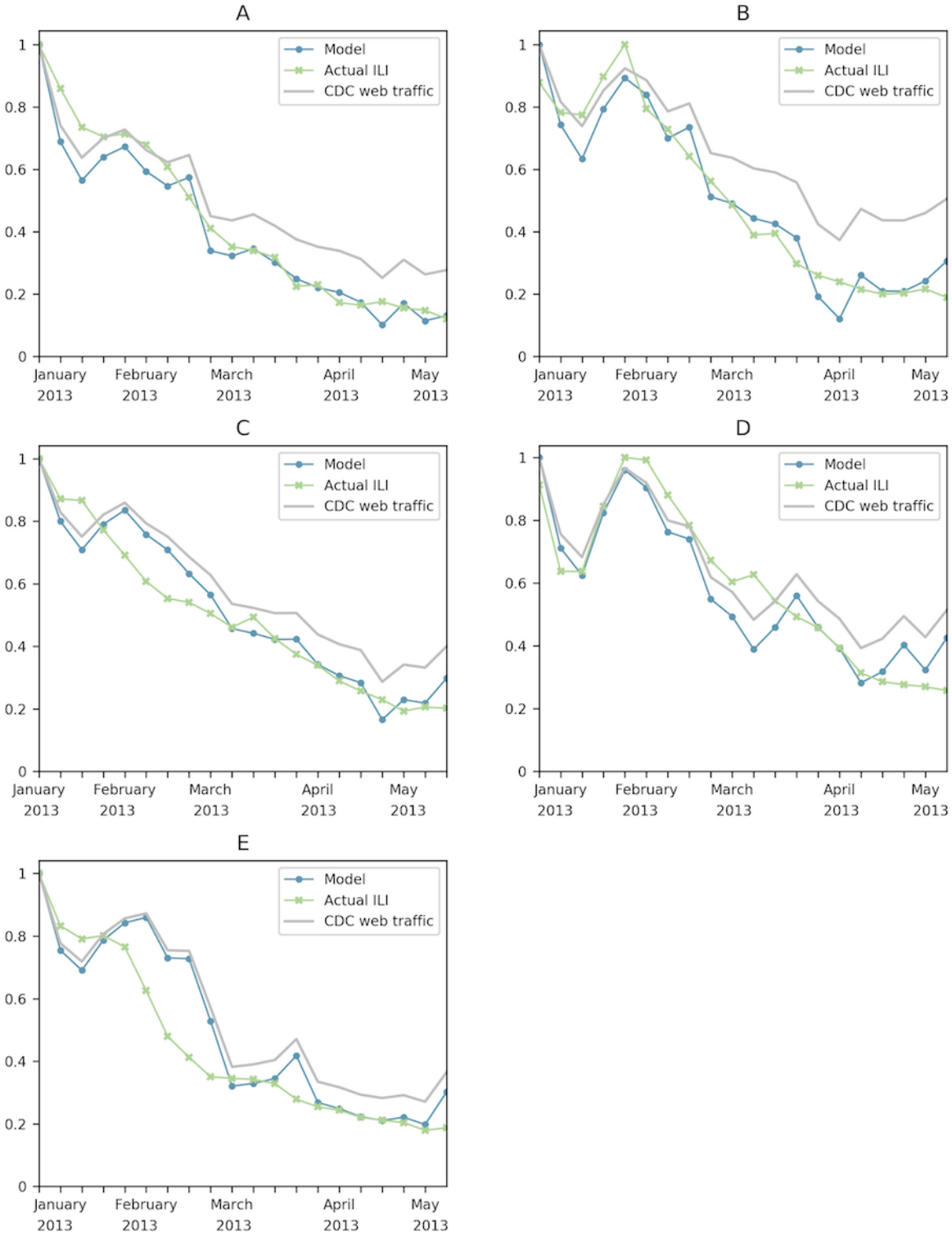


Census Division Results

Using the data for each of the 9 census divisions, we were able to achieve an r^2 value greater than 0.7 for at least one case for each division. We considered all seasons together and separately, with the better results obtained from modeling each individual season. We considered all 10 pages together as well as combinations of one or more of these pages. In the most successful case, the model was able to closely match the 2015 to 2016 influenza season for the West North Central division, with an r^2 value of 0.955 using the FluView, Symptoms, and Treatment pages. Although we had successes using all 10 pages, the most successful model for each division involved only these 3 pages. Figure 5 shows some of these models, and Table 2 highlights these successes.

XSL•FO
RenderX

Figure 5. Census division model successes using the FluView, Symptoms, and Treatment pages for the 2012 to 2013 influenza season. (A) West North Central, 2012 to 2013. (B) Mountain, 2012 to 2013. (C) East North Central, 1-week shift, 2012 to 2013. (D) Pacific, 2012 to 2013. (E) West South Central 2012 to 2013. These plots represent the census division models that had the highest r^2 value in the 2012 to 2013 influenza season. CDC: Centers for Disease Control and Prevention; ILI: influenza-like illness.



XSL•FO
RenderX

Table 2. The 9 census divisions and the season and shift for which the division's model had the highest r^2 value. The table also shows the root mean square error and the normalized root mean square error. The results presented correspond to the FluView, Symptoms, and Treatment pages aggregated.

Division	Season	Shift	r^2	Root mean square error	Normalized root mean square error
West North Central	2012-2013	None	0.955	0.367	0.057
Mountain	2012-2013	None	0.921	0.336	0.077
New England	2015-2016	None	0.920	0.096	0.096
East North Central	2012-2013	1 week	0.899	0.331	0.076
South Atlantic	2015-2016	None	0.893	0.218	0.065
Middle Atlantic	2015-2016	None	0.861	0.302	0.073
Pacific	2012-2013	None	0.849	0.503	0.094
West South Central	2012-2013	None	0.828	0.986	0.105
East South Central	2015-2016	1 week	0.793	0.365	0.082

State Results

We found r^2 for each of the states considered in this study, using a variety of pages and page combinations. Table 3 lists the most successful models for each state, the season, the data shift, and the r^2 value.

Table 3. The most successful results for each state considered in this study.

State	Page(s)	Season	Shift	r^2	Root mean square error	Normalized root mean square error
Texas	All[a]	2012-2013	1 week	0.930	0.667	0.067
Wisconsin	FVST[b]	2012-2013	None	0.833	0.533	0.127
New Jersey	All	2012-2013	1 week	0.832	0.767	0.117
Missouri	FVST	2012-2013	1 week	0.823	0.801	0.127
North Carolina	FVST	2015-2016	1 week	0.781	0.455	0.106
New Mexico	All	2015-2016	1 week	0.771	1.184	0.197
California	FVST	2012-2013	1 week	0.758	0.777	0.125
Maine	Antivirals	2012-2013	None	0.662	0.445	0.171

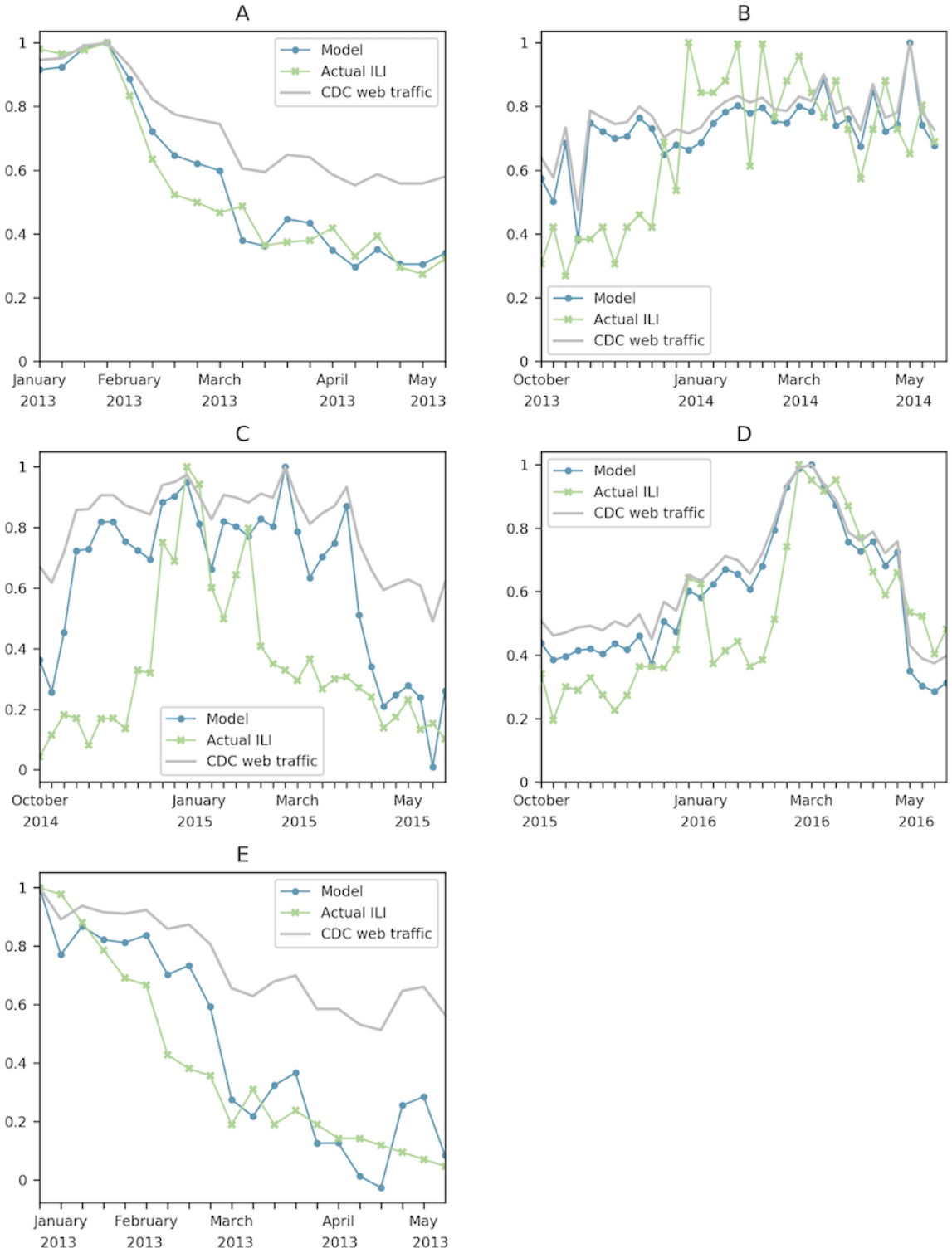
[a]*All* refers to the aggregation of all 10 pages.

[b]*FVST* refers to the aggregation of the FluView, Symptoms, and Treatment pages.

Figure 6 shows both successes and failures at the state level. Adding all the pages together, we were able to obtain r^2 values of 0.930 and 0.801 for Texas and Wisconsin, respectively, for the 2012 to 2013 influenza season. For the 2013 to 2014 season, the highest r^2 value was 0.187 for Wisconsin. For the 2014 to 2015 season, the highest r^2value was 0.322 for Missouri. For the 2015 to 2016 season, the highest r^2 value was 0.647 for North Carolina.

XSL•FO
RenderX

Figure 6. Different states during different seasons. (A) Texas, 1-week shift, 2012 to 2013. (B) Wisconsin, 2013 to 2014. (C) Missouri, 2014 to 2015. (D) North Carolina, 2015 to 2016. (E) Wisconsin 2012 to 2013. The r^2 values of each of these models ranged from 0.187 to 0.930. These models aggregated all 10 pages, and the success varied by state. CDC: Centers for Disease Control and Prevention; ILI: influenza-like illness.



We were not surprised that Texas had the best fit. Texas was the only state we included that provided ILI data not only for the typical influenza season but also for the off-season. These additional data likely contributed to the success of the Texas models. In keeping with our nowcasting scenario, we only included data available during the study period. During that period, Texas was the only state that provided off-season ILI data. These data have since been made available from other states, but the availability was not present during the study. The lack of success we encountered in modeling Maine was also expected because of Maine's outlier behavior in ILI, having values considerably lower and out of pattern with other states. The models in Figure 6 included all 10 pages aggregated together. However, as indicated by the individual state results,

XSL•FO
RenderX

this does not always lead to the best fit. Successful models often include a combination of select pages (such as FluView, Symptoms, and Treatment) but not an aggregation of all 10. Furthermore, aside from Texas, we did not have ILI data for the states outside of the typical flu season. Without these additional data, we are unable to determine how strongly the lower page views in the off-season correlate with off-season ILI.

We then shifted the ILI data forward by 1 week. The regression analysis yielded 7 state/season combinations with r^2 values greater than 0.7 (Table 4). The table also includes both the regular RMSE and NRMSE.

Table 4. States with models that had an r^2 value greater than 0.7 when aggregating all 10 pages and shifting the influenza-like illness data forward by 1 week. The regular and normalized root mean square errors are also displayed.

State	Season	r^2	Root mean square error	Normalized root mean square error
Texas	2012-2013	0.930	0.667	0.067
New Jersey	2012-2013	0.832	0.767	0.117
New Mexico	2015-2016	0.771	1.184	0.197
California	2012-2013	0.746	0.797	0.129
Wisconsin	2012-2013	0.727	0.626	0.153
North Carolina	2015-2016	0.708	1.028	0.204
Missouri	2012-2013	0.702	1.039	0.165

Adding only the FluView, Symptoms, and Treatment pages, we obtained an r^2 value of 0.7 or greater for 6 state/season combinations. For the 2013 to 2014 season, the highest r^2 values were 0.612 for California and 0.568 for Wisconsin. Although this is still less than desired, it is a vast improvement from the r^2 values found from adding all 10 pages. For the 2014 to 2015 season, the highest r^2 was 0.575 for Missouri. Again, although the correlation appears to be weak, it is a stronger correlation than taking all 10 pages together. Using the same 3 pages and implementing a 1-week shift, we obtained an r^2 value of 7 or greater for 10 state/season combinations. For the 2014 to 2015 season, the highest r^2 value was 0.548 for Missouri.

State Influenza-Like Illness Data Availability

The purpose of this study was to demonstrate the viability of near real-time nowcasting during the influenza seasons from 2013 to 2016. To maintain the premise of nowcasting, we chose states with publicly available data, or data available on request, during the period of the study. During the study period, state ILI data were not readily available on the CDC website. Instead, we had to rely on data available through state health-related organizations for each state. In addition, throughout the course of influenza seasons, ILI numbers are often updated as delayed data are reported and made available. However, because we are focusing our study on a nowcasting scenario, we do not consider the ILI data from those seasons as they are reported today but rather as they were reported during the study period.

Model Failures

We generally found the models to be successful when considering pages most closely related to typical health-seeking behavior and when considering each flu season individually. When trying to model multiple influenza seasons together, we had a number of unsuccessful models. Considering all pages and national ILI data, the model combining the 2012 to 2013 and 2013 to 2014 influenza seasons had an r^2 value of 0.061 and RMSE of 0.553. The combined 2013 to 2014 and 2014 to 2015 model had an r^2 value of 0.241 and RMSE of 0.208. The combined 2014 to 2015 and 2015 to 2016 model had an r^2 value of 0.251 and RMSE of 0.286. At the state level, combining all pages resulted in a number of unsuccessful models. For the 2013 to 2014 season, the Wisconsin model had an r^2 value of 0.187 and RMSE of 0.523. For the 2014 to 2015 season, the Missouri model had an r^2 value of 0.322 and RMSE of 1.845. Model failures not included in this section can be found in Multimedia Appendix 6.

We speculate that a number of factors could contribute to these negative results. Although influenza is a seasonal disease, similar strains can span multiple years, affecting the susceptible populations in subsequent years. Our data stream may be biased toward individuals with more awareness of the CDC. Furthermore, individuals who search for influenza information in one season may not search for that information the next year. Finally, with the exception of Texas, we only have ILI data for the influenza season itself. Thus, although we do have internet data for off-season influenza page views, we do not have corresponding ILI data.

Discussion

Conclusions

Internet surveillance data have proven beneficial in predicting ILI incidence during flu seasons. However, our results show that the benefit of internet data streams on informing disease is inconclusive; that is, this study shows that the CDC website traffic can be informative in some cases (eg, national level) but not in others (eg, state level). To determine the extent, we must return to our original research questions.

Research Question 1

Given the successes of some of our models, we can conclude that CDC page view data can be used as an additional data

XSL•FO
RenderX

source for monitoring disease incidence in some cases (eg, at the national level). The degree to which these data can be used appears to rely on the page selection and time frame. The results of the best models varied across geographic and temporal resolutions, but some trends were consistent in most cases. We obtained successful nowcasts when selecting pages related to topics most commonly used for web-based health queries (specific diseases and treatments) during the time span of a typical influenza season. Longer time spans and pages less associated with specific diseases and treatments led to less successful models. Outlier behavior, such as the ILI data in Maine, affected our models and resulted in less successful models than states with ILI curves exhibiting expected behavior. These results can assist others in selecting appropriate supplemental datasets for disease surveillance as well as appropriate spatial and temporal resolutions.

Research Question 2

We obtained our most successful results using a 1-week shift. Moreover, 2-week shifts were successful in some cases but were overall less correlated than 1-week shifts (Multimedia Appendix 6). Using no shift at all proved successful in some cases but not in others. We surmise that the shift required for the best fit depends on the incubation period for the disease in question as well as the period of reporting. The CDC internet data are available daily; however, ILI data are available weekly, so we are limited in the types of shifts we can apply to the datasets. Another factor that could contribute to the need for a 1-week shift is the amount of time between the page view and the subsequent visit to a health care center. If there are one or more days between the page view and the visit, then these 2 events could occur during different weeks. Shifting the data by 1 week accounts for this behavior.

Future Work

We conclude that more studies on internet data streams are needed to understand when and why internet data work. Our methods are consistent with other feasibility studies and provide insight into the conditions under which internet data streams may inform influenza models. Future work should include rigorously testing the predictive power of the models by separating data into training and testing sets [6].

More studies on geographic resolution could provide a better insight into why some models outperform others at various spatial resolutions. National models across single influenza seasons performed well, with each season included in the study having at least one model with an r^2 value greater than 0.75. We attribute the national model successes to the representation of all 50 states. Internet access may not be as prevalent in all states, but the inclusion of all 50 states allows for more data to be considered. Likewise, the census division models performed well, with overall r^2 values greater than those achieved from national models. Each census division had at least one model in the study with an r^2 value greater than 0.79. We attribute these successes to not only the inclusion of all states but also the division into geographic areas. There are instances in which a person may live in one state and seek medical care in another, perhaps because of working in a neighboring state. These instances are not accounted for by simply looking at states but can be accounted for by considering several neighboring states for 1 model. At the state level, models were overall less successful than at national and census division levels, but each state considered in the study had at least one model with an r^2 value greater than 0.65, and all but Maine had models with an r^2 value greater than 0.75. We attribute the overall lower success of state models to a combination of varying levels of internet access across populated and rural areas, the possibility of people living near neighboring states seeking health care in another state, and the inconsistencies in data availability during the study period. As our study focused on using data sources available during the study, we were limited in the states we could model because of the scarcity of the data.

More studies on temporal resolution could provide a better insight into how best to model seasonal diseases over multiple seasons. Models across multiple seasons were not successful, which we attribute in part to the off-season ILI data being unavailable during the study period. As influenza is a seasonal disease, modeling multiple seasons with 1 model may not be the correct approach, and our multiseason models support this idea. However, more exhaustive studies are needed to draw definitive conclusions on the appropriate spatial resolution for modeling influenza.

Acknowledgments

This work was supported by the Los Alamos National Laboratory Information Science and Technology Institute and the National Institutes of Health, the National Institute of General Medical Sciences, and the Modeling for Infectious Disease Agent Study program under grant U01-GM097658-071. Matthew Biggerstaff of the CDC provided integral insight and support. Curt Canada and the Data Science at Scale Summer School were integral to this project. Stephen Wirkus, Abigail Hunter, and Catherine S Plesko provided recommendations and clarifications. Jonathan Woodring, David H Rogers, and Francesca Samsel provided technical and visual assistance. Los Alamos National Laboratory is operated by Triad National Security, Limited Liability Company, for the National Nuclear Security Administration for the US Department of Energy (contract number: 89233218NCA000001).

Authors' Contributions

WC, GF, and SD conceptualized the project and performed data analysis. GF performed data curation. WC wrote the manuscript. GF and SD edited the manuscript. WC created the visualizations.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Names and date ranges of web pages used in this study.
[DOCX File , 14 KB - jmir_v22i7e14337_app1.docx]

Multimedia Appendix 2
Sources for state influenza-like illness data used in this study.
[DOCX File , 15 KB - jmir_v22i7e14337_app2.docx]

Multimedia Appendix 3
Influenza-like illness data for the nine census divisions.
[XLS File (Microsoft Excel File), 55 KB - jmir_v22i7e14337_app3.xls]

Multimedia Appendix 4
State influenza-like illness data.
[XLS File (Microsoft Excel File), 33 KB - jmir_v22i7e14337_app4.xls]

Multimedia Appendix 5
The nine US census divisions, listing all states in each division.
[DOCX File , 15 KB - jmir_v22i7e14337_app5.docx]

Multimedia Appendix 6
Comprehensive list of all models in this study not included in the main text. The list includes model successes and model failures.
[DOCX File , 47 KB - jmir_v22i7e14337_app6.docx]

References

1. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. PLoS Comput Biol 2015 May;11(5):e1004239 [FREE Full text] [doi: 10.1371/journal.pcbi.1004239] [Medline: 25974758]
2. Molinari NM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 2007 Jun 28;25(27):5086-5096. [doi: 10.1016/j.vaccine.2007.03.046] [Medline: 17544181]
3. Putri WC, Muscatello DJ, Stockwell MS, Newall AT. Economic burden of seasonal influenza in the United States. Vaccine 2018 Jun 22;36(27):3960-3966. [doi: 10.1016/j.vaccine.2018.05.057] [Medline: 29801998]
4. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google flu trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
5. Centers for Disease Control and Prevention. Glossary of Influenza (Flu) Terms URL: https://www.cdc.gov/flu/about/glossary.htm [accessed 2019-10-11]
6. Priedhorsky R, Osthus D, Daughton AR, Moran KR, Generous N, Fairchild G, et al. Measuring global disease with Wikipedia: success, failure, and a research agenda. CSCW Conf Comput Support Coop Work 2017;2017:1812-1834 [FREE Full text] [doi: 10.1145/2998181.2998183] [Medline: 28782059]
7. Generous N, Fairchild G, Deshpande A, del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLoS Comput Biol 2014 Nov;10(11):e1003892 [FREE Full text] [doi: 10.1371/journal.pcbi.1003892] [Medline: 25392913]
8. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]
9. Polgreen PM, Chen Y, Pennock DM, Nelson FD, Weinstein RA. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
10. Kim E, Seok JH, Oh JS, Lee HW, Kim KH. Use of Hangeul Twitter to track and predict human influenza infection. PLoS One 2013;8(7):e69305 [FREE Full text] [doi: 10.1371/journal.pone.0069305] [Medline: 23894447]
11. Baltrusaitis K, Brownstein JS, Scarpino SV, Bakota E, Crawley AW, Conidi G, et al. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. BMC Infect Dis 2018 Aug 15;18(1):403 [FREE Full text] [doi: 10.1186/s12879-018-3322-3] [Medline: 30111305]

XSL•FO
RenderX

12. Brownstein JS, Chu S, Marathe A, Marathe MV, Nguyen AT, Paolotti D, et al. Combining participatory influenza surveillance with modeling and forecasting: three alternative approaches. JMIR Public Health Surveill 2017 Nov 1;3(4):e83 [FREE Full text] [doi: 10.2196/publichealth.7344] [Medline: 29092812]
13. Pew Research Center. 2014 Feb 28. Health Fact Sheet URL: https://www.pewresearch.org/internet/health-fact-sheet-copy/ [accessed 2016-07-11]
14. Fox S. Pew Research Center. 2011. Health Topics URL: http://pewInternet.org/2011/02/01/health-topics-2/ [accessed 2016-07-11]
15. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
16. Chretien J, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. PLoS One 2014;9(4):e94130 [FREE Full text] [doi: 10.1371/journal.pone.0094130] [Medline: 24714027]
17. Perrotta D, Tizzoni M, Paolotti D. Using Participatory Web-Based Surveillance Data to Improve Seasonal Influenza Forecasting in Italy. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia p. 303-310. [doi: 10.1145/3038912.3052670]
18. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. Proc Natl Acad Sci U S A 2012 Dec 11;109(50):20425-20430 [FREE Full text] [doi: 10.1073/pnas.1208772109] [Medline: 23184969]
19. Lampos V, Zou B, Cox IJ. Enhancing Feature Selection Using Word Embeddings: The Case of Flu Surveillance. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia p. 695-704. [doi: 10.1145/3038912.3052622]
20. Xu D, Liu Y, Zhang M, Ma S, Cui A, Ru L. Predicting Epidemic Tendency through Search Behavior Analysis. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. 2011 Presented at: IJCAI'11; July 16-22, 2011; Barcelona, Spain URL: https://www.ijcai.org/Proceedings/11/Papers/393.pdf
21. Jia-xing B, Ben-fu L, Geng P, Na L. Gonorrhea Incidence Forecasting Research Based on Baidu Search Data. In: International Conference on Management Science and Engineering 20th Annual Conference Proceedings. 2013 Presented at: ICMSE'13; July 17-19, 2013; Harbin, China p. 36-42. [doi: 10.1109/icmse.2013.6586259]
22. Bardak B, Tan M. Prediction of Influenza Outbreaks by Integrating Wikipedia Article Access Logs and Google Flu Trend Data. In: Proceedings of the 15th International Conference on Bioinformatics and Bioengineering. 2015 Presented at: BIBE'15; November 2-4, 2015; Belgrade, Serbia p. 1-6. [doi: 10.1109/bibe.2015.7367640]
23. Biggerstaff M, Alper D, Dredze M, Fox S, Fung IC, Hickmann KS, Influenza Forecasting Contest Working Group. Results from the centers for disease control and prevention's predict the 2013-2014 Influenza Season Challenge. BMC Infect Dis 2016 Jul 22;16:357 [FREE Full text] [doi: 10.1186/s12879-016-1669-x] [Medline: 27449080]
24. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
25. Zhang Q, Perra N, Perrotta D, Tizzoni M, Paolotti D, Vespignani A. Forecasting Seasonal Influenza Fusing Digital Indicators and a Mechanistic Disease Model. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3-7, 2017; Perth, Australia p. 311-319. [doi: 10.1145/3038912.3052678]
26. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6:- [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
27. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
28. Lampos V, Miller AC, Crossan S, Stefansen C. Advances in nowcasting influenza-like illness rates using search query logs. Sci Rep 2015 Aug 3;5:12760 [FREE Full text] [doi: 10.1038/srep12760] [Medline: 26234783]
29. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci USA 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
30. Johnson HA, Wagner MM, Hogan WR, Chapman W, Olszewski RT, Dowling J, et al. Analysis of web access logs for surveillance of influenza. Stud Health Technol Inform 2004;107(Pt 2):1202-1206. [Medline: 15361003]
31. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: SCIPY'10; June 28 - July 3, 2010; Austin, Texas p. 92-96 URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf

Abbreviations

CDC: Centers for Disease Control and Prevention
ILI: influenza-like illness
NRMSE: normalized root mean square error
OLS: ordinary least squares
RMSE: root mean square error

XSL•FO
RenderX

Edited by G Eysenbach; submitted 10.04.19; peer-reviewed by M Biggerstaff, E Murray, E Lau; comments to author 10.09.19; revised version received 29.01.20; accepted 22.03.20; published 03.07.20.

Please cite as:
Caldwell WK, Fairchild G, Del Valle SY
Surveilling Influenza Incidence With Centers for Disease Control and Prevention Web Traffic Data: Demonstration Using a Novel Dataset
J Med Internet Res 2020;22(7):e14337
URL: https://www.jmir.org/2020/7/e14337
doi:10.2196/14337
PMID:32437327

©Wendy K Caldwell, Geoffrey Fairchild, Sara Y Del Valle. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of Volume of News Reporting and Opioid-Related Deaths in the United States: Comparative Analysis Study of Geographic and Socioeconomic Differences

Yulin Hswen[1,2], ScD; Amanda Zhang[3], BA; Clark Freifeld[4], PhD; John S Brownstein[2,5], PhD

[1]University of California San Francisco, Department of Epidemiology and Biostatistics, Bakar Computational Health Sciences Institute, San Francisco, CA, United States

[2]Computational Epidemiology Lab, Harvard Medical School, Boston, MA, United States

[3]Harvard University, Cambridge, MA, United States

[4]Northeastern University, Boston, MA, United States

[5]Innovation Program, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Yulin Hswen, ScD
University of California San Francisco
Department of Epidemiology and Biostatistics
Bakar Computational Health Sciences Institute
505 Parnassus Ave
San Francisco, CA, 94143
United States
Phone: 1 6177751889
Email: yhswen@gmail.com

Abstract

Background: News media coverage is a powerful influence on public attitude and government action. The digitization of news media covering the current opioid epidemic has changed the landscape of coverage and may have implications for how to effectively respond to the opioid crisis.

Objective: This study aims to characterize the relationship between volume of online opioid news reporting and opioid-related deaths in the United States and how these measures differ across geographic and socioeconomic county-level factors.

Methods: Online news reports from February 2018 to April 2019 on opioid-related events in the United States were extracted from Google News. News data were aggregated at the county level and compared against opioid-related death counts. Ordinary least squares regression was used to model opioid-related death rate and opioid news coverage with the inclusion of socioeconomic and geographic explanatory variables.

Results: A total of 35,758 relevant news reports were collected representing 1789 counties. Regression analysis revealed that opioid-related death rate was positively associated with news reporting. However, opioid-related death rate and news reporting volume showed opposite correlations with educational attainment and rurality. When controlling for variation in death rate, counties in the Northeast were overrepresented by news coverage.

Conclusions: Our results suggest that regional variation in the volume of opioid-related news reporting does not reflect regional variation in opioid-related death rate. Differences in the amount of media attention may influence perceptions of the severity of opioid epidemic. Future studies should investigate the influence of media reporting on public support and action on opioid issues.

(J Med Internet Res 2020;22(7):e17693) doi:10.2196/17693

KEYWORDS

opioid epidemic; news media; geographic; socioeconomic; addiction; overdose

XSL•FO
RenderX

Introduction

Each day in the United States, more than 130 people die from opioid overdose and its death toll only continues to rise [1]. The unprecedented magnitude and growth of the opioid crisis have given it a prominent place in news reporting. As mass media is a powerful influence on public opinion, monitoring trends and differences in the volume of opioid-related news stories may offer important insights for addressing this crisis.

Analyzing the content of popular news coverage is an established method for studying the public discourse surrounding health and social issues and how this discourse may influence public attitudes [2-10]. Large bodies of research have demonstrated the agenda-setting function of news media [11,12]. By focusing coverage on some topics rather than on others, news media can influence which issues audiences perceive to be more important and in more need of response. For example, an analysis of over 25 years of General Social Survey data showed that greater rates of newspaper reading and television viewing were associated with higher likelihoods of public desire of funding to be directed to addiction [6]. In another study, a time-series analysis of changing public opinion about the importance of illegal drug use in the United States showed that public views were largely driven by news media describing drug abuse as a *crisis* [7].

Of particular interest is how differences in the volume of media attention may translate into differences in government policy, funding, and resource deployment, and thereby affect particular geographic regions and social groups. For example, a content analysis of 100 popular press articles revealed that differential media coverage of whites and blacks in the opioid epidemic of the 2000s may have led to different public responses and policy interventions [13]. Another study found evidence that biased media coverage directed policy to be disproportionately aimed at specific stereotypes of drug users [14]. In these ways, the uneven landscape of opioid media reporting may play a major role in the differences in public health outcomes of the current US opioid epidemic.

As the opioid crisis is a relatively new public health problem, previous research on media coverage of the opioid epidemic in the United States is limited. Borwein et al [15] found that print news media in North America portrayed the opioid analgesic oxycodone as a social problem, which coincided with the reduction in oxycodone prescriptions by doctors in Nova Scotia, Canada, but the authors were not able to determine a causal relationship. McGinty et al [9] studied how opioid analgesic use was depicted by the US print and television media from 1998 to 2012, revealing that in the United States, the issue was commonly framed as a criminal justice problem requiring legal solutions rather than a remediable public health issue. Lillian Seklir and Lori Dorfman [5] found contrarily that in media coverage in Northern California, opioid addiction is routinely presented as both a public health and a criminal justice issue, but public health advocates and medical practitioners are largely absent from the coverage.

In this study, we developed a platform called DrugMap.org, a novel multistream real-time surveillance system that aggregates online news reports on opioid-related events, to investigate geographic and socioeconomic differences in the volume of opioid-related news reporting and opioid-related deaths in the United States. Our study expands on current research in four important ways: (1) Our real-time data set allows us to study the most recent (2018-2019) reports to uncover relevant trends in this fast-evolving public health issue. (2) Our expanded data set includes online news sources, which are significant because 93% of US adults report reading news online [16]. (3) The scope of our study encompasses all but 4 US states (Montana, North Dakota, South Dakota, and Wyoming). And (4) we dig more deeply into news trends by associating them with epidemiological data and study how differences in socioeconomic status and geography may contribute to potential discrepancies between news media and the true nature of the opioid crisis.

Methods

DrugMap Data

News Reports Mentioning Opioids From DrugMap

News reports about opioids from February 2018 to April 2019 were collected from DrugMap.org, a multistream news archive platform we developed that captures news media data using automated scripts. Using the Google News API, DrugMap extracts all online news articles from Google News whose body text contained key opioids and opioid-related terms that were developed in consultation with medical toxicologists: *opioid, oxycontin, dilaudid, suboxone, oxycodone, morphine, methadone, buprenorphine, hydrocodone, heroin, fentanyl, naloxone, vicodin, percocet, actiq, duragesic, sublimaze, naltrexone, codeine, tylox, tramadol, tapentadol, ultram, percodan, robitussin, demerol, roxicet, avinza, butrans, dolophine, embeda, exalgo, nucynta, opana, kadian, oxymorphone, levorphanol, dihydrocodeine,* and *zohydro.* Although we included a detailed list of terms, this search list may not be fully comprehensive. We did not limit search topics to opioid drug overdose or deaths and included all opioid-related news, in line with previous drug news studies that modeled drug epidemics based on volume of all media related to the drug of interest [17].

Geographic coordinates (latitude and longitude) were assigned to each report using a previously established geolocation engine, CLIFF-CLAVIN, a text mining tool that parses news articles to retrieve place mentions [18]. The location where the reported opioid-related event occurred was determined using place names mentioned in the news report text. US county information was derived by inputting latitude and longitude into the census geocode Python library [19].

The number of opioid-related news reports retrieved from DrugMap was 92,124, spanning from February 7, 2018 to April 8, 2019. Duplicate online news reports, news reports with the same body text and URL, and articles in geographies outside the United States were excluded. News reports that reported a granularity larger than state level (eg, New England) were also removed from the analysis as these articles could not be defined to a specific county. After data cleaning, the total number of

XSL•FO
RenderX

reports mentioning opioids was 35,758 and represented all 50 states and 1789 unique counties. The data covered 56.93% (1789/3142) of all US counties. Of the 1789 counties represented by the DrugMap data, 14.76% (464/3142) also contained Centers for Disease Control and Prevention (CDC) opioid-related death data, and this subset was used in the regression analysis. The 464 counties represent 46 states plus Washington DC, with the 4 states not included being Montana, North Dakota, South Dakota, and Wyoming. To account for variation in county population, adjusted news volume was defined as the number of news reports per population and was calculated by dividing total report count by county population times 100,000 total population.

County-Level Opioid-Related Deaths

Population-adjusted rates of opioid-related deaths in 2017 were obtained from the Multiple Cause of Death data set produced by the Division of Vital Statistics of the National Center for Health Statistics at the CDC, accessible through the WONDER portal of the CDC [1]. This CDC data set was compiled from death certificates for US residents, with each entry containing a single underlying cause of death and up to 20 additional contributing causes. Because death certificates for drug overdose are not complete in identifying drugs involved (in 2017, 12% of drug overdose death certificates did not provide information on specific drugs involved), we aimed to be broad in our inclusion criteria for opioid-related deaths, querying for all deaths that contained an opioid in the multiple-cause-of-death category rather than the underlying-cause-of-death category. Specifically, the criteria for inclusion were the following ICD-10 codes: opioids (T40.0, T40.1, T40.2, T40.3, T40.4, or T40.6); natural and semisynthetic opioids (T40.2); methadone (T40.3); synthetic opioids, other than methadone (T40.4); and heroin (T40.1). Death rates were given in units of deaths per 100,000 total population in county. Retail opioid prescriptions dispensed per 100 persons were collected from the CDC Opioid Prescribing Rate database [20].

County-Level Demographic and Socioeconomic Data

Demographic and socioeconomic variables were used to assess how media reporting and death rate differed in their relationship to external variables. Data on census region membership (Northeast, South, Midwest, and West), county population, high-school attainment in county, percent white in county, median income in county, and percent rurality in county were obtained from the US Census Bureau of Statistics [21].

Descriptive and Visual Analysis

To examine spatial trends, population-adjusted opioid news volume from DrugMap was visualized on a geographic map of the United States. To examine social and demographic trends, 2 heatmaps were created: one showing opioid news volume per 100,000 total population in county on the axes of rurality and educational attainment and another showing opioid-related death rates per 100,000 total population in county on the same axes. These maps were created in the following manner: first, we identified counties that contained both news data and death data. Of this subset of counties, we created 25 groups based on quintiles of percent rurality in county and high-school graduation rate in county. For each group, we computed the average opioid news rate and average opioid-related death rate by taking the mean of the individual rates of the counties in each group. This allowed us to visualize differences in associations with socioeconomic factors for opioid-related deaths versus opioid reporting.

Statistical Analyses

Ordinary least squares regression was used to model county death rate and county report rate. A Poisson model was initially considered with population as an offset, but upon evaluation this model was rejected because of overdispersion, a violation of model assumptions. We then chose ordinary least squares after checking that the assumptions of normality of the residual terms were met. Counties with both DrugMap and CDC death data were used for this analysis. In Model 1, we modeled opioid-related death rate in county as a linear function of county geographic, demographic, and socioeconomic variables to identify area-level characteristics of greater opioid-related mortality. In Model 2, we used ordinary least squares regression to model news reporting rate in county as a linear function of county death rate and county variables to identify areas of higher volume of news reporting. A number of studies have described county characteristics associated with opioid use [22,23]. These include educational attainment, racial composition, physician supply, opioid prescribing rate, income, and rurality. These variables were included as controls in Model 1 and Model 2. All analyses were conducted with Python statsmodels [24].

Results

Descriptive and Visual

The median number of reports per state was 565.5 (range 7-4169) and the median number of reports per county was 5 (range 1-1631). At the county level, the median was 9.35 reports per 100,000 total population (range 0.38-430.57 per 100,000 total population). Figure 1 displays the population-adjusted volume of DrugMap reporting across the United States by percentile category. As can be seen, opioid reporting volume is higher in the Northeast, especially in the areas of Maine, New Hampshire, Vermont, and Massachusetts. Other states with high opioid reporting relative to population size are Alaska and Nevada.

XSL•FO
RenderX

Figure 1. US Map of DrugMap reports in county divided by county population, 2018-2019.

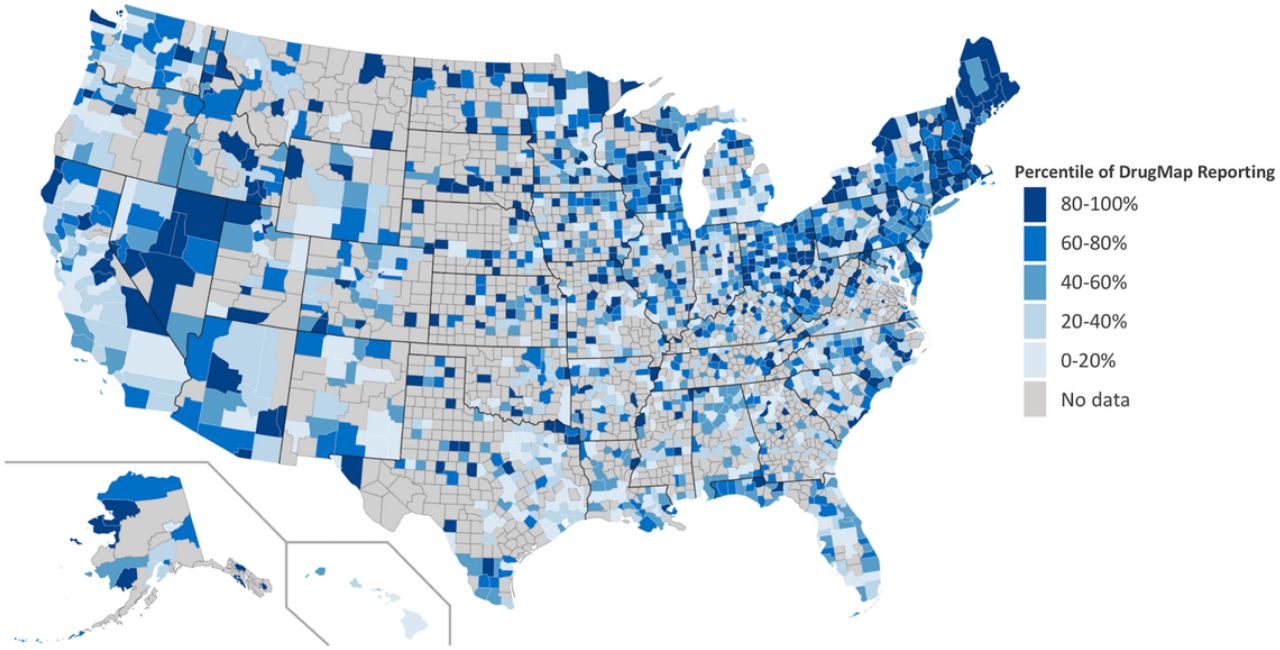
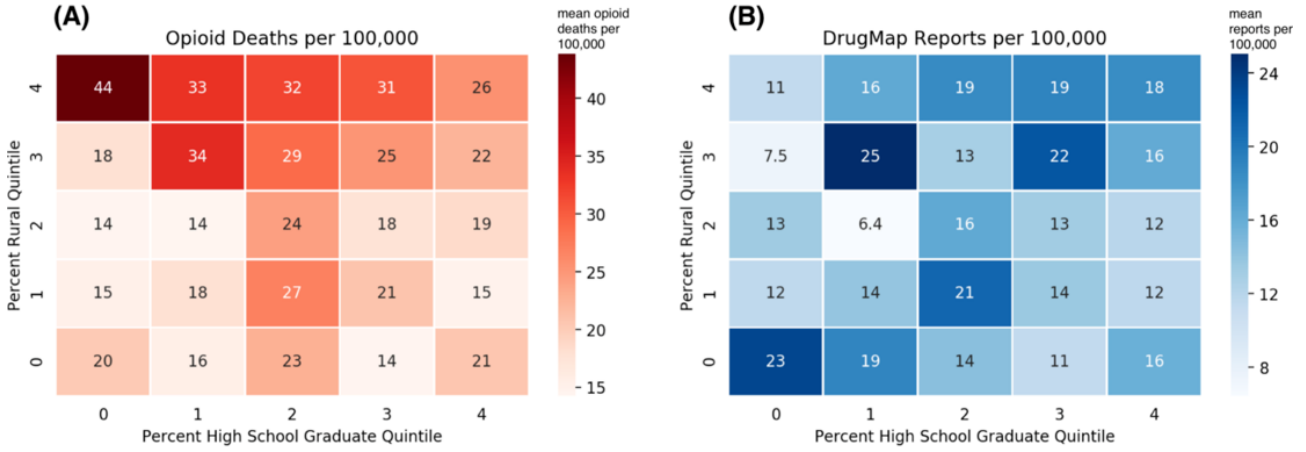


Figure 2A shows the mean value of opioid-related death rate on the axes of percent rurality quintile and high-school graduation rate quintile, and Figure 2B shows the mean value of opioid reporting on the same axes. In the mortality heatmap, counties in higher quintiles of rurality and lower quintiles of education are shaded more darkly, indicating greater opioid-related death rates in rural and less educated populations. By contrast, the heatmap of DrugMap news reporting shows more variation and does not reflect the pattern observed for opioid-related death. When we compare the distribution of online opioid news from DrugMap with the US populations, we find that for rurality, the news data represented 53.7% of the first quintile (least rural), 15.3% of the second quintile, 4% of the third quintile, and 0.4% of the fourth/fifth quintile (most rural). For high-school graduation rate, the news data represented 2.6% of the first quintile (least educated), 8.5% of the second quintile, 18.2% of the third quintile, 23.9% of the fourth quintile, and 18.9% of the fifth quintile (most educated). These differences show that there exists greater news coverage about opioids relative to opioid mortality in urban and more educated areas.

Figure 2. Heatmap showing distribution of opioid-related death rate (A) 2017 and DrugMap report rate (B) 2018-2019 by percent rural and percent high-school graduate.



Regression

Model 1 regression results are presented in Table 1. For every percentage point increase in percent white at the county level, there was on average 28.02 less opioid-related deaths per 100,000 total population (P<.001), controlling for median income, rurality, high-school graduate rate, opioid prescribing rate, and geographic region. For every percentage point increase in percent rurality at the county level, there was on average 0.14 more opioid-related deaths per 100,000 total population (P<.01), controlling for all other variables. For every percentage point increase in percent high-school graduation rate at the county level, there was on average 0.86 less opioid-related deaths per 100,000 total population (P<.001), controlling for all other variables. Compared with the Midwest, the West had on average 9.99 less opioid-related deaths per 100,000 total population (P<.001), whereas there was no significant differences for the

XSL•FO
RenderX

Northeast or South. Median income and opioid prescribing rate at the county level were not significantly associated with opioid-related death rate.

Model 2 regression results are presented in Table 1. Opioid mortality was positively associated with opioid news reporting; for every additional death per 100,000 total population at the county level, there was on average 0.40 more news reports per 100,000 total population (P<.001), controlling for median income, percent white, rurality, high-school graduate rate, opioid prescribing rate, and geographic region. For every dollar increase in median income at the county level, there was on average 0.0003 more news reports per 100,000 total population (P<.001), controlling for all other variables. For every percentage point increase in percent rurality at the county level, there was on average 0.17 less news reports per 100,000 total population (P<.01), controlling for all other variables. For every percentage point increase in percent high-school graduation rate at the county level, there was on average 0.65 more news reports per 100,000 total population (P<.01), controlling for all other variables. For every additional opioid prescription per 100 persons at the county level, there was on average 0.112 more news reports per 100,000 total population (P<.01), controlling for all other variables. Compared with the Midwest, the Northeast had 10.12 more news reports per 100,000 total population (P<.001), whereas the South and West exhibited no significant differences. Percent white at the county level was not significantly associated with news reporting frequency.

Table 1. Summary of regression results of Models 1 and 2[a,b].

Dependent variable	Model 1				Model 2			
	Death rate				Report rate			
	Coefficient	Standard error	*P* value	95% CI	Coefficient	Standard error	*P* value	95% CI
Intercept	84.202	18.575	<.001	47.701 to 120.703	–24.040	19.152	.21	–61.678 to 13.598
Death rate	N/A[c]	N/A	N/A	N/A	0.397	0.047	<.001	0.305 to 0.489
Median income	–0.0001	6.28×10^{-05}	.11	–0.000 to 2.32×10^{-05}	–0.0003	6.37×10^{-05}	<.001	–0.000 to –0.000
Percent white	28.019	6.431	<.001	15.382 to 40.656	0.410	6.652	.95	–12.662 to 13.483
Percent rural	0.143	0.054	<.01	0.036 to 0.249	–0.172	0.056	<.01	–0.282 to –0.061
Percent high-school graduate	–0.857	0.242	<.001	–1.332 to –0.381	0.649	0.248	<.01	0.162 to 1.135
Opioid prescribing rate	0.024	0.046	.55	–0.065 to 0.114	–0.112	2.438	.02	–0.202 to –0.022
Northeast	–0.786	2.054	.70	–4.823 to 3.251	10.115	2.078	<.001	6.031 to 14.200
South	–1.919	1.911	.32	–5.675 to 1.836	0.779	1.954	.69	–3.062 to 4.619
West	–9.991	2.418	<.001	–14.743 to –5.238	0.129	2.485	.96	–4.755 to 5.013

[a]Midwest was used as the reference group in both models.

[b]For both models, N=464.

[c]N/A: Not applicable.

Discussion

Results from our study show that opioid-related mortality and opioid-related news reporting were positively correlated with each other: increases in opioid-related death rate at the county level were associated with increases in news reports. However, opioid-related mortality and opioid-related news reporting differed in how they related to geographic, demographic, and socioeconomic factors. Opioid-related deaths are less common in the geographic region of the West and more common in communities that are less educated, have a greater proportion of whites, and are more rural. While this finding regarding how opioid *mortality* relates to geography, education, and racial breakdown is consistent with prior studies [25], our analysis suggests that opioid-related *news reporting* follows different trends. Areas with greater news reporting tended to be more educated, more urbanized, and located in the Northeast. These discrepancies in news reporting rate and death rate, which can be interpreted as a sign of the magnitude of the opioid crisis in a certain area, highlight potential differences in the relative attention of news coverage across regions, rural areas, and socioeconomic status. More urban and educated areas receive greater media attention even though the opposite pattern is observed for opioid-related deaths. Previous research has shown that greater media attention and coverage increase public and governmental responses [6,13]. This increased attention may lead the public and the government to devote more resources and develop policy intervention targeted to areas that are less affected by the opioid epidemic.

In recent years, mass media has become increasingly relevant because its online transmission is immediate and its reach is broad. Because online news is often the first point of contact people have with current events, online sources prime people's perceptions of the reported issues going forward and potentially override conflicting information received afterward [26]. For example, during the 1980s War on Drugs, media depiction of the cocaine epidemic as a "ghetto" problem is widely believed to have contributed to the public's perception of African

XSL•FO
RenderX

Americans and Latinos as addicts and criminals and white drug users as casualties and victims [13,27,28]. Another devastating case of differential media attention is the widely circulated news associating the measles vaccine with autism. News headlines prominently reported this association, and even though academics have long dispelled and discredited this link, this belief still persists because of the heightened media attention received by the original report and the lower media attention received by the studies that dismissed this association [29-31]. Through examining media trends for the current US opioid epidemic, we see that patterns of opioid media attention do not reflect the distribution of opioid-related deaths. This may lead to exacerbations of health disparities as previous studies have linked media attention with resource allocation. Lower media volume in areas with greater opioid use risk may reduce attention from government officials or public health agencies for these regions, leading to an increase in disparities. For instance, opioid-related deaths occur most in rural areas with lower education, suggesting that these areas are being affected the most. However, our results indicate that there may be less opioid-related news reporting and coming from these areas. This relative lack of news discussion could signal less awareness, interest, and investment in opioid issues in regions that are at greater risk.

Limitations

Results from this study were obtained at the area level based on county variables, so conclusions must be drawn at the ecological and not at the individual level. First, issues of collinearity across these demographic and socioeconomic variables prevented us from including all socioeconomic and demographic variables into the model. The specific variables selected in this study were based on previous research on opioid-related events. Second, our regression models provide information on associations and cannot be deemed causal. Third, while we adjusted news volume to consider differences in county population, the measures of baseline news output and readership may have been more direct control variables; however, these were unavailable during the study. Using these measures may have yielded different results.

There are benefits and drawbacks to our methodology of using opioid-related deaths as opposed to opioid-related overdose deaths for our analysis. Although the variable of opioid overdose deaths is more specific and perhaps more directly interpretable, overdose death data from the CDC are limited in important ways. Specifically, CDC data on drug overdose deaths are derived from information provided on death certificates, and not all death certificates for overdose include information on drugs involved.

We chose not to specify drug overdose as the underlying cause of death for opioid-related mortality in order to encompass a broader view of the impact of the opioid epidemic across the United States. Although our data are still subject to the aforementioned data limitations, the broader scope of our inclusion criteria allows us to achieve a more comprehensive analysis. These limitations in national databases are reflected in our analysis as counties with poorer reporting, which tend to also be more rural, and are thus excluded. However, even with greater representation in more urban areas, our results demonstrate a relative lack of media opioid reporting in rural and less educated areas, thereby supporting the existence of biases in media reporting on opioids.

Finally, our study did not examine the language or content of opioid-related news reporting, such as positive and negative connotation or types of vocabulary used. Textual examination of media content should be conducted in follow-up studies as previous research has shown that framing can impact public perceptions and judgments toward certain socioeconomic groups [8,12,13,32]. Furthermore, while our study supports the existence of differences in reporting volume, it does not measure the effect of such trends on government support and action on opioid issues. Future studies should be carried out to empirically investigate the relationship between variation in news reporting volume and variation in realized efforts.

Conclusion

Our study brings to light potential differences in the volume of media portrayal of the current opioid epidemic. Previous research and theory have demonstrated a strong link between media attention and its impact on public perception, future resource allocation, and population-level opioid use behaviors. Although our study did not directly test these consequences, our results provide evidence supporting the existence of these media disparities across the United States. Given the importance of news media in shaping public attitudes and government efforts, it is critical that news reporting on opioids is distributed equitably and accurately to prevent misperceptions. Future studies should investigate trends in the content of news reporting on the opioid epidemic and how framing of news stories may be affecting government support and action on opioid issues.

Acknowledgments

Funding for this study was provided by an NIH/NLM Institute grant (No. 3R01LM011965-03S1). The NIH/NLM had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Authors' Contributions

YH retrieved funding to conduct research; conducted literature searches; summarized previous research studies; designed the study; and wrote the protocol, the first draft, and the final manuscript. AZ conducted literature searches, provided summaries of previous research studies, designed the study, conducted the statistical analysis, and contributed to writing the manuscript. CF

XSL•FO
RenderX

retrieved the data, built the database, and edited the manuscript. JSB retrieved funding to conduct research and contributed to the edits. All authors have approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention (CDC). Multiple Cause of Death 1999-2018 on CDC WONDER Online Database. URL: https://wonder.cdc.gov/wonder/help/mcd.html [accessed 2020-06-12]
2. Danovaro-Holliday MC, Wood AL, LeBaron CW. Rotavirus Vaccine and the News Media, 1987-2001. JAMA 2002 Mar 20;287(11):1455. [doi: 10.1001/jama.287.11.1455]
3. Stryker JE, Solky BA, Emmons KM. A Content Analysis of News Coverage of Skin Cancer Prevention and Detection,1979 to 2003. Arch Dermatol 2005 Apr 01;141(4):491-496. [doi: 10.1001/archderm.141.4.491]
4. Fishman J, Ten Have T, Casarett D. Cancer and the media: how does the news report on treatment and outcomes? Arch Intern Med 2010 Mar 22;170(6):515-518. [doi: 10.1001/archinternmed.2010.11]
5. Lillian SP, Dorfman L. The Opioid Epidemic in the News: Findings From a Preliminary Analysis of Northern California Coverage. Berkeley, CA: Berkeley Media Studies Group; 2016.
6. Nielsen AL, Bonn S. Media Exposure and Attitudes Toward Drug Addiction Spending, 1975–2004. Deviant Behavior 2008 Oct 03;29(8):726-752. [doi: 10.1080/01639620701839492]
7. Fan DP. News Media Framing Sets Public Opinion That Drugs is the Country's Most Important Problem. Substance Use & Misuse 2009 Jul 03;31(10):1413-1421. [doi: 10.3109/10826089609063984]
8. Zhang MWB, Hong YX, Husain SF, Harris KM, Ho RCM. Analysis of print news media framing of ketamine treatment in the United States and Canada from 2000 to 2015. PLoS ONE 2017 Mar 3;12(3):e0173202. [doi: 10.1371/journal.pone.0173202]
9. McGinty EE, Kennedy-Hendricks A, Baller J, Niederdeppe J, Gollust S, Barry CL. Criminal Activity or Treatable Health Condition? News Media Framing of Opioid Analgesic Abuse in the United States, 1998–2012. PS 2016 Apr;67(4):405-411. [doi: 10.1176/appi.ps.201500065]
10. Soroka S, Maioni A, Martin P. What Moves Public Opinion on Health Care? Individual Experiences, System Performance, and Media Framing. Journal of Health Politics, Policy and Law 2013 Jun 21;38(5):893-920. [doi: 10.1215/03616878-2334656]
11. McCombs MC, Shaw DL. The agenda-setting function of mass media. Public Opinion Quarterly 1972 May;36(2):176-187.
12. Walsh-Childers K. Mass Media and Health: Examining Media Impact on Individuals and the Health Environment. London, UK: Routledge; 2016.
13. Netherland J, Hansen HB. The War on Drugs That Wasn't: Wasted Whiteness, "Dirty Doctors," and Race in Media Coverage of Prescription Opioid Misuse. Cult Med Psychiatry 2016 Jun 6;40(4):664-686. [doi: 10.1007/s11013-016-9496-5]
14. Taylor S. Outside the outsiders: Media representations of drug use. Probation Journal 2008 Dec;55(4):369-387. [doi: 10.1177/0264550508096493]
15. Borwein A, Kephart G, Whelan E, Asbridge M. Prescribing Practices Amid the OxyContin Crisis: Examining the Effect of Print Media Coverage on Opioid Prescribing Among Physicians. The Journal of Pain 2013 Dec;14(12):1686-1693.e1. [doi: 10.1016/j.jpain.2013.08.012]
16. Stocking G. Digital News Fact Sheet. Washington, DC: Pew Research Center; 2018 Jun. URL: https://www.journalism.org/fact-sheet/digital-news/ [accessed 2020-06-12]
17. Dasgupta N, Mandl KD, Brownstein JS. Breaking the News or Fueling the Epidemic? Temporal Association between News Media Report Volume and Opioid-Related Mortality. PLoS ONE 2009 Nov 18;4(11):e7758. [doi: 10.1371/journal.pone.0007758]
18. D'Ignazio C, Bhargava R, Zuckerman E, Beck L. CLIFF-CLAVIN: Determining geographic focus for news. NewsKDD: Data Science for News Publishing. 2014. URL: https://www.media.mit.edu/publications/cliff-clavin-determining-geographic-focus-for-news-articles/ [accessed 2020-06-12]
19. Freeman N. Census Geocode. 0.4.3.post1 ed 2020:2019.
20. US Centers for Disease Control and Prevention. U.S. County Prescribing Rates, 2017. 2017. URL: https://www.cdc.gov/drugoverdose/maps/rxcounty2017.html [accessed 2020-06-12]
21. U.S. Census Bureau. 2010 Decennial Census of Population and Housing. URL: https://www.census.gov/programs-surveys/decennial-census/decade/decennial-publications.2010.html [accessed 2020-06-12]
22. Edlund MJ, Steffick D, Hudson T, Harris KM, Sullivan M. Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain. Pain 2007;129(3):355-362. [doi: 10.1016/j.pain.2007.02.014]
23. Ghertner R, Groves L. The opioid crisis and economic opportunity: geographic and economic trends. ASPE Research Brief 2018:1-22 [FREE Full text]
24. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. 2010 Presented at: Proceedings of the 9th Python in Science Conference (SciPy 2010); July 2010; Austin, TX URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf

XSL•FO
RenderX

25. Dasgupta N, Beletsky L, Ciccarone D. Opioid Crisis: No Easy Fix to Its Social and Economic Determinants. Am J Public Health 2018 Feb;108(2):182-186. [doi: 10.2105/ajph.2017.304187]
26. Iyengar S, Kinder DR. News That Matters: Television and American Opinion. Chicago, IL: University of Chicago Press; 2010.
27. Reeves J, Campbell R. Cracked Coverage: Television News, the Anti-Cocaine Crusade, and the Reagan Legacy. Durham, NC: Duke University Press; 1994.
28. Cobbina J. Race and class differences in print media portrayals of crack cocaine and methamphetamine. Journal of Criminal Justice and Popular Culture 2008;15(2):145-167.
29. Smith MJ. Vaccination and autism 10 years on: why are parents still worried? Pediatric Health 2008 Dec;2(6):667-669. [doi: 10.2217/17455111.2.6.667]
30. Plotkin S, Gerber J, Offit P. Vaccines and Autism: A Tale of Shifting Hypotheses. CLIN INFECT DIS 2009 Feb 15;48(4):456-461. [doi: 10.1086/596476]
31. Taylor LE, Swerdfeger AL, Eslick GD, Arumugham V. Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. Vaccine 2014 Jun;32(29):3623-3629. [doi: 10.1016/j.vaccine.2014.04.085]
32. Orsini MM. Frame Analysis of Drug Narratives in Network News Coverage. Contemporary Drug Problems 2017 Aug 03;44(3):189-211. [doi: 10.1177/0091450917722817]

Edited by G Eysenbach; submitted 04.01.20; peer-reviewed by P Ning, J Black; comments to author 10.02.20; revised version received 15.04.20; accepted 13.05.20; published 10.07.20.

Please cite as:
Hswen Y, Zhang A, Freifeld C, Brownstein JS
Evaluation of Volume of News Reporting and Opioid-Related Deaths in the United States: Comparative Analysis Study of Geographic and Socioeconomic Differences
J Med Internet Res 2020;22(7):e17693
URL: http://www.jmir.org/2020/7/e17693/
doi:10.2196/17693
PMID:32673248

©Yulin Hswen, Amanda Zhang, Clark Freifeld, John S Brownstein. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Eating Disorder Awareness Campaigns: Thematic and Quantitative Analysis Using Twitter

Iranzu Viguria[1*], MD; Miguel Angel Alvarez-Mon[1,2,3*], MD, PhD; Maria Llavero-Valero[4], MD; Angel Asunsolo del Barco[5], MD, PhD; Felipe Ortuño[1], MD, PhD; Melchor Alvarez-Mon[2,6,7], MD, PhD

[1]Department of Psychiatry and Medical Psychology, Clinica Universidad de Navarra, Pamplona, Spain

[2]Department of Medicine and Medical Specialities, University of Alcala, Alcalá de Henares, Spain

[3]Department of Psychiatry and Medical Psychology, Hospital Universitario Infanta Leonor, Madrid, Spain

[4]Department of Endocrinology and Nutrition, Clinica Universidad de Navarra, Pamplona, Spain

[5]Department of Surgery, Medical and Social Sciences, University of Alcala, Madrid, Spain

[6]Internal Medicine and Immune System Diseases-Rheumatology Service, University Hospital Príncipe de Asturias, Alcala de Henares, Spain

[7]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Instituto Ramón y Cajal de Investigaciones Sanitarias, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Miguel Angel Alvarez-Mon, MD, PhD
Department of Medicine and Medical Specialities
University of Alcala
Campus Universitario - C/ 19, Av de Madrid, Km 33,600
Alcalá de Henares, 28871
Spain
Phone: 34 918854505
Email: maalvarezdemon@icloud.com

Abstract

Background: Health awareness initiatives are frequent but their efficacy is a matter of controversy. We have investigated the effect of the Eating Disorder Awareness Week and Wake Up Weight Watchers campaigns on Twitter.

Objective: We aimed to examine whether the Eating Disorder Awareness Week and Wake Up Weight Watchers initiatives increased the volume and dissemination of Twitter conversations related to eating disorders and investigate what content generates the most interest on Twitter.

Methods: Over a period of 12 consecutive days in 2018, we collected tweets containing the hashtag #wakeupweightwatchers and hashtags related to Eating Disorder Awareness Week (#eatingdisorderawarenessweek, #eatingdisorderawareness, or #EDAW), with the hashtag #eatingdisorder as a control. The content of each tweet was rated as medical, testimony, help offer, awareness, pro-ana, or anti-ana. We analyzed the number of retweets and favorites generated, as well as the potential reach and impact of the hashtags and the characteristics of contributors.

Results: The number of #wakeupweightwatchers tweets was higher than that of Eating Disorder Awareness Week and #eatingdisorder tweets (3900, 2056, and 1057, respectively). The content of tweets was significantly different between the hashtags analyzed ($P<.001$). Medical content was lower in the awareness campaigns. Awareness and help offer content were lower in #wakeupweightwatchers tweets. Retweet and favorite ratios were highest in #wakeupweightwatchers tweets. Eating Disorder Awareness Week achieved the highest impact, and very influential contributors participated.

Conclusions: Both awareness campaigns effectively promoted tweeting about eating disorders. The majority of tweets did not promote any specific preventive or help-seeking behaviors.

(J Med Internet Res 2020;22(7):e17626) doi:10.2196/17626

KEYWORDS

awareness campaigns; eating disorders; Twitter; social media

XSL•FO
RenderX

Introduction

Eating disorders belong to a wide group of disorders characterized by weight concern, body image preoccupation, and persistent disturbances in eating that impair health and psychosocial functioning [1]. These diseases occur worldwide, mostly commonly in young women, with an estimated lifetime prevalence of 0.8% for anorexia nervosa and binge eating disorder and 0.3% for bulimia nervosa [2,3]. The etiology of these disorders is complex and not fully understood. Nonetheless, many aspects of Western culture, such as the "thin ideal" or media pressure, promote an obsession with weight loss and body image [4].

Several strategies have been established to reduce the prevalence of eating disorders [5]. Health awareness initiatives have become a frequent intervention strategy, as high-visibility programs have stimulated the discussion of certain health risks and enhanced public exposure to these issues [6]. For instance, Eating Disorder Awareness Week is an annual event sponsored by a wide range of partners that aims to increase public recognition of the preventive and supportive options for individuals suffering from eating disorder behaviors [7]. However, some initiatives are not scheduled in advance, as was the case with Wake Up Weight Watchers. On February 6, 2018, many health care providers were concerned when Weight Watchers announced free memberships for teenagers, because dieting as a teenager increases the risk of developing eating disorders and risky behaviors [8,9]. Thus, the Wake Up Weight Watchers initiative was developed to counteract the potentially negative effects of adolescent dieting [10]. Although the efficacy of health awareness initiatives is a matter of controversy, analytical studies of those related to eating disorders are particularly scarce [11].

Notably**,** research using Twitter has been an effective assessment strategy for analyzing patient attitudes toward various medical topics, including mental health [12,13]. Additionally, Twitter has been used to assess the impact of disease awareness programs on users, especially young adults and adolescents [14]. Thus, this platform could help health care providers effectively reach many young people troubled by eating disorders [15]. It has been shown that medical interventions conducted through Twitter can modify health-related behaviors [16]. Therefore, this platform could provide a good setting for the promotion of healthy lifestyles among at-risk individuals and encourage them to get in touch with health providers [17]. Previous research has found that engaging in help-seeking behavior has been an important factor in protecting the mental health of young people [18]**.** Despite these efforts to help, evidence nevertheless suggests that those experiencing eating disorder thoughts and behaviors are less likely to seek help for their mental health difficulties [19]. One recent study even reported that body image was the second leading personal concern for which people between the ages of 18 and 25 years sought help online [20].

The aims of this study are the following: (1) Examine the volume of Twitter conversations related to eating disorders generated by the Eating Disorder Awareness Week and Wake Up Weight Watchers initiatives, and compare the dissemination of these tweets with tweets related to eating disorders that did not use the official hashtags of these campaigns. (2) Describe the main content themes of tweets from both initiatives and compare them to those of typical tweets related to eating disorders. (3) Investigate what content generates the most interest on Twitter. (4) Identify and categorize the most active and influential users among those that participate in these conversations.

Methods

Research Strategy

In this observational quantitative and qualitative study, we focused on searching for tweets that referred to two eating disorder campaigns in particular: Eating Disorder Awareness Week (#eatingdisorderawarenessweek, #EDAW, #eatingdisorderawareness) and Wake Up Weight Watchers (#wakeupweightwatchers). As a control, we simultaneously studied those tweets related to eating disorders stemming from the hashtag with the name of the disease (#eatingdisorder). The inclusion criteria for tweets were the following: (1) tweeted by a public account, (2) included the previously mentioned hashtags, (3) posted between February 24 and March 7, 2018, and (4) posted in English. The 12-day period was chosen to align our research with the dates of the two campaigns included in this study. Eating Disorder Awareness Week 2018 started February 26, 2018, and the Wake Up Weight Watchers initiative began in February 2018. In addition, we obtained the number of retweets and favorites each tweet generated, the date and time of each tweet, the potential reach and impact of each hashtag, as well as the user's profile description.

Search Tool and Data Collection

For this study, we used the Twitter Firehose (Gnip) data stream, which allows access to 100% of all public tweets that match a certain criteria (query) [21]. In our study, the search criteria were the previously mentioned hashtags. Tweet Binder, the search engine we employed, uses automatic machine learning text analysis algorithms, as well as node.js and PHP language; this enabled us to analyze tweets in a JavaScript Object Notation (JSON) format, which is used by Gnip.

Content Analysis Process

All 7468 retrieved tweets were directly inspected by two raters (authors IV and MAAM). First, we scanned all of the tweets, and excluded 455 tweets that provided information that was too limited (ie, tweets consisting mainly of hashtags), contained only pictures, or included hashtags of both initiatives. All remaining tweets were considered for content analysis. Second, we created a codebook based on our research questions, our previous experience in analyzing tweets, and what we determined to be the most common tweet themes. Third, IV and MAAM analyzed 300 tweets separately to test the suitability of the codebook. Discrepancies were discussed between the raters and with another author (MLV). After revising the codebook, the interrater reliability was reassessed with a different set of 159 tweets. As this resulted in adequate κ values (range 0.89-0.99), the raters then proceeded to perform a content

XSL•FO
RenderX

analysis of 1500 tweets, including 500 tweets each from the three different groups of hashtags, randomly selected. Each tweet, depending on the content, was rated as medical, personal testimony, pro-ana (tweets glorifying or encouraging ED behaviors or portraying them as a lifestyle choice), anti-ana (tweets confronting pro-ana ideas), awareness, or help offer. The coding categories were not mutually exclusive. If the contents of a tweet were repeated exactly or almost identically in other tweets, they were classified in the same way as the first tweet encountered. The classification criteria we used and examples of tweets according to category are shown in Table 1.

Table 1. Category definitions and examples of classification[a].

Categories and definitions	Examples of classification
Medical (clinical or epidemiological information)	"For #EatingDisorderAwarenessWeek remember that1) ppl with eating difficulties are not always underweight 2) it's not about wanting to diet and be thin 3) recovery is not as simple as "eat a bit more to recover" 4) ED's are life threatening illnesses"
Testimony (expression of experience)	• "When I was about to become a teenager, my parents put me on Weight Watchers. I have been battling Anorexia since then. Dieting is one of the most common antecedents of disordered eating in young people. #WakeUpWeightWatchers" • "We are warriors whether we are still battling this war or we have overcome and conquered it. We stand together in a battle which costs lives and relationships but we stand together. #EatingDisorderAwareness-Week" • "My younger sister has this too - as if she is cured and we can now resume normal life and carry on from where we left off. But everything has changed. #StateNotWeight #EatingDisorderAwarenessWeek" • "As someone who suffered from an ED after dieting I am so worried about promoting weight watchers to teenagers. It may be dangerous. #WakeUpWeightWatchers"
Pro-ana (pro–eating disorder content, promoting compensatory behaviors or disordered eating for weight loss)	• "Ohh wow my dad really bought me the scale I wanted. I'm so excited but at the same time it's better he doesn't know why I really wanted it so bad. #eatingdisorder #proana #ed #rexi" • "I can't eat because I'm allergic to calories. #ana #mia #anorexia #anorexic #anoressia #anoressica #anorexie #anorexique #vita #lavie #mystory #Mylife #conversation #quote #quotes #meme #memes #MEMES #anamia #proana #eatingdisorder #diary #mydiary #me #life" • "Skip dinner, wake up thinner. #Eatingdisorder. " • "Just think how all of this pain will be worth it in the end #thinspo #thin #skinny #eatingdisorder #anorexic #anorexia #mia"
Anti-ana (expressing concerns regarding pro-ana content)	"As the mom of a teen who struggles with an eating disorder, seeing WW target teens makes me worry. It is not correct to tell kids that a number on the scale is what gives them their value. #livelife #smile #behappy #BodyPositivity #WakeUpWeightWatchers"
Help offer (promoting help- and treatment-seeking behaviors)	• "It's #EatingDisorderAwarenessWeek If you have concerns about your loved ones or if you feel keen to talk to a professional counselor about your difficulties, call our confidential helpline at XXX #EDAW18 #NotAlone #HeretoHelp" • "Are you concerned that your child may have an #eatingdisorder? Help and support is available from @XXX"
Awareness (aim to raise awareness, combat misconceptions, or educate)	• "Sign this petition against @WeightWatchers' insidious campaign to recruit teens to their dangerous business #WakeUpWeightWatchers" • "Eating disorder experts, dietitians, physicians, and advocates around the world are standing together against @weightwatchers marketing dangerous "free" summer memberships to vulnerable young people. A season of fat-shaming, a lifetime of weight cycling." • "Spotting the signs: "If you suspect an eating disorder you should talk to them directly, and express your concerns –because the sooner you get help the better" #EatingDisorderAwarenessWeek #SpeakYourMind"

[a]Usernames and personal names have been removed.

Measuring Influence on Twitter: Retweets, Favorites, and the Reach and Impact of Hashtags

We analyzed the number of retweets and favorites generated by each tweet as an indicator of user interest in a given topic. We also measured the potential reach and impact of all analyzed hashtags to best assess tendencies in the dissemination of tweets. For the purpose of this study, impact is defined as a numerical value representing the potential views a tweet may receive, while reach is defined as a numerical value measuring the potential audience of the hashtag. These metrics are analyzed by Tweet Binder, a Twitter tool that automatically tracks hashtags. A more detailed explanation is shown in Multimedia Appendix 1.

User Categorization

The number of tweets that each Twitter user (contributor) posted with a certain hashtag was automatically quantified by Tweet Binder. We defined the contributors as very active if posting at least 6 tweets with a certain hashtag. The number of followers of the contributors was automatically quantified by Tweet Binder at the time of the study. We classified those with at least 5000 followers at the time of our analysis as very influential

XSL•FO
RenderX

contributors. Furthermore, we collected the profile description of the ten most influential and active users that participated in these conversations. The user profile description was used to determine whether a tweet was posted by health providers (such as physicians or nurses), health organizations (hospitals, journals, medical societies, or research institution), or by non–health-related users.

Ethical Considerations

This study was approved by the Research Ethics Committee of University of Navarra. This study did not directly involve human subjects and did not include interventions, but instead used publicly available tweets. Nevertheless, we have taken care to not reveal any username and to avoid citing the tweets that could reveal usernames.

Statistical Analysis

A descriptive study of the sample was performed, describing the variables by their absolute and relative frequencies. The percentages found were compared using the chi-square test. The mean and median numbers of retweets or favorites per original tweet about the different hashtags studied were compared using the Kruskal-Wallis test. All statistical analyses were performed using SPSS (Version 20, IBM Corp).

Results

Public Awareness Initiatives Succeeded in Spreading Their Tweets

Our search tool found that the total number of tweets generated with the hashtags associated with the Eating Disorder Awareness Week and Wake Up Weight Watchers awareness campaigns and the control hashtag #eatingdisorder were 7468**.** Of these, 455 were excluded according to the criteria of the study. The number of #wakeupweightwatchers tweets (3900) was double that of the Eating Disorder Awareness Week tweets (2056 tweets included the hashtags #eatingdisorderawarenessweek, #eatingdisorderawareness, or #EDAW). There were 1057 tweets using the control hashtag #eatingdisorder.

As shown in Table 2, we found that the potential impact and reach of the tweets was higher among Eating Disorder Awareness Week hashtags (54,143,739 and 9,547,326, respectively) and #wakeupweightwatchers (28,371,411 and 9,332,345, respectively) than observed with the #eatingdisorder hashtag (7,384,797 and 4,175,017, respectively).

Table 2. Potential impact, potential reach, number of contributors, percentage of very active users, and percentage of very influential users in the 3 groups.

Parameters	Eating disorder	Eating Disorder Awareness Week	Wake Up Weight Watchers
Potential impact, n	7,384,797	54,143,739	28,371,411
Potential reach, n	4,175,017	9,547,326	9,332,345
Contributors, n	1243	2531	4609
Very active contributors, n (%)	70 (5.6)	59 (2.3)	383 (8.3)
Very influential contributors, n (%)	127 (10.2)	186 (7.4)	289 (6.3)

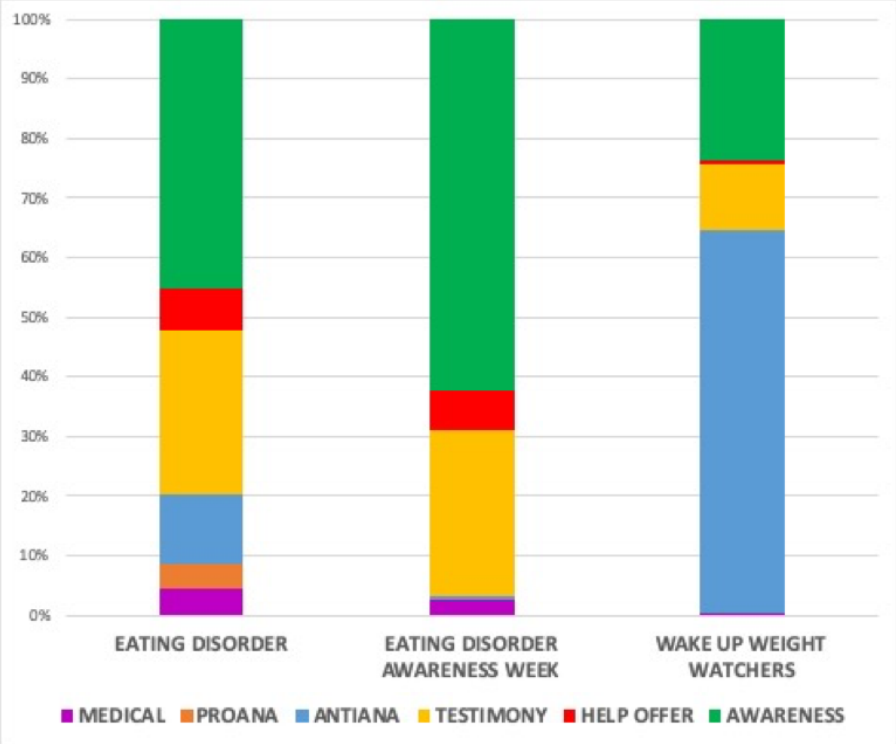
Eating Disorder Awareness Campaigns Fail to Promote Medical Content, Help Offers, Or Personal Testimonies But Can Increase Pro-Ana Criticism

We performed a content analysis of 1500 tweets, including 500 randomly selected tweets each from the three different groups of hashtags. First, we studied the medical contents of the tweets. Although the frequency was low, there were significant differences between the 3 groups (P<.001; Figure 1). The percentage of tweets with medical content was higher in those with the #eatingdisorder hashtag than in those with the awareness campaign hashtags. The lowest frequency was found in the #wakeupweightwatchers group.

We also analyzed the content of the 1500 tweets according to the other five established categories: personal testimony, help offer, awareness, pro-ana, and anti-ana. Different patterns of distribution among the five categories were found between the 3 groups of hashtags (P<.001). The percentage of tweets with personal testimony content was similar in #eatingdisorder tweets and tweets using Eating Disorder Awareness Week hashtags, while the percentage was lower in #wakeupweightwatchers tweets (Figure 1). The frequency of awareness tweets was higher in the tweets with Eating Disorder Awareness Week hashtags and lower in the tweets with the #wakeupweightwatchers hashtag (Figure 1). Help offer tweets were more prevalent in tweets with the #eatingdisorder hashtag and Eating Disorder Awareness Week hashtags. The frequency of anti-ana tweets was clearly higher in the #wakeupweightwatchers group and minimal in Eating Disorder Awareness Week tweets (Figure 1). Pro-ana content was absent in #wakeupweightwatchers tweets and the frequency was highest in #eatingdisorder tweets (Figure 1).

XSL•FO
RenderX

Figure 1. Different percentages of tweets in each category generated about the awareness campaigns #eatingdisorderawarenessweek (including #eatingdisorderawareness and #EDAW) and #wakeupweightwatchers, with #eatingdisorder as the control. Percentages were calculated with respect to the 500 tweets analyzed for each of the three categories of hashtags. Chi-square test; *P*<.001.



Pro-Ana and Medical Contents Did Not Attract Much Attention Among Twitter Users

We measured the number of retweets and favorites generated by each hashtag and category. Since these hashtags have different contribution numbers, we used ratios, dividing by the total number of tweets in each campaign. We found that the retweet-to-tweet and favorite-to-tweet ratio between the three hashtag groups were significantly different (*P*<.001; Table 3). As whole group, #wakeupweightwatchers was the hashtag with the highest retweet-to-tweet and favorite-to-tweet ratio. We found that the retweet-to-tweet and favorite-to-tweet ratio between the six categories were significantly different (*P*<.05; Table 4). Regarding the medical content, the highest ratios were found in #wakeupweightwatchers tweets, while they were low among the #eatingdisorder and Eating Disorder Awareness Week hashtags. Furthermore, tweets with personal testimony, help offer, and awareness content showed higher retweet-to-tweet and favorite-to-tweet ratio in #wakeupweightwatchers than those found in the other two groups. The tweets with anti-ana content generated higher ratios in #eatingdisorder tweets and lower ratios in Eating Disorder Awareness Week tweets. We found that the pro-ana tweets generated no or low numbers of retweets and likes in the 3 groups of tweets studied.

Table 3. Retweet-to-tweet and favorite-to-tweet ratio per hashtag analyzed.

Parameter and value	Eating disorder	Eating Disorder Awareness Week	Wake Up Weight Watchers	*P* value[a]
Retweet				
Mean (SD)	0.84 (7.21)	0.1 (0.79)	4.23 (21.51)	N/A[b]
Median (IQR)	0 (0-0)	0 (0-0)	1 (0-3)	<.001
Favorite				
Mean (SD)	1.76 (16.46)	0.2 (1.43)	9.52 (58.09)	N/A
Median (IQR)	0 (0-0)	0 (0-0)	2 (0-7)	<.001

[a]The Kruskal-Wallis test was used.

[b]N/A: not applicable.

XSL•FO
RenderX

Table 4. Retweet-to-tweet and favorite-to-tweet ratio per category in each hashtag analyzed.

Category and parameter	Eating disorder, median (IQR)	Eating Disorder Awareness Week, median (IQR)	Wake Up Weight Watchers, median (IQR)	*P* value[a]
Medical				
Retweet	0 (0-0)	0 (0-0)	4 (3)	.03
Favorite	0 (0-1)	0 (0-0)	4 (1-4)	.02
Pro-ana				
Retweet	0 (0-0)	0 (0-0)	—[b]	.88
Favorite	0 (0-1.75)	0 (0-0)	—	.72
Anti-ana				
Retweet	2 (0-7.5)	0 (0-0)	1 (0-3)	.002
Favorite	5 (1.5-9.5)	0 (0-0)	3 (1-7)	.001
Personal testimony				
Retweet	0 (0-0)	0 (0-0)	1 (0-2)	<.001
Favorite	0 (0-2)	0 (0-0)	3 (2-6.25)	<.001
Help offer				
Retweet	0 (0-1)	0 (0-0)	2 (1.25-2)	.002
Favorite	0 (0-1)	0 (0-1)	4 (0.5-6.75)	.047
Awareness				
Retweet	0 (0-1)	0 (0-0)	1 (0-3)	<.001
Favorite	0 (0-2)	0 (0-0)	3 (1-6.75)	<.001

[a]The Kruskal-Wallis test was used.

[b]Not available.

Eating Disorder Awareness Campaigns Increased the Participation of Twitter Users, But Not of Those Considered Very Active or Influential

We also analyzed the number and type of users that participated in conversations using any of the three hashtag groups (Table 2). The total number of contributors was higher in the #wakeupweightwatchers group (4609) than in the Eating Disorder Awareness Week (2531) and #eatingdisorder groups (1243). The frequency of very active users was higher in the #wakeupweightwatchers group (8.3%) than in the #eatingdisorder (5.63%) and Eating Disorder Awareness Week groups (2.33%). The percentage of very influential participants was highest in #eatingdisorder group (10.22%) than in the #wakeupweightwatchers (6.27%) and Eating Disorder Awareness Week groups (7.35%). The percentage of health providers or health organizations among the ten most active or influential users that participated in these conversations was higher in the #wakeupweightwatchers group (40%) than in the #eatingdisorder (30%) and Eating Disorder Awareness (20%) groups.

Finally, we analyzed the 5 hashtags more frequently associated with the 3 groups of hashtags studied (Table 5). We found that the hashtags most frequently associated with the 3 groups analyzed were different.

Table 5. Top 5 hashtags associated with the 3 groups of hashtags analyzed.

Hashtag group	The top 5 hashtags and the number of tweets per hashtag
Eating disorder	Anorexia (n=199), HAES (Health At Every Size; n=195), HelpfaithHeal (n=193), bodypositive (n=171), recovery (n=156)
Eating Disorder Awareness Week	GlobalGoodEmi (n=1201), Winsday (n=289), Wednesday (n=287), Giveaway (n=196), Flash (n=180)
Wake Up Weight Watchers	HAES (n=341), simplyNYC (m=283), MNYFashionweek (n=281), sheopenedthedoor (n=281), weightwatchers (n=198)

Discussion

Principal Findings

In this study, we have demonstrated that the eating disorder awareness campaign initiatives Eating Disorder Awareness Week and Wake Up Weight Watchers generated a greater number of tweets, with more far-reaching dissemination than the control eating disorder hashtag. Nonetheless, they failed to increase the frequency of tweets containing medical content, personal testimonies of recovery, or offers for treatment.

XSL•FO
RenderX

Moreover, those Twitter users classified as "very influential" did not command a strong presence in driving awareness of the aforementioned initiatives; on the whole, medical institutions also did not have a strong presence. Lastly, hashtags commonly used by pro-ana social networks are not used to any great extent by the awareness campaigns studied, thus making the campaigns less likely to reach at-risk individuals.

Our findings show that awareness initiatives generated a greater number of eating disorder–related tweets than those spontaneously posted. This finding suggests that awareness initiatives trigger conversations and social responses to eating disorders that would not have occurred otherwise. The total number of tweets related to the 3 groups of hashtags we analyzed over a period of 12 days was higher than the numbers of tweets associated with other mental health conditions [22]. However, the interest in eating disorders shown by the American media overall can be classified as low [23]. Notwithstanding, our results show that the Eating Disorder Awareness Week and Wake Up Weight Watchers campaigns had a positive effect on Twitter users. Furthermore, the potential impact and reach of the tweets related to the two campaigns were clearly higher than those achieved by our eating disorder control group.

According to the analysis of the generated tweets, both awareness initiatives have limitations. They failed to increase the percentage of tweets related to medical content, in contrast to previous evidence showing that tweets with medical content generated special interest among Twitter users [24]. A study performed within a population of young people highlighted the use of scientific data as being a key indicator of credibility in online resources [20]. The scarcity of medical content found by our study has also been reported in other awareness initiatives focused on other medical diseases, such as several prevalent types of cancer [25]. For example, Breast Cancer Awareness Month tweets most commonly contained content related to the practices of wearing pink, promoting walks/runs, and promoting fundraisers, as opposed to direct medical advice and preventive measures [25]. Furthermore, the frequency of tweets generated with help content was not higher in tweets related to awareness campaign hashtags. These results are concerning due to the existence of evidence that interventions that included help resources increased participants' knowledge of treatment resources [7].

Interestingly, one-third of the tweets analyzed have personal testimonies. This finding supports the value of Twitter as a means of communicating personal content related to eating disorder symptoms [26,27]. One-third of US teenagers use Twitter and this share has remained consistent over the past several years [28]. In general, Twitter users tend to be younger, have more education, and have higher incomes than the overall US population, but these differences are minimal among teenagers [29]. In addition, similar results have been reported in other countries. Twitter is a predominantly public-facing platform. It gives anonymity to user testimonies, encourages communication by people with real or perceived personal or social restrictions, and prevents the potential stigmatization of those with mental health diseases [30]. In addition, it provides an insight into the potential value of Twitter as a tool to reach these patients and encourage them to get in touch with health care providers. Unfortunately, neither campaign increased the percentage of tweets that included testimonies. It has been shown that young people value resources that allow them to access the personal stories of peers with lived experiences [31]. As expected, Eating Disorder Awareness Week had the highest proportion of awareness tweets. There is no evidence that this type of content increases the odds of seeking help or changing behaviors.

The awareness initiatives increased the frequency of tweets with anti-ana content. Additionally, pro-ana contributors were absent. The analysis of the eating disorder tweets showed a very low percentage of pro-ana content. Several reasons that are not mutually exclusive may explain this finding. First, Twitter has a policy of excluding pro-ana content [32]. Second, the pro-ana community may communicate through other web-based means, such as pro-ana websites [33]. Third, pro-ana social media posts are often tagged using specific hashtags, such as #thinspo, #thinspiration, #Thin15, #edprobs, and #proana. Pro-ana communities on social media have been described as tighter and more cohesive than others, as well as more secretive and exclusive [34]. Our findings show that these pro-ana hashtags have not been included in tweets related to the awareness campaigns or tweets tagged with #eatingdisorder. These findings suggest that Twitter awareness campaigns should include hashtags commonly used by these pro-ana communities to increase the likelihood of reaching people in need of interventions for disordered eating behaviors [35,36].

Retweet frequency is a parameter that indicates user interest in the topic of each tweet [37]. We also measured the number of favorites generated as an additional parameter of user interest, as previously shown [38]. Our results show that the Wake Up Weight Watchers hashtag obtained higher retweet and favorites frequencies than Eating Disorder Awareness Week hashtags and the control #eatingdisorder hashtag. Other Twitter metrics of diffusion (such as the potential impact and reach of tweets) were higher in Eating Disorder Awareness Week and Wake Up Weight Watchers tweets than in #eatingdisorder tweets. Furthermore, the awareness campaigns were associated with an increased number of contributors and tweets generated; there was an increased frequency of posting by very active users. However, these campaigns failed to promote the activity of very influential participants that were identified by the number of followers. It has been shown that verified accounts influence Twitter users [39].

Strengths and Limitations

This study has some limitations. First, Twitter may not be reflective of the general population. Second, researchers cannot directly measure eating disorder behavior or clinical outcomes from tweets. Third, the codebook design and text analysis imply a degree of subjectivity. However, this methodology is consistent with previous medical research studies using Twitter and could be applied to different topics by different authors [40-42]. Furthermore, to address this issue, the study comprised a series of steps including the initial review, design of the codebook, and agreement between the coders. Although computerized machine learning methods have been tested to automatically identify and classify topics in medical research

XSL•FO
RenderX

in social media [43], we used an analysis strategy based on the raters' clinical expertise in mental health and endocrinology, which constitutes a qualitative advantage compared to automated strategies. On the other hand, it can be argued that #eatingdisorder content may be influenced by awareness campaigns. However, to our knowledge, there is no evidence of content seasonality. The time frame was selected according to awareness campaign dates. To collect data, we used the Twitter Firehose data stream, which is managed by Gnip and allows access to 100% of all public tweets that match a set of search criteria, which enables the creation of a very representative database. For the content analysis, we randomly collected a number of tweets that enabled us to make inferences from the text to summarize the content [44]. Finally, we did not determine whether a campaign prepared in advance (eg, Eating Disorder Awareness Week) affected Twitter activity differently (eg, user engagement) than a campaign that was more spontaneous (eg, Wake Up Weight Watchers).

Conclusions

The initiatives Eating Disorder Awareness Week and Wake Up Weight Watchers increased the volume of eating disorder–related tweets, but the content of most tweets focused on raising awareness rather than promoting treatment-seeking behaviors, which would likely have a more significant outcome. Our findings support the need to reconsider the message and communication strategy employed by eating disorder awareness campaigns. The involvement of health institutions appears to be desirable because they are an important indicator of credibility and generated more tweets focused on medical information. Furthermore, the testimony of patients who have recovered from eating disorders should be promoted in these campaigns. Awareness campaigns may also wish to consider a strategy in which hashtags targeted toward those with eating disorders are included.

Acknowledgments

This work was partially supported by grants from the Fondo de Investigación de la Seguridad Social, Instituto de Salud Carlos III (PI18/01726), Spain, and the Programa de Actividades de I+D de la Comunidad de Madrid en Biomedicina (B2017/BMD-3804), Madrid, Spain.

MAAM was affiliated with Clinica Universidad de Navarra at the time of the study and is currently affiliated with Hospital Universitario Infanta Leonor.

Authors' Contributions

IV and MAAM participated equally as principal contributors in the research design and content analysis, as well as manuscript writing, review, and submission. MLV participated in the content analysis and review. AAdB conducted and reported statistical analysis. MAM and FO supervised all stages of the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Potential reach and potential impact.
[DOCX File , 14 KB - jmir_v22i7e17626_app1.docx]

References

1. Treasure J, Claudino AM, Zucker N. Eating disorders. Lancet 2010 Mar 13;375(9714):583-593. [doi: 10.1016/S0140-6736(09)61748-7] [Medline: 19931176]
2. Keski-Rahkonen A, Hoek H, Susser E, Linna M, Sihvola E, Raevuori A. Epidemiology and course of anorexia nervosa in the community. Am J Psychiatry Internet 2007:1259-1265 [FREE Full text] [doi: 10.1037/e721392007-004]
3. Udo T, Grilo CM. Prevalence and Correlates of DSM-5-Defined Eating Disorders in a Nationally Representative Sample of U.S. Adults. Biol Psychiatry 2018 Sep 01;84(5):345-354 [FREE Full text] [doi: 10.1016/j.biopsych.2018.03.014] [Medline: 29859631]
4. Schaefer LM, Burke NL, Anderson LM, Thompson JK, Heinberg LJ, Bardone-Cone AM, et al. Comparing internalization of appearance ideals and appearance-related pressures among women from the United States, Italy, England, and Australia. Eat Weight Disord 2019 Oct;24(5):947-951. [doi: 10.1007/s40519-018-0544-8] [Medline: 30019258]
5. Fitzsimmons-Craft EE, Balantekin KN, Eichen DM, Graham AK, Monterubio GE, Sadeh-Sharvit S, et al. Screening and offering online programs for eating disorders: Reach, pathology, and differences across eating disorder status groups at 28 U.S. universities. Int J Eat Disord 2019 Oct;52(10):1125-1136. [doi: 10.1002/eat.23134] [Medline: 31268183]
6. Sampogna G, Bakolis I, Evans-Lacko S, Robinson E, Thornicroft G, Henderson C. The impact of social marketing campaigns on reducing mental health stigma: Results from the 2009-2014 Time to Change programme. Eur Psychiatry 2017 Feb;40:116-122 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.08.008] [Medline: 27997875]

XSL•FO
RenderX

7. Tillman KS, Arbaugh T, Balaban MS. Campus programming for National Eating Disorders Awareness Week: an investigation of stigma, help-seeking, and resource knowledge. Eat Behav 2012 Aug;13(3):281-284. [doi: 10.1016/j.eatbeh.2012.03.010] [Medline: 22664413]
8. Golden NH, Schneider M, Wood C, Committee on Nutrition, Committee on Adolescence, Section on Obesity. Preventing Obesity and Eating Disorders in Adolescents. Pediatrics 2016 Sep;138(3):1-12 [FREE Full text] [doi: 10.1542/peds.2016-1649] [Medline: 27550979]
9. Raffoul A, Leatherdale ST, Kirkpatrick SI. Dieting predicts engagement in multiple risky behaviours among adolescent Canadian girls: a longitudinal analysis. Can J Public Health 2018 Feb;109(1):61-69 [FREE Full text] [doi: 10.17269/s41997-018-0025-x] [Medline: 29981072]
10. Loth KA, MacLehose R, Bucchianeri M, Crow S, Neumark-Sztainer D. Predictors of dieting and disordered eating behaviors from adolescence to young adulthood. J Adolesc Health 2014 Nov;55(5):705-712 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.04.016] [Medline: 24925491]
11. Purtle J, Roman L. Health Awareness Days: Sufficient Evidence to Support the Craze? Am J Public Health 2015 Jun;105(6):1061-1065 [FREE Full text] [doi: 10.2105/ajph.2015.302621]
12. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
13. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):51-58. [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
14. Lyles CR, López A, Pasick R, Sarkar U. "5 mins of uncomfyness is better than dealing with cancer 4 a lifetime": an exploratory qualitative analysis of cervical and breast cancer screening dialogue on Twitter. J Cancer Educ 2013 Mar;28(1):127-133 [FREE Full text] [doi: 10.1007/s13187-012-0432-2] [Medline: 23132231]
15. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
16. Jessup D, Glover Iv M, Daye D, Banzi L, Jones P, Choy G, et al. Implementation of Digital Awareness Strategies to Engage Patients and Providers in a Lung Cancer Screening Program: Retrospective Study. J Med Internet Res 2018 Feb 15;20(2):e52 [FREE Full text] [doi: 10.2196/jmir.8932] [Medline: 29449199]
17. Booth RG, Allen BN, Bray Jenkyn KM, Li L, Shariff SZ. Youth Mental Health Services Utilization Rates After a Large-Scale Social Media Campaign: Population-Based Interrupted Time-Series Analysis. JMIR Ment Health 2018 Apr 06;5(2):e27 [FREE Full text] [doi: 10.2196/mental.8808] [Medline: 29625954]
18. McLean SA, Caldwell B, Roberton M. Reach Out and Recover: Intentions to seek treatment in individuals using online support for eating disorders. Int J Eat Disord 2019 Oct;52(10):1137-1149. [doi: 10.1002/eat.23133] [Medline: 31298791]
19. Ali K, Farrer L, Fassnacht D, Gulliver A, Bauer S, Griffiths K. Perceived barriers and facilitators towards help-seeking for eating disorders: A systematic review. Int J Eat Disord 2017 Jan;50(1):9-21 [FREE Full text] [doi: 10.1002/eat.22598] [Medline: 27526643]
20. Pretorius C, Chambers D, Cowan B, Coyle D. Young People Seeking Help Online for Mental Health: Cross-Sectional Survey Study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
21. Colditz J, Chu K, Emery S, Larkin C, James A, Welling J, et al. Toward Real-Time Infoveillance of Twitter Health Messages. Am J Public Health 2018 Aug;108(8):1009-1014 [FREE Full text] [doi: 10.2105/ajph.2018.304497]
22. Alvarez-Mon M, Llavero-Valero M, Sánchez-Bayona R, Pereira-Sanchez V, Vallejo-Valdivielso M, Monserrat J, et al. Areas of Interest and Stigmatic Attitudes of the General Public in Five Relevant Medical Conditions: Thematic and Quantitative Analysis Using Twitter. J Med Internet Res 2019 May 28;21(5):e14110 [FREE Full text] [doi: 10.2196/14110] [Medline: 31140438]
23. Alvarez-Mon M, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing Interest of Mass Communication Media and the General Public in the Distribution of Tweets About Mental Disorders: Observational Study. J Med Internet Res 2018 May 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]
24. Attai D, Cowher M, Al-Hamadani M, Schoger J, Staley A, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
25. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
26. Arseniev-Koehler A, Lee H, McCormick T, Moreno M. #Proana: Pro-Eating Disorder Socialization on Twitter. J Adolesc Health 2016 Jun;58(6):659-664 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
27. Cavazos-Rehg P, Krauss M, Costello S, Kaiser N, Cahn E, Fitzsimmons-Craft E, et al. "I just want to be skinny.": A content analysis of tweets expressing eating disorder symptoms. PLoS One 2019;14(1):e0207506 [FREE Full text] [doi: 10.1371/journal.pone.0207506] [Medline: 30650072]

XSL•FO
RenderX

28. Anderson M, Jiang J. Teens, social media & technology. Pew Res Cent Internet Am Life Proj. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-06-30]
29. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2020-06-30]
30. Fitzsimmons-Craft EE, Balantekin KN, Eichen DM, Graham AK, Monterubio GE, Sadeh-Sharvit S, et al. Screening and offering online programs for eating disorders: Reach, pathology, and differences across eating disorder status groups at 28 U.S. universities. Int J Eat Disord 2019 Oct;52(10):1125-1136. [doi: 10.1002/eat.23134] [Medline: 31268183]
31. Dawson L, Mullan B, Touyz S, Rhodes P. Are recovery stories helpful for women with eating disorders? A pilot study and commentary on future research. J Eat Disord 2018;6:21 [FREE Full text] [doi: 10.1186/s40337-018-0206-2] [Medline: 30128153]
32. Bert F, Gualano M, Camussi E, Siliquini R. Risks and Threats of Social Media Websites: Twitter and the Proana Movement. Cyberpsychol Behav Soc Netw 2016 Apr;19(4):233-238 [FREE Full text] [doi: 10.1089/cyber.2015.0553] [Medline: 26991868]
33. Oksanen A, Garcia D, Räsänen P. Proanorexia Communities on Social Media. Pediatrics 2016 Jan;137(1):1-5 [FREE Full text] [doi: 10.1542/peds.2015-3372] [Medline: 26676052]
34. Tiggemann M, Churches O, Mitchell L, Brown Z. Tweeting weight loss: A comparison of #thinspiration and #fitspiration communities on Twitter. Body Image 2018 Jun;25:133-138. [doi: 10.1016/j.bodyim.2018.03.002] [Medline: 29567619]
35. Harris J, Duncan A, Men V, Shevick N, Krauss M, Cavazos-Rehg P. Messengers and Messages for Tweets That Used #thinspo and #fitspo Hashtags in 2016. Prev Chronic Dis 2018 Jan 04;15:E01 [FREE Full text] [doi: 10.5888/pcd15.170309] [Medline: 29300696]
36. Branley DB, Covey J. Pro-ana versus Pro-recovery: A Content Analytic Comparison of Social Media Users' Communication about Eating Disorders on Twitter and Tumblr. Front Psychol 2017;8:1356 [FREE Full text] [doi: 10.3389/fpsyg.2017.01356] [Medline: 28848472]
37. Tsugawa S, Kito K. Retweets as a Predictor of Relationships among Users on Social Media. PLoS One 2017;12(1):e0170279 [FREE Full text] [doi: 10.1371/journal.pone.0170279] [Medline: 28107489]
38. Pereira-Sanchez V, Alvarez-Mon M, Asunsolo Del Barco A, Alvarez-Mon M, Teo A. Exploring the Extent of the Hikikomori Phenomenon on Twitter: Mixed Methods Study of Western Language Tweets. J Med Internet Res 2019 May 29;21(5):e14167 [FREE Full text] [doi: 10.2196/14167] [Medline: 31144665]
39. Schwenk ES, Jaremko KM, Park BH, Stiegler MA, Gamble JG, Chu LF, et al. I Tweet, Therefore I Learn: An Analysis of Twitter Use Across Anesthesiology Conferences. Anesth Analg 2019 Jan 28:333-340. [doi: 10.1213/ANE.0000000000004036] [Medline: 31124801]
40. Ghaznavi J, Taylor L. Bones, body parts, and sex appeal: An analysis of #thinspiration images on popular social media. Body Image 2015 Jun;14:54-61 [FREE Full text] [doi: 10.1016/j.bodyim.2015.03.006] [Medline: 25880783]
41. Tufts C, Polsky D, Volpp K, Groeneveld P, Ungar L, Merchant R, et al. Characterizing Tweet Volume and Content About Common Health Conditions Across Pennsylvania: Retrospective Analysis. JMIR Public Health Surveill 2018 Dec 06;4(4):e10834 [FREE Full text] [doi: 10.2196/10834] [Medline: 30522989]
42. Albalawi Y, Nikolov N, Buckley J. Trustworthy Health-Related Tweets on Social Media in Saudi Arabia: Tweet Metadata Analysis. J Med Internet Res 2019 Oct 08;21(10):e14731 [FREE Full text] [doi: 10.2196/14731] [Medline: 31596242]
43. Saha K, Torous J, Ernala S, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med 2019 Nov 25;9(6):1197-1207 [FREE Full text] [doi: 10.1093/tbm/ibz028] [Medline: 30834942]
44. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for Coding Tobacco-Related Twitter Data: A Systematic Review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]

Abbreviations

HAES: Health At Every Size

Edited by G Eysenbach; submitted 29.12.19; peer-reviewed by R Rodriguez-Jimenez, B Hoffman, C Ralph-Nearman, C Pretorius; comments to author 21.01.20; revised version received 06.04.20; accepted 13.06.20; published 14.07.20.

Please cite as:
Viguria I, Alvarez-Mon MA, Llavero-Valero M, Asunsolo del Barco A, Ortuño F, Alvarez-Mon M
Eating Disorder Awareness Campaigns: Thematic and Quantitative Analysis Using Twitter
J Med Internet Res 2020;22(7):e17626
URL: http://www.jmir.org/2020/7/e17626/
doi:10.2196/17626
PMID:32673225

XSL•FO
RenderX

©Iranzu Viguria, Miguel Angel Alvarez-Mon, Maria Llavero-Valero, Angel Asunsolo del Barco, Felipe Ortuño, Melchor Alvarez-Mon. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Conversations and Misconceptions About Chemotherapy in Arabic Tweets: Content Analysis

Abdulrahman Alghamdi[1,2], MBBS; Khalid Abumelha[1,2], MBBS; Jawad Allarakia[1,2], MBBS; Ahmed Al-Shehri[1,2,3], MBBS, ABIM, FRCPC

[1]College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia

[2]King Abdullah International Medical Research Center, Jeddah, Saudi Arabia

[3]Department of Medical Oncology, Princess Noorah Oncology Center, Ministry of the National Guard - Health Affairs, Jeddah, Saudi Arabia

Corresponding Author:
Abdulrahman Alghamdi, MBBS
College of Medicine
King Saud Bin Abdulaziz University for Health Sciences
King Abdulaziz Medical City, Makkah-Jeddah Hwy
Jeddah, 22384
Saudi Arabia
Phone: 966 012 224 5000
Email: abdurrahman.essa@gmail.com

Abstract

Background: Although chemotherapy was first introduced for the treatment of cancer more than 60 years ago, the public understanding and acceptance of chemotherapy is still debatable. To the best of our knowledge, no study has assessed the conversations and misconceptions about chemotherapy as a treatment for cancer on social media platforms among the Arabic-speaking populations.

Objective: The aim of this study was to assess the types of conversations and misconceptions that were shared on Twitter regarding chemotherapy as a treatment for cancer among the Arabic-speaking populations.

Methods: All Arabic tweets containing any of the representative set of keywords related to chemotherapy and written between May 1, 2017 and October 31, 2017 were retrieved. A manual content analysis was performed to identify the categories of the users, general themes of the tweets, and the common misconceptions about chemotherapy. A chi-square test for independence with adjusted residuals was used to assess the significant associations between the categories of the users and the themes of the tweets.

Results: A total of 402,157 tweets were retrieved, of which, we excluded 309,602 retweets and 62,651 irrelevant tweets. Therefore, 29,904 tweets were included in the final analysis. The majority of the tweets were posted by general users (25,774/29,904, 86.2%), followed by the relatives and friends of patients with cancer (1913/29,904, 6.4%). The tweets were classified into 9 themes; prayers and wishes for the well-being of patients undergoing chemotherapy was the most common theme (20,288/29,904, 67.8%), followed by misconceptions about chemotherapy (2084/29,904, 7.0%). There was a highly significant association between the category of the users and the themes of the tweets (χ^2_{40}= 16904.4, P<.001).

Conclusions: Our findings support those of the previous infodemiology studies that Twitter is a valuable social media platform for assessing public conversations, discussions, and misconceptions about various health-related topics. The most prevalent theme of the tweets in our sample population was supportive messages for the patients undergoing chemotherapy, thereby suggesting that Twitter could play a role as a support mechanism for such patients. The second most prevalent theme of the tweets in our study was the various misconceptions about chemotherapy. The findings of our exploratory analysis can help physicians and health care organizations tailor educational efforts in the future to address different misconceptions about chemotherapy, thereby leading to increased public acceptance of chemotherapy as a suitable mode of treatment for cancer.

(J Med Internet Res 2020;22(7):e13979) doi:10.2196/13979

KEYWORDS

internet; chemotherapy; cancer; Twitter; social media; Arab world; misconceptions; infodemiology; infoveillance

XSL•FO
RenderX

Introduction

Although chemotherapy linguistically refers to any type of treatment with chemicals, it is now specifically used to describe cytotoxic anticancer medications, which are used to treat hematological and solid tumors [1,2]. Chemotherapeutic agents can be used alone or in combination with surgery or radiotherapy to achieve the goal of the treatment, which varies among patients based on many factors. The goal of the treatment may be to cure the cancer completely, control the cancer, stop it from spreading, or to use these agents as a palliative means to ease the symptoms and improve the patients' quality of life [3]. However, the natural mechanism of action of these chemotherapeutic agents results in relatively a wide spectrum of side effects, which might be unpleasant to the patients.

Despite continuous improvements in chemotherapy for increasing its effectiveness in treating cancer, the public understanding and acceptance of chemotherapy is still debatable. Many misconceptions can occur, thereby causing compliance issues and other problems related to the acceptance of chemotherapy as a treatment modality [4]. Most of the previous studies that assessed the perceptions and experiences of chemotherapy were performed with selected groups of patients who were either surveyed or interviewed [5-8]. However, large-scale studies attempting to understand the public views and conversations on chemotherapy are lacking. Retrieving and analyzing the contents of various social media platforms on this matter serves as an evolving method to gather information about the perceptions, misconceptions, and experiences with chemotherapy as a treatment for cancer. One of the most popular social media platforms that has been explored is Twitter, which is a large microblog wherein users can write up to 280-character messages called tweets and share them with their followers and the public. There are almost 313 million monthly active users who write around 500 million tweets daily [9,10]. In March 2017, it was estimated that the number of the monthly active Twitter users in the Arab world was around 11.1 million [11]. Moreover, the number of their tweets per month in March 2016 was estimated to be 849.1 million tweets, and 72% of them were written in Arabic [11].

Twitter can be useful for exploring public opinions on health-related matters and it may play an important role in clinical settings by providing an avenue for patients to increase their knowledge regarding their diseases and by leading to a positive impact on their medical outcomes [12-14]. For example, a study showed that the mental health of the patients was affected by social media content, as positive messages were shown to reduce anxiety in patients with breast cancer who were using Twitter [12]. Moreover, social media contents show the public impression and knowledge in addition to their emotions and stress regarding diseases and their associated factors in real time [15,16].

The aim of this study was to report a content analysis of the Arabic tweets regarding chemotherapy to determine the categories of the contents and to explore the possible cases of misconceptions about chemotherapy in the Arab world. This study addresses the common areas of misconceptions on cytotoxic chemotherapies in Arabic Twitter conversations, which may help to direct the future educational efforts to address these areas of misconception.

Methods

Data Collection

This observational study was conducted on Arabic tweets about chemotherapy over a 6-month period, and the main themes to which these tweets belonged were determined. A search was conducted on every publicly available tweet in Arabic containing any of the keywords mentioned in the supplementary file (Multimedia Appendix 1). Keywords were chosen on a linguistic basis; therefore, they represented all the possible spellings and abbreviations of the word "chemotherapy" in Arabic. The Twitter Archiver add-on to Google Sheets was used to collect the data [17]. This tool searches for new public tweets every 15 minutes that include any prespecified keywords and allows users to download them on a Microsoft Excel spreadsheet. Further, this tool collects the related data of every user whose tweet has been collected such as the number of people who follow those users, whom they follow, and a brief biography that users may share about themselves on the profile page. All Arabic tweets from May 1, 2017 to October 31, 2017 that contained any of the predetermined keyword(s) were retrieved. Among these, we excluded all retweets because they might deviate from the thematic analysis as they do not necessarily imply endorsement and we excluded irrelevant tweets containing the keyword but that were used in a context other than chemotherapy as a medical intervention (eg, chemical weapon attacks in Syria).

Data Analysis

To help expand the knowledge on the tweets circulated on chemotherapy, the main questions asked to generate the coding scheme were if there were misconceptions or not about chemotherapeutic agents as cancer treatments and what other patterned meanings appeared in the data set. After a thorough study of a sample of 150 tweets, the primary author generated a preliminary list of the themes of the tweets and the categories of the users. This list was used by the rest of the authors to code the same 150-tweet sample for the themes and the categories of the users, which resulted in the modification and addition of a few themes. Thereafter, another 150 tweets were used to finally test the comprehensiveness of the themes and the categories of the users along with assessing the inter-rater reliability of the coders of the main data set (JA and KA) by using the Cohen kappa coefficient. The Cohen kappa coefficients were estimated as 0.90 and 0.92 for themes and categories of users, respectively. Table 1 describes the themes of the tweets and the illustrative quotes translated from Arabic. Table 2 describes the different categories of the users who were identified on the basis of the information provided in their biographies mentioned in their Twitter profile page or on the basis of their tweets.

Using the coding scheme, 2 of the investigators (JA and KA) independently assigned each tweet and the user who shared it to the proper mutually exclusive theme and user category. Cases of ambiguity were resolved by consensus among the authors.

XSL•FO
RenderX

Tweets under the misconception theme were further analyzed to establish the most common misconceptions among the studied groups of users.

Table 1. Descriptions of the different themes of the tweets on chemotherapy and the illustrative tweets.

Theme of the tweet	Description	Examples of the translated tweets
Advice and information	Disseminating true information and advice about chemotherapy	*…We use chemotherapy because its efficacy is proven—that is why it is called as evidence-based medicine.*
Experience	Sharing an experience of chemotherapy either by the patients themselves or by the people who surround them	*…My brother is on chemotherapy—the first session was on his birthday. It was like a birthday present—a painful and sad present.*
Misconception	Sharing a false concept or a false idea about chemotherapy	*…a long time ago I heard that a cure for cancer was discovered, but the person who discovered it was killed because companies wanted to benefit from the profits that are generated from selling chemotherapeutic drugs.*
Prayers and wishes	Saying a prayer or a wish for recovery for people receiving chemotherapy	*…They have been worn out by cancer, by chemotherapy…oh Allah heal our patients and all Muslim patients.*
Seeking medical information/advice	Asking for medical information about chemotherapy	*…If a patient has started chemotherapy for treating cancer, does he lose only the scalp hair or does he lose the hair all over his body?*
Seeking medical/financial help	Asking for medical or financial help for patients receiving chemotherapy	*…My mother is supposed to receive chemotherapy this week but we cannot afford it. All radiographic images and reports are attached. We need your help.*
Offering medical intervention/financial help	Offering medical or financial help for patients needing or receiving chemotherapy.	*…If anyone is in need of paclitaxel, which is a chemotherapeutic drug that is well over 3000 Egyptian pounds (US $187), please do not hesitate to contact me.*
Analogy	Using chemotherapy in an analogic way to share an idea or a concept.	*…Whatever concerns you now is nothing compared to those who are waiting for chemotherapy or hemodialysis tomorrow.*
Miscellaneous	Tweets that did not fit any of the other categories.	*…I hope I can discover a cancer treatment better than chemotherapy.*

Table 2. Description of the categories of the different Twitter users who tweeted on chemotherapy.

Category of the Twitter user accounts	Description of the Twitter user accounts
Patients with cancer/survivors	Accounts of patients with cancer receiving chemotherapy
Relatives/friends of patients with cancer	Accounts of a relative or a friend of a patient receiving chemotherapy
Cancer specialists	Accounts of people working in the oncology field
Health-related accounts	Accounts of an organization in the medical field or of a person working in a medical field other than oncology
Media-related accounts	Accounts of media platforms such as newspapers and news channels
General users	Accounts of anyone who does not belong to the other categories or who belongs to unidentified accounts.

Statistical Analysis

Statistical analysis was performed using the SPSS software (release version 23.0.0.0, IBM Corp). Descriptive statistical analyses were used to report the frequencies for each theme and the category of the users. The chi-square test for independence was used to assess any significant associations between the types of users and the themes of the tweets. To further identify the cells that contributed to the overall significant results, adjusted residuals (z-scores) were calculated for each cell along with the *P* values, as described by Beasley and Schumacker [18]. Taking into consideration the multiple comparisons in this analysis that might lead to type I error, a Bonferroni-adjusted $P<.001$ was utilized to indicate statistical significance.

Results

Frequency of Themes and Categories of Users

A total of 402,157 tweets were retrieved. Of these, we excluded 309,602 retweets and 62,651 irrelevant tweets (Figure 1). Therefore, 29,904 tweets were included in the final analysis. Table 3 provides the frequencies of all the themes and the user categories. Under the 9 previously identified themes, over two-thirds of the tweets (20,288/29,904; 67.8%) expressed “prayers and wishes” for the recovery of patients with cancer on chemotherapy. Approximately 7.0% (2084/29,904) of the tweets contained misconceptions regarding various aspects of chemotherapy. Providing “advice and information” was the third most frequent theme (1888/29,904, 6.3%), followed by tweets expressing “experience” (1847/29,904, 6.2%) with

XSL•FO
RenderX

chemotherapy for cancer management and, lastly, tweets that used 1 or more of the keywords for "analogy" (1556/29,904, 5.2%).

Figure 1. Overview of the process of collecting and filtering the data.

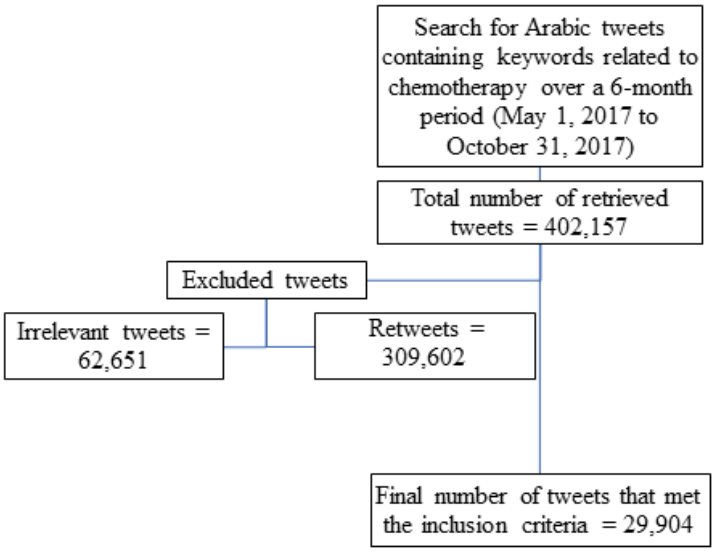


Table 3. Frequency analysis of the themes of the tweets on chemotherapy and the Twitter user categories (N=29,904).

Variable	n (%)
Themes	
Miscellaneous	932 (3.1)
Advice and information	1888 (6.3)
Experience	1847 (6.2)
Misconception	2084 (7.0)
Prayers and wishes	20,288 (67.8)
Seeking medical information/advice	487 (1.6)
Seeking medical/financial help	754 (2.5)
Offering medical/financial help	66 (0.2)
Analogy	1556 (5.2)
Category of users	
General users	25,774 (86.2)
Patients with cancer/survivors	854 (2.9)
Relatives/friends of patients with cancer	1913 (6.4)
Cancer specialists	222 (0.7)
Health-related accounts	459 (1.5)
Media-related accounts	680 (2.3)

In terms of the proportions of the categories of the users, general users represented the majority of the users in the analyzed data set (25,774/29,904, 86.2%). The relatives and friends of patients with cancer were the second most common contributors (1913/29,904, 6.4%), followed by patients with cancer and survivors (854/29,904, 2.9%). Cancer specialists, health-related accounts, and media-related accounts represented the remaining less than 5% (1361/29,904, 4.5%) of the tweets. Tweets under the theme "misconceptions" were further analyzed to capture the most common misconceptions, as described in detail in Table 4. Falsified and unrealistic side effects of chemotherapy were the most common misconceptions (1271/2084, 60.9%) in the studied data set. The other misconceptions were that chemotherapy causes cancer to spread (282/2084, 13.5%) and that chemotherapy has no therapeutic benefit (214/2084, 10.3%).

We compared the volumes of the tweets of misconception with the volumes of the tweets by cancer specialists and health-related accounts for the same period, and we found that the 2 noticeable spikes on June 1-16, 2017 and July 1-16, 2017 in the volume

XSL•FO
RenderX

of the tweets containing misconceptions seemed to drive both categories of users to tweet more (Table 5).

Table 4. Frequency of the main misconceptions about chemotherapy in the Arabic tweets (n=2084).

Misconceptions	n (%)
Falsified and unrealistic side effects about chemotherapy; the main shared tweet was that "a drop of chemotherapy spilled on a healthy man's skin would severely burn the skin."	1271 (60.9)
Chemotherapy causes cancer to spread.	282 (13.5)
Chemotherapy has no therapeutic effect on cancers.	214 (10.3)
Claims that there are natural products/preparations (eg, olive oil, *Nigella sativa*, ginger, almonds, camel milk or urine) or other practices (eg, blood cupping, carbohydrate restriction), which are more effective than chemotherapy for treating cancers	170 (8.2)
Chemotherapy is prescribed so that pharmaceutical companies and physicians can make huge profits.	67 (3.2)
Claims about some pharmaceutical products (eg, vitamin B17, antibiotics, vitamin C) being more effective than chemotherapy for treating cancers	47 (2.3)
Claims that there are few religious practices (eg, Roqya, Zamzam water intake, seclusion in mosque, fasting from dawn to sunset), which are more effective than chemotherapy for treating cancers	33 (1.6)

Table 5. Comparison of the number of tweets containing misconceptions by any user category and the number of tweets by cancer specialists and health-related accounts during the study period.

Time period (2017)	Tweets containing misconceptions (n)	Tweets by cancer specialists and health-related accounts (n)
May 1-16	1616	35
May 17-31	1992	33
June 1-16[a]	4365	108
June 17-30	2561	41
July 1-16[a]	3441	104
July 17-31	2854	84
August 1-16	2486	21
August 17-31	2758	59
September 1-16	2026	49
September 17-30	2670	51
October 1-16	1372	52
October 17-31	1761	44

[a]Noticeable spikes in the number of tweets containing misconceptions and tweets by cancer specialists and health-related accounts.

Comparison of the Themes by the Categories of the Users

The initial Pearson chi-square test of independence showed a highly significant association between the category of the users and the themes of tweets (χ^2_{40}=16904.4, P<.001). Further post-hoc analyses revealed the cells that contributed the overall significance (Table 6).

Table 6 shows that general users were more likely to support patients with cancer by tweeting prayers and wishes and less likely to tweet about experiences, seeking medical intervention, and advice and information themes. Patients with cancer and the relatives and friends of patients with cancer tweeted significantly more about experiences and seeking medical advice and intervention than all other themes. They were less likely to tweet misconceptions. Cancer specialists were more likely to tweet advice and information about chemotherapy and less likely to tweet either prayers and wishes or misconceptions. Health-related accounts showed a similar statistically significant tendency as the cancer specialists to tweet more about advice and information, but they had a greater tendency to tweet about offering medical interventions. Finally, media-related accounts were more likely to tweet about advice and information but also more likely to share misconceptions than all others. In contrast, they were less likely to share tweets on prayers and wishes or tweets containing analogy.

XSL•FO
RenderX

Table 6. Distribution of the themes by the source category.

Total number of tweets in each user category (n)	Theme-wise tweets								
	Miscellaneous, n (%)	Advice and information n (%)	Experience n (%)	Misconception n (%)	Prayers and wishes n (%)	Seeking medical advice n (%)	Seeking medical/financial help n (%)	Offering medical/financial help n (%)	Analogy n (%)
General users, n=25,774	760 (2.9)[a]	1120 (4.3)[a]	613 (2.4)[a]	1864 (7.2)[a]	19,294 (74.9)[a]	260 (1.0)[a]	291 (1.1)[a]	35 (0.1)[a]	1537 (5.9)[a]
Patients with cancer/survivors, n=854	26 (3.0)	24 (2.8)[a]	464 (54.3)[a]	7 (0.8)[a]	64 (7.5)[a]	3 (8.5)[a]	183 (21.4)[a]	9 (1.1)[a]	4 (0.5)[a]
Relatives/friends of patients with cancer, n=1913	34 (1.8)[a]	42 (2.2)[a]	615 (32.1)[a]	16 (0.8)[a]	821 (42.9)[a]	137 (7.2)[a]	245 (12.8)[a]	0 (0)	3 (0.2)[a]
Cancer specialists, n=222	4 (1.8)	206 (92.8)[a]	5 (2.2)	0 (0)	0 (0)	4 (1.8)	2 (0.9)	1 (0.4)	0 (0)
Health-related accounts, n=459	46 (10.0)[a]	215 (46.8)[a]	46 (10.0)[a]	21 (4.6)	85 (18.5)[a]	13 (2.8)	20 (4.4)	7 (1.5)[a]	6 (1.3)[a]
Media-related accounts, n=680	62 (9.1)[a]	281 (41.3)[a]	104 (15.3)[a]	176 (25.9)[a]	24 (3.5)[a]	0 (0)	13 (1.9)	14 (2.1)[a]	6 (0.9)[a]

[a]Statistically significant at Bonferroni-adjusted $P<.001$.

Discussion

The aim of this study was to assess the types of conversations and misconceptions regarding chemotherapy among Twitter users in Arabic-speaking populations. Given the large number of tweets retrieved, with their further spread via retweets, our study solidified the conclusions of previous studies that showed Twitter as a rich social media platform for obtaining health-related information [19-23]. Therefore, understanding these social media websites and having a glimpse of what is shared on these websites might serve as an important step for physicians to improve the health care delivered to their patients.

Twitter also represents a growing venue for researchers to analyze on what is shared about various topics of interest in health care either related to oncology [13-19] or related to topics on other health care fields (eg, antibiotics [24], vaccinations [25], smoking [26,27]). Twitter provides researchers and health care organizations with an opportunity to outreach a wide group of participants, overcome barriers pertaining to research resources, and track emerging health-related discussions and problems in real time [28,29].

In our manual content analysis, the most common theme of the tweets was “prayers and wishes.” This finding highlights a possible role of social media platforms as a support mechanism for patients with cancer receiving chemotherapy. Previous studies have also shown that social media platforms play a positive role in optimizing health care interventions [30], particularly in oncology settings [31].

Tweets containing misconceptions accounted for 7.0% (2084/29,904) of the total tweets in our sample population, making it the second most common theme. A further analysis of the misconceptions showed that most of these misconceptions were on the unrealistic side effects of chemotherapeutic agents, and people who do not have the knowledge or the experience of chemotherapy may tend to exaggerate their harmful effects, as reported in a previous study [5]. Therefore, health care organizations and professionals must tailor their awareness activities to target such common misconceptions in a given population.

Most of the tweets on media-related accounts delivered misconceptions about chemotherapy. We also found that when there was a spike in the misconceptions during the study period, the contributions of the cancer specialists and health-related accounts also increased proportionately to correct these misconceptions and to answer the questions of other users. This proportionate increase can limit the dissemination of misconceptions because cancer specialists and health-related accounts check the medical information that is tweeted and provide reliable information through tweets prior to sharing with the followers on Twitter. Further, heath care organizations can promote the true information shared by cancer specialists and health-related accounts by using the payable option offered by Twitter to show a certain tweet in search results and user feeds so that wide groups of the targeted population can be reached with reliable information [32].

Among the themes of the studied tweets, those containing prayers and wishes were the most common. This finding corroborates that reported in previous studies that showed that religious coping strategies were the most commonly followed strategies by patients with cancer [33]. A study on Arab women in Israel who had breast cancer showed that most of them used religious coping skills to actively cope with their disease [34].

This study had the following limitations. First, even though we searched for tweets containing representative spelling variations and abbreviations for the word “chemotherapy” in Arabic, there is still a possibility of some tweets being missed. For example, some tweets may have not mentioned chemotherapy directly but may have tweeted about it indirectly. Second, even with the large number of tweets retrieved and analyzed, the

XSL•FO
RenderX

representativeness of the users and their age groups to the whole Arabic-speaking populations cannot be guaranteed. Third, the period of the study was 6 months. Longer study periods (eg, 1 year) will possibly give a broader view and include seasonal spikes such as special occasions such as Ramadan wherein prayers and wishes for recovery can increase, which might add valuable information to the current findings. Fourth, lacking automated content analysis might limit the reproducibility of the findings to some extent. Fifth, the generalizability of the findings is limited as tweets were subjected to multiple external factors that could not be controlled such as the timing of the tweets and the geographical place wherein the users lived [35]. Lastly, some users may have appeared multiple times within the data set which, to some extent, limits the interpretation of the statistical analysis, as their views on the matter being discussed might be overrepresented. Future studies should account for bots and assess for their contributions in conversations about chemotherapy on Twitter.

In conclusion, our findings corroborate those of the previous studies that showed Twitter as a valuable social media platform to assess public perception and misconceptions about various health-related topics. Most of the tweets in our sample population showed supportive messages for patients undergoing chemotherapy, thereby suggesting that Twitter could play a supportive role for such patients, while the second most prevalent theme of the tweets in our study was misconceptions about chemotherapy. The findings of our exploratory analysis can help physicians and health care organizations tailor educational efforts in the future to address the common areas of misconceptions about chemotherapy, thereby leading to increased public acceptance of chemotherapy as a suitable mode of treatment for cancer.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Arabic keywords used for searching tweets related to chemotherapy. All the keywords are the different possible spellings/abbreviations of the word "chemotherapy".
[DOCX File , 19 KB - jmir_v22i7e13979_app1.docx]

References

1. Chabner BA, Roberts TG. Timeline: Chemotherapy and the war on cancer. Nat Rev Cancer 2005 Jan;5(1):65-72. [doi: 10.1038/nrc1529] [Medline: 15630416]
2. Williams GH, Stoeber K. The cell cycle and cancer. J Pathol 2012 Jan;226(2):352-364. [doi: 10.1002/path.3022] [Medline: 21990031]
3. Chemotherapy to Treat Cancer. National Cancer Institute. 2017. URL: https://www.cancer.gov/about-cancer/treatment/types/chemotherapy [accessed 2017-05-17]
4. Puts MTE, Tu HA, Tourangeau A, Howell D, Fitch M, Springall E, et al. Factors influencing adherence to cancer treatment in older adults with cancer: a systematic review. Ann Oncol 2014 Mar;25(3):564-577 [FREE Full text] [doi: 10.1093/annonc/mdt433] [Medline: 24285020]
5. Carelle N, Piotto E, Bellanger A, Germanaud J, Thuillier A, Khayat D. Changing patient perceptions of the side effects of cancer chemotherapy. Cancer 2002 Jul 01;95(1):155-163 [FREE Full text] [doi: 10.1002/cncr.10630] [Medline: 12115329]
6. Lindley C, McCune JS, Thomason TE, Lauder D, Sauls A, Adkins S, et al. Perception of chemotherapy side effects cancer versus noncancer patients. Cancer Pract 1999;7(2):59-65. [doi: 10.1046/j.1523-5394.1999.07205.x] [Medline: 10352062]
7. Gumusay O, Cetin B, Benekli M, Gurcan G, Ilhan MN, Bostankolu B, et al. Factors Influencing Chemotherapy Goal Perception in Newly Diagnosed Cancer Patients. J Cancer Educ 2016 Jun;31(2):308-313. [doi: 10.1007/s13187-015-0827-y] [Medline: 25851203]
8. McIlfatrick S, Sullivan K, McKenna H, Parahoo K. Patients' experiences of having chemotherapy in a day hospital setting. J Adv Nurs 2007 Aug;59(3):264-273. [doi: 10.1111/j.1365-2648.2007.04324.x] [Medline: 17590208]
9. What happens in an internet second: 54,907 Google searches, 7,252 tweets, 125,406 YouTube video views and 2,501,018 emails sent. DailyMail.com. 2017. URL: http://www.dailymail.co.uk/sciencetech/article-3662925/What-happens-internet-second-54-907-Google-searches-7-252-tweets-125-406-YouTube-video-views-2-501-018-emails-sent.html [accessed 2017-05-10]
10. About Twitter. URL: https://about.twitter.com/company [accessed 2017-05-17]
11. Salem F. Social Media and the Internet of Things: Towards Data-Driven Policymaking in the Arab World-Potential, Limits and Concerns. The Arab Social Media Report series. Dubai: Mohammed Bin Rashid School of Government; 2017. URL: https://www.academia.edu/31250728/Social_Media_and_the_Internet_of_Things_Towards_Data-Driven_Policymaking_in_the_Arab_World_-_Potential_Limits_and_Concerns [accessed 2019-09-08]
12. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]

XSL•FO
RenderX

13. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging Social Media to Promote Public Health Knowledge: Example of Cancer Awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17 [FREE Full text] [doi: 10.2196/publichealth.5205] [Medline: 27227152]
14. Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising Awareness About Cervical Cancer Using Twitter: Content Analysis of the 2015 #SmearForSmear Campaign. J Med Internet Res 2017 Dec 16;19(10):e344 [FREE Full text] [doi: 10.2196/jmir.8421] [Medline: 29038096]
15. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
16. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. J Med Internet Res 2014;16(5):e137 [FREE Full text] [doi: 10.2196/jmir.3298] [Medline: 24867458]
17. G Suite Marketplace. Twitter Archiver: Google Sheets add-on. URL: https://chrome.google.com/webstore/detail/twitter-archiver/pkanpfekacaojdncfgbjadedbggbbphi?hl=en [accessed 2017-05-11]
18. Beasley TM, Schumacker RE. Multiple Regression Approach to Analyzing Contingency Tables: Post Hoc and Planned Comparison Procedures. J Exp Educ 1995 Oct;64(1):79-93. [doi: 10.1080/00220973.1995.9943797]
19. Pemmaraju N, Utengen A, Gupta V, Kiladjian J, Mesa R, Thompson MA. Rare Cancers and Social Media: Analysis of Twitter Metrics in the First 2 Years of a Rare-Disease Community for Myeloproliferative Neoplasms on Social Media-#MPNSM. Curr Hematol Malig Rep 2017 Dec;12(6):598-604. [doi: 10.1007/s11899-017-0421-y] [Medline: 29105027]
20. Teoh D, Shaikh R, Vogel RI, Zoellner T, Carson L, Kulasingam S, et al. A Cross-Sectional Review of Cervical Cancer Messages on Twitter During Cervical Cancer Awareness Month. J Low Genit Tract Dis 2018 Jan;22(1):8-12 [FREE Full text] [doi: 10.1097/LGT.0000000000000363] [Medline: 29271850]
21. Crannell WC, Clark E, Jones C, James TA, Moore J. A pattern-matched Twitter analysis of US cancer-patient sentiments. J Surg Res 2016 Dec;206(2):536-542. [doi: 10.1016/j.jss.2016.06.050] [Medline: 27523257]
22. Sutton J, Vos SC, Olson MK, Woods C, Cohen E, Gibson CB, et al. Lung Cancer Messages on Twitter: Content Analysis and Evaluation. J Am Coll Radiol 2018 Jan;15(1 Pt B):210-217. [doi: 10.1016/j.jacr.2017.09.043] [Medline: 29154103]
23. Diddi P, Lundy LK. Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. J Health Commun 2017 Mar;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
24. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
25. Keim-Malpass J, Mitchell EM, Sun E, Kennedy C. Using Twitter to Understand Public Perceptions Regarding the #HPV Vaccine: Opportunities for Public Health Nurses to Engage in Social Marketing. Public Health Nurs 2017 Jul;34(4):316-323. [doi: 10.1111/phn.12318] [Medline: 28261846]
26. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
27. Rocheleau M, Sadasivam RS, Baquis K, Stahl H, Kinney RL, Pagoto SL, et al. An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of tweets. J Med Internet Res 2015 Jan 14;17(1):e18 [FREE Full text] [doi: 10.2196/jmir.3768] [Medline: 25589009]
28. Tang L, Bie B, Zhi D. Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease. Am J Infect Control 2018 Dec;46(12):1375-1380 [FREE Full text] [doi: 10.1016/j.ajic.2018.05.019] [Medline: 29929837]
29. Odlum M, Yoon S. Health Information Needs and Health Seeking Behavior During the 2014-2016 Ebola Outbreak: A Twitter Content Analysis. PLoS Curr 2018 Mar 23:10 [FREE Full text] [doi: 10.1371/currents.outbreaks.fa814fb2bec36e29b718ab6af66124fa] [Medline: 29707416]
30. Mattingly TJ. Innovative patient care practices using social media. J Am Pharm Assoc (2003) 2015;55(3):288-293. [doi: 10.1331/JAPhA.2015.14171] [Medline: 25859756]
31. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
32. What are Promoted Tweets? Twitter Business. 2018. URL: https://business.twitter.com/en/help/overview/what-are-promoted-tweets.html [accessed 2018-07-04]
33. Cohen M. An integrated view of cultural perceptions of cancer among Arab people in Israel. Health Psychol Rev 2014;8(4):490-508. [doi: 10.1080/17437199.2013.816205] [Medline: 25211212]
34. Goldblatt H, Cohen M, Azaiza F, Manassa R. Being within or being between? The cultural context of Arab women's experience of coping with breast cancer in Israel. Psychooncology 2013 Apr;22(4):869-875. [doi: 10.1002/pon.3078] [Medline: 22473753]

XSL•FO
RenderX

35. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Edited by G Eysenbach; submitted 12.03.19; peer-reviewed by R Gore, A Cyr, J Colditz, B Hoffman, N Cesare; comments to author 03.10.19; revised version received 27.02.20; accepted 13.06.20; published 29.07.20.

Please cite as:
Alghamdi A, Abumelha K, Allarakia J, Al-Shehri A
Conversations and Misconceptions About Chemotherapy in Arabic Tweets: Content Analysis
J Med Internet Res 2020;22(7):e13979
URL: http://www.jmir.org/2020/7/e13979/
doi:10.2196/13979
PMID:32723724

©Abdulrahman Alghamdi, Khalid Abumelha, Jawad Allarakia, Ahmed Al-Shehri. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study

Muhammad Syamsuddin[1], PhD; Muhammad Fakhruddin[1], MSc; Jane Theresa Marlen Sahetapy-Engel[1], BSc; Edy Soewono[1], Prof Dr

Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia

Corresponding Author:
Edy Soewono, Prof Dr
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Jalan Ganesha 10
Bandung, 40132
Indonesia
Phone: 62 8122000534
Email: esoewono@math.itb.ac.id

Abstract

Background: The popularity of dengue can be inferred from Google Trends that summarizes Google searches of related topics. Both the disease and its Google Trends have a similar source of causation in the dengue virus, leading us to hypothesize that dengue incidence and Google Trends results have a long-run equilibrium.

Objective: This research aimed to investigate the properties of this long-run equilibrium in the hope of using the information derived from Google Trends for the early detection of upcoming dengue outbreaks.

Methods: This research used the cointegration method to assess a long-run equilibrium between dengue incidence and Google Trends results. The long-run equilibrium was characterized by their linear combination that generated a stationary process. The Dickey-Fuller test was adopted to check the stationarity of the processes. An error correction model (ECM) was then adopted to measure deviations from the long-run equilibrium to examine the short-term and long-term effects. The resulting models were used to determine the Granger causality between the two processes. Additional information about the two processes was obtained by examining the impulse response function and variance decomposition.

Results: The Dickey-Fuller test supported an implicit null hypothesis that the dengue incidence and Google Trends results are nonstationary processes (P=.01). A further test showed that the processes were cointegrated (P=.01), indicating that their particular linear combination is a stationary process. These results permitted us to construct ECMs. The model showed the direction of causality of the two processes, indicating that Google Trends results will Granger-cause dengue incidence (not in the reverse order).

Conclusions: Various hypothesis testing results in this research concluded that Google Trends results can be used as an initial indicator of upcoming dengue outbreaks.

(J Med Internet Res 2020;22(7):e17633) doi:10.2196/17633

KEYWORDS

dengue; Google Trends; infodemiology; infoveillance; vector error correction model; Granger causality

Introduction

Dengue is known as an infectious disease, which is caused by the dengue virus from *Flaviviridae* and genus *Flavivirus* families. This virus has four serotypes, namely DEN-1, DEN-2, DEN-3, and DEN-4 [1-3]. Infection by one of these four serotypes does not give cross-protective immunity. Hence, people who live in endemic areas can be reinfected by the other three serotypes throughout their lifetime [4]. *Aedes aegypti* and *Aedes albopictus* mosquitoes are vector transmitters of dengue. The disease is transmitted by mosquitoes through arthropod vectors in tropical and subtropical areas around the world [5]. As the most rapidly spreading mosquito-borne disease in the

XSL•FO
RenderX

world, dengue fever has affected the lives of approximately 1.8 billion people in Southeast Asia alone. In the dengue-endemic region, Indonesia is one of the largest countries, with a population of 267 million [6]. Since the first dengue incidents reported in 1968, the number and range of dengue incidents in Indonesia have increased nationwide [7]. All 34 provinces in Indonesia have been reported to have dengue cases, showing the extensive range of the disease [8]. Early detection of disease activity can reduce the impact of the disease [9].

Bandung is one of the crowded cities in Indonesia. It has the highest dengue incidence, especially in West Java. Daily habits, landscape structures, weather, and the ecosystem in the city play roles in dengue vector breeding as primary factors for dengue transmission. The climate in Bandung is a mountainous climate (humid and cold), with an average temperature of 23.5°C. The average rainfall is 200.4 mm, and there are on average 21.3 rainy days per month. It is an ideal environment for *Aedes aegypti*. According to the Extraordinary Early Childhood Awareness System (SKDKLB-DBD) report, Bandung had the highest dengue incidence from 2002 to 2006, with a total of 22,335 infected people. In January 2019, the West Java Provincial Health Office recorded 236 dengue cases in Bandung. For these reasons, we selected Bandung as our study area to investigate and analyze the association between dengue data from Google Trends and dengue incidence data from a reputable hospital in Bandung.

In this modern world, it is impossible to say that technology, especially the internet, does not influence human lives. Over the years, research has been performed to investigate the accuracy of using internet search engine data to predict real-life phenomena, such as influenza epidemics and flu trends [9], stock markets [10-12], house prices [13,14], and tourism demand [15-17]. Google Trends is a public website belonging to Google Inc that offers data based on Google Search, which shows how frequently a particular search term is entered.

According to StatCounter, in 2016, Google was the most used text search engine in Indonesia. About 97% of people who use the internet in Indonesia use Google. It is assumed that Indonesian people show the trends to find information about dengue on the internet [18]. Therefore, we hypothesized that the popularity of dengue on Google has a correlation with the dengue incidence in Indonesia. Several studies about Google Trends results and the relationship with various diseases have been carried out, such as a study on dengue fever in Indonesia by using moving average analysis [19]. Other Google Trends–related work is presented in the following sentences. Dengue in several countries (Bolivia, Brazil, India, Indonesia, and Singapore) has been studied to maximize a fitting model by using a univariate linear model [20]. A spatiotemporal analysis of dengue incidence has been performed by using an exponential generalized autoregressive condition heteroscedastic model [21]. In other studies, we can see a web-based search for the early identification of the disease prevalence of coronary heart disease [22], forecasting of influenza cases using internet data [23], the use of Google Trends in health care research by using correlation analysis [24], infodemiology and infoveillance [25-27], a framework of social media data and quality assessment for a reporting standard [28], the spread of pertussis in Europe [29], and the spread of AIDS in the United States [30]. This led us to hypothesize that dengue incidence and Google Trends results have a long-run equilibrium.

We started our research with an initial hypothesis that the popularity of dengue on Google correlates with dengue cases in Bandung. We then investigated the relationship between these two data by using the Dicky-Fuller test, error correction model (ECM), impulse response function, and variance decomposition. We hoped that information from Google Trends can be used for the early detection of upcoming dengue outbreaks so that policymakers can prepare for the early prevention or control of the epidemic.

Methods

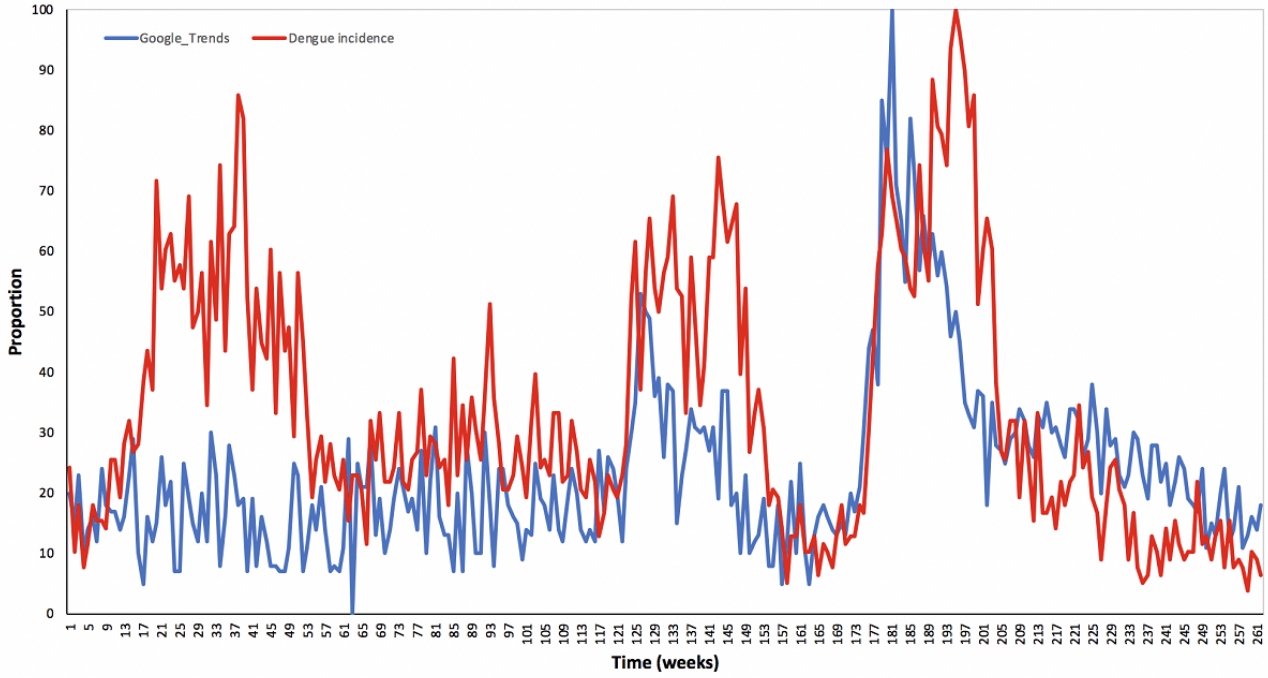
Collected Data

Google Trends is a website that analyzes the popularity of a topic in various countries and various languages based on search requests. The data source is over the internet and open source and can be easily accessed by everyone. In Google Trends, a user can enter a keyword in the form of words or phrases related to the selected topic or cases. Google Trends is not case sensitive but takes into account spelling errors that might occur. Users can specify the duration of time they want to review by selecting a time range or specifying a date. In addition, users can specify the area to be reviewed by selecting the appropriate country, city, or province or state. They can also see the popularity of these keywords globally by selecting the option *worldwide*.

Data used in this study are time-series data of dengue incidence from Santo Borromeus Hospital in Bandung, as well as popularity data taken from Google Trends via the website (Figure 1). The time range used in this paper is from September 9, 2012 until September 7, 2017. We consider weekly data over the time interval, and there were 261 data points. We assumed that the incidence data represent all dengue cases in Bandung owing to the location of the hospital in the city center. Google Trends data were obtained from the Google Trends website on google [31] by entering the keyword “demam berdarah dengue” plus “dbd.” With regard to the technique of taking data, we followed previous guidelines [32] when using Google Trends for valid results in our study by selecting the appropriate keyword(s), region(s), period, and category. Quotation marks were used so that the search results only showed the popularity of keywords in that exact order. The data represented the popularity of the keyword “demam berdarah dengue” plus “dbd” found on Google. Google Trends normalizes its popularity data by dividing each data point by the total amount of searches at a given time and location. This results in a proportion scaled in a range from 0 to 100. This scale shows the popularity of specific keywords relative to the time and location of the query.

XSL•FO
RenderX

Figure 1. Dengue data plot from Google Trends and reported cases in Bandung.



Stationary Test

We performed a stationary test for the time series data of Google Trends (X_t) and the dengue incidence (Y_t). A time series $\{y_t\}$ is said to be stationary if it satisfies the following conditions: (1) $E[y_t] = \mu < \infty$; (2) $Var[y_t] = \sigma^2 < \infty$; and (3) $Cov[y_t, y_{t+s}] = \gamma_s < \infty$, for s>0.

Differencing a series produces another set of observations, such as the first differenced values, where $\triangle y_t = y_t - y_{t-1}$. Generalizing this operation and performing the difference operation as much as n times can be written as $\triangle^n y_t = y_t - y_{t-n}$. If a series is stationary without any differencing, it is said to be integrated of the order 0 or $I(0)$. However, if it is stationary only after differencing once, it is said to be integrated of the order 1 or $I(1)$. The Dickey-Fuller test was used to detect the presence of a unit root and determine the stationarity of Google Trends and dengue fever incidence series.

Cointegration Test

For cointegration, Engle and Granger [33] used the stationarity test of the residual series obtained from the long-run equilibrium equation. If the residual series, denoted by $\{e_t\}$, is stationary, given that y_t and z_t are first-order stationary, they are cointegrated.

After finding the Google Trends and dengue incidence series to be first-order difference stationary, the long-run equilibrium relationship can be stated in the following form:

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad (1)$$

where e_t denotes the residual.

Let $\{\hat{e}_t\}$ be the residual sequence. The series $\{\hat{e}_t\}$ contains the estimated values of deviations from the long-run relationship. By using the Dickey-Fuller test to check its stationarity, it was found that the level values of $\{\hat{e}_t\}$ were stationary.

Estimation and Analysis of a Vector ECM

After a cointegrating relationship has been established, an ECM can be built to establish the short-run relationship between two variables. A likelihood ratio test can be used to determine the time lag of the vector ECM or the value of p. The regression equation for an ECM is as follows:



Analysis of cointegration shows that Google Trends and dengue incidence have a long-run equilibrium relationship. However, they are in disequilibrium in the short term. View equations 2 and 3 as a vector autoregression (VAR) model as follows:



Hence, the vector ECM at hand can be written as a VAR model as follows:



Before estimating the vector ECM, the optimal lag order is first determined.

Causal Relationship Between Google Trends and Dengue Incidence

One way to test causality is to see whether the time lag of one variable is relevant for another variable. In a two-equation system with stationary variables y_t and z_t with p lags, it is said that $\{y_t\}$ does not Granger-cause $\{z_t\}$ if and only if the coefficients of y_t in the equation for z_t are equal to zero. In other words, if $\{y_t\}$ does not provide improvement for the forecasting performance of $\{z_t\}$, $\{y_t\}$ does not Granger-cause $\{z_t\}$. Granger

XSL•FO
RenderX

causality only shows the effects of past values of $\{y_t\}$ toward the current values of $\{z_t\}$.

In order to test Granger causality, a standard F test of the restriction $a_{21}(1) = a_{21}(2) = \ldots = a_{21}(p) = 0$ is performed.

In a cointegrated system, X_t does not Granger-cause Y_t if the values of ΔX_{t-i} do not enter in the equation of ΔY_t and if Y_t does not respond to deviation from the long-run equilibrium.

Impulse Response Function and Variance Decomposition

To analyze the dynamic effects of the model in response to shocks and the effects on the two variables, the impulse response function and variance decomposition were examined.

Results

Stationary Test

The stationary test results can be seen in Table 1, Figure 2A, and Figure 2B. The table shows that the level values of the Google Trends and dengue incidence series were nonstationary. However, Google Trends and dengue incidence data were found to be stationary after being differentiated once. This was done to reduce the fluctuations in the data.

Table 1. Dickey-Fuller test for Google Trends data, dengue incidence data, first differenced Google Trends data, and first differenced dengue incidence data.

Variable	Dickey-Fuller test statistic (value)	Dickey-Fuller critical value (N=250)
X_t [a]	−2.42 (.02)	−2.58
Y_t [b]	−2.24 (.03)	−2.58
ΔX_t [c]	−21.76 (.01)	−2.58
ΔY_t [d]	−27.85 (.01)	−2.58

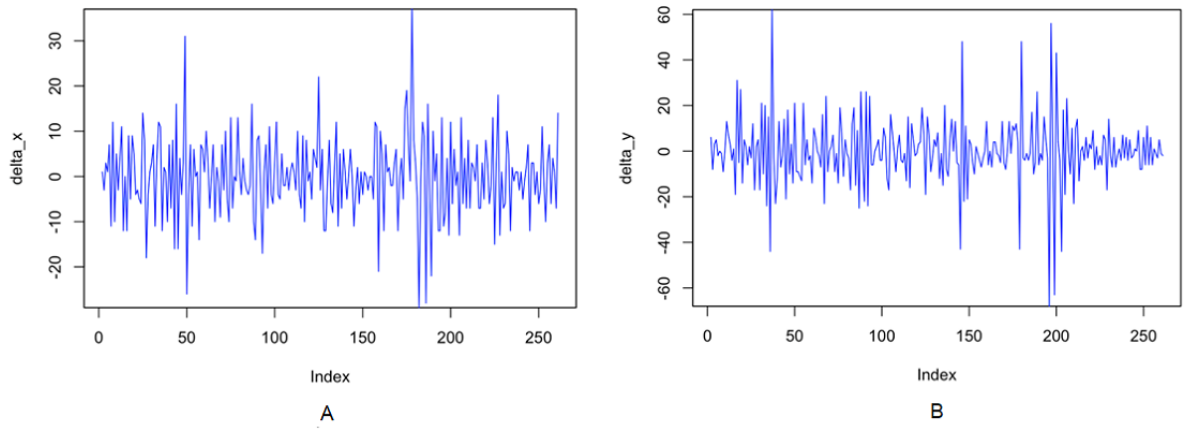
[a]Google Trends data.

[b]Dengue incidence data.

[c]First differenced Google Trends data.

[d]First differenced dengue incidence data.

Figure 2. (A) $\{X_t\}$ and (B) $\{Y_t\}$ plots with one-time difference. X_t: Google Trends data; Y_t: dengue incidence data.



Cointegration Test

The cointegration test results of ordinary least squares regression yielded that the long-run equilibrium relationship can be shown as follows:



with e_t denoting the residual.

Let $\{\hat{e}_t\}$ be the residual sequence. The series $\{\hat{e}_t\}$ contains the estimated values of deviations from the long-run relationship. By using the Dickey-Fuller test to check its stationarity, it was found that the level values of $\{\hat{e}_t\}$ were stationary. The results are shown in Table 2.

From previous results, it was seen that $\{X_t\}$ and $\{Y_t\}$ are $I(1)$, and because the $\{\hat{e}_t\}$ series is stationary,$\{X_t\}$ and $\{Y_t\}$ are cointegrated. Hence, a vector ECM can be constructed.

XSL•FO
RenderX

Table 2. Dickey-Fuller test for the residual sequence.

Variable	Dickey-Fuller test statistic	Dickey-Fuller critical value (N=250)
e_t [a]	−8.77	−2.58

[a]residual estimated as follows: $e_t = Y_t - 12.609 - 0.455X_t$

Likelihood Ratio Test to Find the Time Lag

The longest feasible lag length was set as 8 weeks. Thereafter, the value of the determinant of the variance-covariance matrix of a model with lag length eight was examined (denoted as Σ_8) and compared with that of a model with lag length seven (denoted as Σ_7). The likelihood ratio is $(T - c)(\ln\Sigma_7 - \ln\Sigma_8)$, where T is the number of observations and c is the number of parameters that are estimated in each equation of the unrestricted system. In the case of comparing the eight-lag model to the seven-lag model, the value of c is $1 + 8n$, with n being the number of variables, which is two in this case. If the likelihood ratio is smaller than the critical value (χ^2_4 at a significance of α=1%), the null hypothesis of the restriction A_8=0 is rejected. This is done until lag 1.

The results of this test are shown in Table 3. The likelihood ratio test showed that the optimal number of lags needed for this vector ECM is three.

Table 3. Likelihood ratio test for lag length.

Number	H_0	H_1	Likelihood ratio	χ^2_4	Verdict
1	A_8=0	A_8≠0	7.655	13.277	H_0 rejected
2	A_7=0	A_7≠0|A_8=0	3.291	13.277	H_0 rejected
3	A_6=0	A_6≠0|A_8=A_7=0	0.221	13.277	H_0 rejected
4	A_5=0	A_5≠0|A_8=…A_6=0	2.543	13.277	H_0 rejected
5	A_4=0	A_4≠0|A_8=…A_5=0	6.191	13.277	H_0 rejected
6	A_3=0	A_3≠0|A_8=…A_4=0	19.666	13.277	H_0 rejected
7	A_2=0	A_2≠0|A_8=…A_3=0	27.887	13.277	H_0 rejected
8	A_1=0	A_1≠0|A_8=…A_2=0	60.361	13.277	H_0 rejected

Estimation of the ECM

After finding the optimal number of lags, an ECM model was built. The estimated vector ECM is as follows:



From the equation, it is seen that the speed of the adjustment parameter is −0.1816 for $\{Y_t\}$ and −0.0267 for $\{X_t\}$. This means that when there is a deviation of 1 from the long-run equilibrium in the period $t - 1$, the number of dengue incidences will decrease by 0.1816 and dengue popularity in Google will decrease by 0.0267.

The speed of adjustment parameter for dengue incidence was nine times larger than the value for Google Trends, meaning that dengue incidence is more responsive to deviations from the long-run equilibrium. On the other hand, Google Trends only responds slightly to the aforementioned deviation.

It was found that this model has an R-squared value of 0.4128 for the ΔX_t equation and 0.1511 for the ΔY_t equation, as well as an Akaike Information Criterion (AIC) value of 2370.2. Since the R-squared value is quite low, it can be said that the model cannot explain the data at hand accurately.

Causal Relationship Between Google Trends and Dengue Incidence

Based on the vector ECM in equation 7, Granger causality was tested between Google Trends and dengue incidence. It was noted that at lag 2 and 3, X_t Granger-causes Y_t at a significance of α=5%, since the P obtained was similar (.04). However, Y_t does not Granger-cause X_t. This means that information from past values of Google Trends regarding dengue at a lag of 2 and 3 weeks is useful for explaining the present value of dengue incidence. The results are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Granger causality test for Google Trends data and dengue incidence data.

Part 1: Does Google Trends Granger-cause incidence?



H_0: $\beta_1=\ldots=\beta_j=0$ (Google Trends does not Granger-cause incidence)



Part 2: Does incidence Granger-cause Google Trends?



H_0: $\beta_1=\ldots=\beta_j=0$ (incidence does not Granger-cause Google Trends)



Impulse Response Function and Variance Decomposition

The results for 12 periods (3 months) are obtained as presented below.

Impulse Response Function

As shown in Figure 3B, a positive shock in dengue popularity on Google Trends has a relevant impact on dengue incidence. Dengue incidence shows a large increase after two periods. Thereafter, it shows a slight decrease, but then, it increases again slowly. Its effects seem to be long term, since the incidence keeps increasing until the end of the 12 periods. This phenomenon suggests that shocks in dengue fever popularity on Google have a relevant impact on dengue fever incidence.

On the other hand, through analysis of the response of dengue incidence to a positive shock, it was found that dengue popularity increases slightly and then remains constant. This behavior is presented in Figure 3A. After a positive shock in dengue incidence, there is little fluctuation in its popularity on Google Trends. This suggests that shocks in dengue incidence do not have a relevant effect on its popularity on Google Trends. Its effects are only short term and do not remain in the long run.

Generally, the impulse response function shows that Google Trends has a relevant impact on dengue fever incidence and has a long-term effect. On the contrary, dengue incidence has only a short-term and small effect on the popularity of dengue on Google.

Variance Decomposition

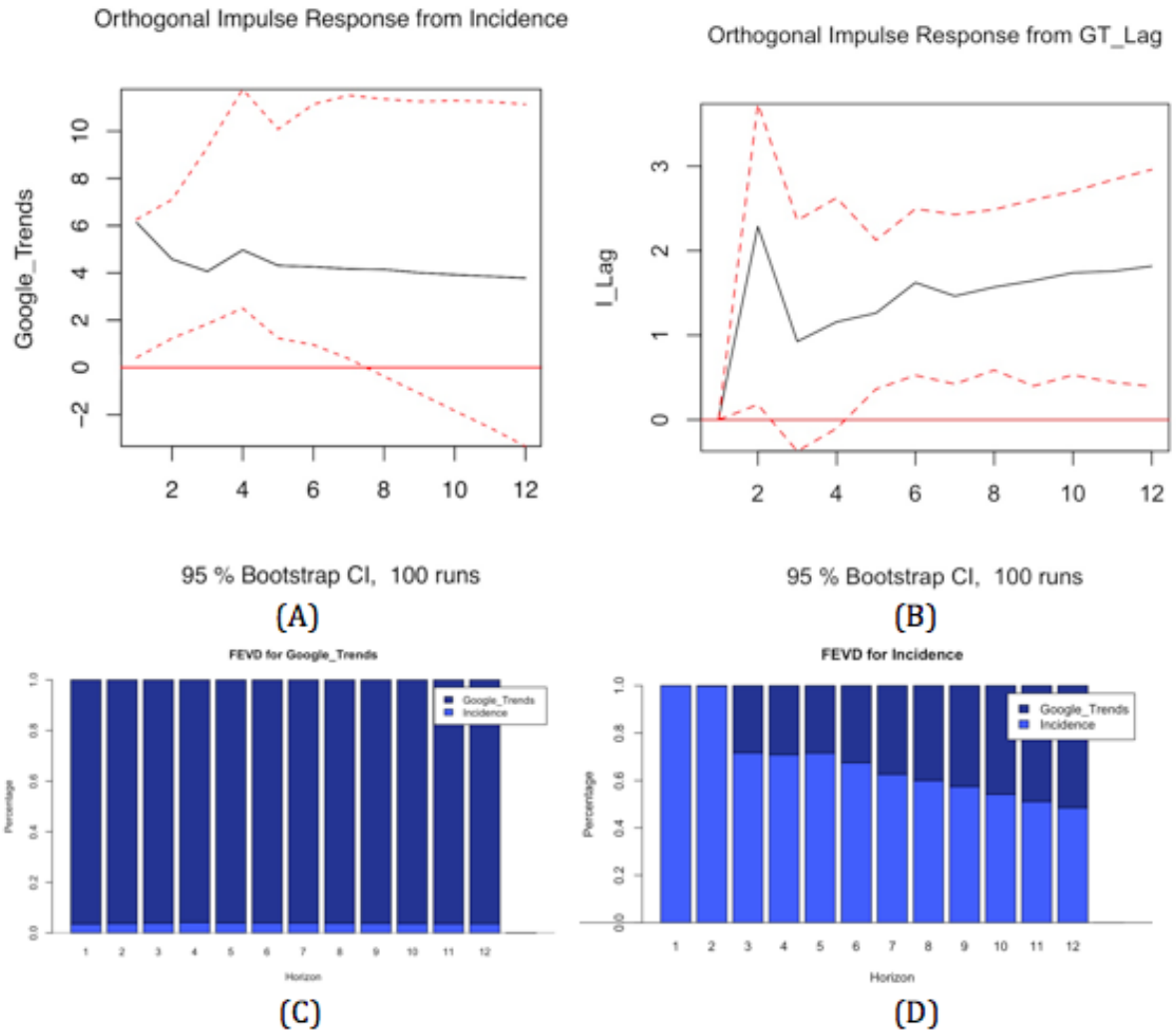
Variance decomposition estimates the contribution of shocks in a variable toward the response of another variable. As shown in Figure 3D, the contribution of dengue incidence to its variance gradually declines in the first two periods. Thereafter, it declines further until the contribution of dengue incidence is finally only around 40%. In the first period, Google Trends only has a small contribution to dengue incidence variance with only 0%. Thereafter, during the second period, it increases to around 28% and then continues to climb gradually. After the 12 periods, Google Trends has up to 60% contribution to dengue incidence variance.

On the other hand, Figure 3C shows that Google Trends variance mainly comes from itself, where dengue incidence only contributes at a rate of approximately 4%. This rate increases in the first four periods (from 3.6% in the first period to 4.2% in the fourth period). Thereafter, the contribution rate from dengue incidence remains around 3.7%. This means that dengue incidence very slightly influences Google Trends in the short term, but does not influence the popularity of Google Trends in the long term.

In summary, it can be seen that Google Trends influences dengue incidence in the long term, but dengue incidence only influences Google Trends in the short term and not in the long term. As presented in the model, dengue incidence is related to not only the popularity of dengue in Google but also its lagged value of up to 1 week.

XSL•FO
RenderX

Figure 3. Impulse response function of (A) dengue data from Google Trends with respect to reported cases and (B) reported cases with respect to dengue data from Google Trends. Forecast error variance decomposition of (C) dengue data from Google Trends and (D) reported cases.



Discussion

Principal Findings

Our results show that there is indeed a causal relationship between dengue popularity in Google Trends and dengue incidence in Bandung. A Granger cointegrated relationship between dengue popularity in Google Trends and dengue incidence in Bandung was noted. This is justified because both data sets were found to be *I*(1), and the residual from the ordinary least squares regression was also found to be stationary.

Based on the ECM, it can be seen that there is a relationship between Google Trends results and dengue incidence. Through Granger analysis, it was seen that Google Trends Granger-causes dengue incidence in Bandung at a lag of 2 and 3 weeks. This was further supported by the impulse response function, where shocks in dengue popularity in Google cause dengue incidence to increase. It was also supported by the variance decomposition, where after 1 week, the contribution from Google Trends to dengue incidence variance increases. Granger analysis also showed that dengue incidence does not Granger-cause its popularity in Google.

The vector ECM also showed that dengue incidence is more responsive to deviations from the long-run equilibrium, since it has a larger value of the speed of adjustment, which is nine times the value for Google Trends.

Limitations

The results showed a causal relationship between dengue popularity in Google Trends and dengue incidence in Bandung. However, this exact ECM cannot be used for forecasting or early detection owing to the low R-squared values of 0.4128 for the Google Trends equation and 0.1511 for the dengue incidence equation. A further improved model will need to be built for future forecasting.

The results of this study can help provide a more real-time indication of dengue outbreaks in Bandung. Owing to Indonesia's standard and traditional approach to dengue surveillance, the data of dengue cases have several weaknesses, such as low accuracy and timeliness [20]. In addition, data available from Santo Borromeus Hospital were only from 2008 until 2017. Owing to the limitations of Google Trends, it was preferable to use weekly data with a 5-year period (giving 260 data points) rather than data from 2008 to 2014 on a monthly

XSL•FO
RenderX

basis with only 84 data points. Another limitation is that people searching for dengue-related information may not necessarily have the disease, as they could be searching because a relative or friend is ill. Besides, a search in Bandung does not necessarily mean that the intended sick person is in Bandung. The individual could be searching for someone else who is ill in another city.

Our proposed model used strong assumptions, such as the behavior of the use of gadgets and social media in the community, which is quite high, and a good internet signal in the observation area (Bandung in this case). Therefore, it is risky to implement the findings in areas with low internet access.

Conclusions

Google Trends data may be used as an initial indicator of a dengue outbreak in Bandung. However, further improvements to the ECM need to be made by using more data points to gain more extensive insights.

Acknowledgments

Part of this work was supported by the Indonesian RistekDikti Grant 2020. The second author gratefully acknowledges the financial support provided by the Indonesia Ministry of Research and Technology through the Pendidikan Magister menuju Doktor untuk Sarjana Unggul (PMDSU) Program.

Authors' Contributions

MS performed the statistical analysis, review, and interpretation. MF and JTMSE performed numerical simulation and constructed this paper. ES interpreted and reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Halstead SB, Heinz FX, Barrett A, Roehrig JT. Dengue virus: molecular basis of cell entry and pathogenesis, 25-27 June 2003, Vienna, Austria. Vaccine 2005 Jan 04;23(7):849-856. [doi: 10.1016/j.vaccine.2004.03.069] [Medline: 15603884]
2. Malavige GN, Fernando S, Fernando DJ, Seneviratne SL. Dengue viral infections. Postgrad Med J 2004 Oct;80(948):588-601 [FREE Full text] [doi: 10.1136/pgmj.2004.019638] [Medline: 15466994]
3. Fakhruddin M, Putra PS, Wijaya KP, Sopaheluwakan A, Satyaningsih R, Komalasari KE, Mamenun, Sumiati, et al. Assessing the interplay between dengue incidence and weather in Jakarta via a clustering integrated multiple regression model. Ecological Complexity 2019 Aug;39:100768. [doi: 10.1016/j.ecocom.2019.100768]
4. Westaway E, Brinton M, Gaidamovich SYa, Horzinek M, Igarashi A, Käériäinen L, et al. Flaviviridae. Intervirology 1985;24(4):183-192. [doi: 10.1159/000149642] [Medline: 3000978]
5. Lu L, Lin H, Tian L, Yang W, Sun J, Liu Q. Time series analysis of dengue fever and weather in Guangzhou, China. BMC Public Health 2009 Oct 27;9:395 [FREE Full text] [doi: 10.1186/1471-2458-9-395] [Medline: 19860867]
6. Karyanti MR, Uiterwaal CS, Kusriastuti R, Hadinegoro SR, Rovers MM, Heesterbeek H, et al. The changing incidence of dengue haemorrhagic fever in Indonesia: a 45-year registry-based analysis. BMC Infect Dis 2014 Jul 26;14:412 [FREE Full text] [doi: 10.1186/1471-2334-14-412] [Medline: 25064368]
7. Wahyono TY, Nealon J, Beucher S, Prayitno A, Moureau A, Nawawi S, et al. Indonesian dengue burden estimates: review of evidence by an expert panel. Epidemiol Infect 2017 Aug;145(11):2324-2329 [FREE Full text] [doi: 10.1017/S0950268817001030] [Medline: 28545598]
8. Wardhani P, Aryati A, Yohan B, Trimarsanto H, Setianingsih TY, Puspitasari D, et al. Clinical and virological characteristics of dengue in Surabaya, Indonesia. PLoS One 2017;12(6):e0178443 [FREE Full text] [doi: 10.1371/journal.pone.0178443] [Medline: 28575000]
9. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
10. Da Z, Engelberg J, Gao P. In search of attention. The Journal of Finance 2011 Oct;66(5):1461-1499. [doi: 10.1111/j.1540-6261.2011.01679.x]
11. Preis T, Moat HS, Stanley HE. Quantifying trading behavior in financial markets using Google Trends. Sci Rep 2013;3:1684 [FREE Full text] [doi: 10.1038/srep01684] [Medline: 23619126]
12. Hamid A, Heiden M. Forecasting volatility with empirical similarity and Google Trends. Journal of Economic Behavior & Organization 2015 Sep;117:62-81. [doi: 10.1016/j.jebo.2015.06.005]
13. Wu L, Brynjolfsson E. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. United States: University of Chicago Press; Apr 2015:89-118.
14. Askitas N, Zimmermann KF. Detecting mortgage delinquencies. Social Science Research Network 2011 Jul:1-16 [FREE Full text]
15. Song H, Liu H. Predicting Tourist Demand Using Big Data. In: Analytics in Smart Tourism Design. Switzerland: Springer; Oct 2016:13-29.

XSL•FO
RenderX

16. Pan B, Chenguang Wu D, Song H. Forecasting hotel room demand using search engine data. JHTT 2012 Sep 28;3(3):196-210. [doi: 10.1108/17579881211264486]
17. Yang X, Pan B, Evans JA, Lv B. Forecasting Chinese tourist volume with search engine data. Tourism Management 2015 Feb;46:386-397. [doi: 10.1016/j.tourman.2014.07.019]
18. Prabowo H, Hamdani R, Sanusi Z. The New Face of People Power: An Exploratory Study on the Potential of Social Media for Combating Corruption in Indonesia. AABFJ 2018;12(3):19-40. [doi: 10.14453/aabfj.v12i3.3]
19. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Glob Health Action 2019;12(1):1552652 [FREE Full text] [doi: 10.1080/16549716.2018.1552652] [Medline: 31154985]
20. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011 May;5(5):e1206 [FREE Full text] [doi: 10.1371/journal.pntd.0001206] [Medline: 21647308]
21. Risteski D, Davcev D. Can we use daily Internet search query data to improve predicting power of EGARCH models for financial time series volatility. In: Proceedings of the International Conference on Computer Science and Information Systems (ICSIS). 2014 Oct Presented at: International Conference on Computer Science and Information Systems (ICSIS); October 17-18, 2014; Dubai, United Arab Emirates. [doi: 10.15242/iie.e1014066]
22. Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of Search Engine Queries for Chest Pain With Coronary Heart Disease Epidemiology. JAMA Cardiol 2018 Dec 01;3(12):1218-1221 [FREE Full text] [doi: 10.1001/jamacardio.2018.3459] [Medline: 30422176]
23. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. JMIR Public Health Surveill 2018 Jan 09;4(1):e4 [FREE Full text] [doi: 10.2196/publichealth.8950] [Medline: 29317382]
24. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
25. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
26. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0]
27. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
28. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. J Med Internet Res 2016 Feb 26;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
29. Gianfredi V, Bragazzi N, Mahamid M, Bisharat B, Mahroum N, Amital H, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. Public Health 2018 Dec;165:9-15. [doi: 10.1016/j.puhe.2018.09.001] [Medline: 30342281]
30. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. J Big Data 2018 May 19;5(1). [doi: 10.1186/s40537-018-0126-7]
31. Google Trends. URL: https://trends.google.co.id/trends/?geo=ID [accessed 2018-05-10]
32. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
33. Engle RF, Granger CW. Co-Integration and Error Correction: Representation, Estimation, and Testing. Econometrica 1987 Mar;55(2):251. [doi: 10.2307/1913236]

Abbreviations

ECM: error correction model
VAR: vector autoregression

XSL•FO
RenderX

Edited by G Eysenbach; submitted 30.12.19; peer-reviewed by H Susanto, A Mavragani, S Kardes; comments to author 10.03.20; revised version received 03.05.20; accepted 20.05.20; published 24.07.20.

Please cite as:
Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E
Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study
J Med Internet Res 2020;22(7):e17633
URL: http://www.jmir.org/2020/7/e17633/
doi:10.2196/17633
PMID:32706682

©Muhammad Syamsuddin, Muhammad Fakhruddin, Jane Theresa Marlen Sahetapy-Engel, Edy Soewono. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigation of Geographic and Macrolevel Variations in LGBTQ Patient Experiences: Longitudinal Social Media Analysis

Yulin Hswen[1,2], ScD; Amanda Zhang[3,4], BA; Kara C Sewalk[3], MPH; Gaurav Tuli[3], PhD; John S Brownstein[2,3], PhD; Jared B Hawkins[2,3], MMSc, PhD

[1]Bakar Computational Health Sciences Institute, Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

[2]Computational Epidemiology Lab, Harvard Medical School, Boston, MA, United States

[3]Innovation Program, Boston Children's Hospital, Boston, MA, United States

[4]Pritzker School of Medicine, The University of Chicago, Chicago, IL, United States

Corresponding Author:
Yulin Hswen, ScD
Computational Epidemiology Lab
Harvard Medical School
401 Park Drive
Boston, MA
United States
Phone: 1 6177751889
Email: yhswen@gmail.com

Abstract

Background: Discrimination in the health care system contributes to worse health outcomes among lesbian, gay, bisexual, transgender, and queer (LGBTQ) patients.

Objective: The aim of this study is to examine disparities in patient experience among LGBTQ persons using social media data.

Methods: We collected patient experience data from Twitter from February 2013 to February 2017 in the United States. We compared the sentiment of patient experience tweets between Twitter users who self-identified as LGBTQ and non-LGBTQ. The effect of state-level partisan identity on patient experience sentiment and differences between LGBTQ users and non-LGBTQ users were analyzed.

Results: We observed lower (more negative) patient experience sentiment among 13,689 LGBTQ users compared to 1,362,395 non-LGBTQ users. Increasing state-level liberal political identification was associated with higher patient experience sentiment among all users but had stronger effects for LGBTQ users.

Conclusions: Our findings highlight that social media data can yield insights about patient experience for LGBTQ persons and suggest that a state-level sociopolitical environment influences patient experience for this group. Efforts are needed to reduce disparities in patient care for LGBTQ persons while taking into context the effect of the political climate on these inequities.

(J Med Internet Res 2020;22(7):e17087) doi:10.2196/17087

KEYWORDS

LGBTQ; sexual and gender minorities; health care quality; health care disparities; social media; digital health; sentiment analysis; infodemiology

XSL•FO
RenderX

Introduction

Health Disparities and Discrimination in Health Care Among Lesbian, Gay, Bisexual, Transgender, and Queer Patients

Across several health indicators, lesbian, gay, bisexual, transgender, and queer (LGBTQ) people consistently experience worse health outcomes than their non-LGBTQ counterparts [1]. Discrimination in the health care system is associated with worse health outcomes [2-4]. Previous studies have shown that LGBTQ persons report increased rates of discrimination from medical providers and other staff across a wide variety of health care settings [5,6]. In one study, almost one-third of transgender survey respondents reported postponing medical care because they experienced discrimination [6]. In addition, many LGBTQ persons report reluctance to disclose their sexual orientation or gender identity to their health care providers [7,8]. Such an environment results in a lack of understanding and acceptance of LGBTQ people and their specific health care needs, and leads to inadequate treatment and an erosion of trust in the health care services for this group [7,8]. Although there has been increasing awareness and understanding of LGBTQ patient experiences, many recent studies have consisted of small qualitative inquiries focused on specific LGBTQ populations [9-12]. Furthermore, there is a dearth of valid scales and indices that measure LGBTQ patient experience [13] as well as limited research on the sociopolitical cultural factors that contribute to these discriminatory accounts.

Geographic Disparities in LGBTQ Health Care

There is a large body of evidence characterizing regional variation in health care and patient experiences across the United States [14-16]. Some studies have further stratified this variation by social variables such as socioeconomic status and race [15,17]. However, research has yet to explore the geographic disparities for LGBTQ care due to methodological limitations in the identification of LGBTQ patients [18]. US states vary considerably in their policies and practices that provide protection against sexual orientation and identity-based discrimination. In the areas of employment, housing, public accommodations, and health care services, there is wide variation of legislatures and pre-empt local protections that help protect LGBTQ persons by ensuring fair and equal treatment [19,20]. Although certain states have adopted protection laws for LGBTQ persons, other states have not been committed to the passage or enforcement of local nondiscrimination laws [21]. In fact, some states have policies that explicitly prevent the passage or enforcement of local nondiscriminatory laws, including those that relate to health care [21]. States without LGBTQ protective policies also tend to have higher percentages of conservative voters [22], and the social climate of a geographic area has been shown to be associated with differences in health outcomes for LGBTQ persons [23,24]. Prior research has also shown that residence in areas with a higher percentage of Republican voters is associated with a greater risk for depression among LGBTQ youth at the neighborhood level [25] and health care refusal among transgender patients at the state level [21]. In addition, physician political identity has been shown to be highly correlated with treatment decisions, with physician partisan bias leading to variation in patient care [26]. However, a large-scale, geographically contextual analysis has not been conducted to study the effect of the political environment on LGBTQ patient experience.

Twitter as an Outlet for Patient Feedback

Recently there has been a shift in US health care to emphasize patient experiences, as this has been linked with quality of care [16]. Novel methods based on online data have been applied to studying this field [27]. For instance, the social media platform Twitter has been shown to be an effective resource for obtaining unsolicited feedback on quality of hospital care [16,27-29]. It has also been validated as a method to characterize differences in LGBTQ hospital care across the United States [30]. This previous study assessed how hospitals either supported or did not support LGBTQ care and showed that hospitals deemed as having LGBTQ equitable policies were also shown to be more supportive toward LGBTQ practices on Twitter. However, it did not investigate LGBTQ patient users' experiences or how geopolitical state-level factors may influence the sentiment of these experiences. Other studies have demonstrated the feasibility and promise of real time social media sites to study the patient experiences of LGBTQ communities but only in small-scale content analyses, and they did not consider how geographic variation in political identity may shape these experiences [31,32]. In these ways, online social media information from LGBTQ patients can provide researchers with unfiltered accounts of patient experience and help develop an understanding of the complexities surrounding LGBTQ health disparities [33,34].

This study sought to fill the gap in research by examining the geographic variation in patient experiences among LGBTQ persons in the United States using novel computer science methods to curate a large-scale data set from the online social media platform Twitter. The goal of this research was to identify differences in patient experience among LGBTQ and non-LGBTQ patients, and to statistically model the state-level geopolitical factors that are associated with LGBTQ patient experiences using Twitter. Understanding the political factors associated with LGBTQ patient experience can be used to uncover reasons for these LGBTQ disparities in care and inform the development of targeted interventions to improve equity and advocate for this marginalized group.

Methods

Sample

Health Care Patient Experience Twitter Data

Tweets relating to patient experience from February 16, 2013, to February 15, 2017, were collected. A supervised machine classifier was built from a combination of keywords and rule-based learning algorithms to identify tweets related to health care patient experiences in the areas of medical facility and staff, medical procedures, hospital visits and stays, medications, hospital bills and insurance, care condition, and pain. Tweets repeated more than 5 times from the same user were deemed

XSL•FO
RenderX

irrelevant based on manual inspection and the tweet was removed from the data set. Geolocation was collected through the metadata of Twitter. Using the geolocation inference engine, we identified the latitude and longitude of tweets. We verified the accuracy of our geolocation classifier by using Amazon Mechanical Turk (MTurk) to manually curate 10,000 randomly selected tweets and found that using MTurk validated 91% (n=9100/10,000) of the inferred locations through the geolocation engine were correct (with 87%, n=8700/10,000 agreement between two MTurk curators). Only tweets with geolocation data were used for locational analysis. The detailed methods and validation of the patient experience data set procedure is documented in Multimedia Appendix 1, which has been validated in a previous study [35].

LGBTQ Users

The Twitter user descriptions were collected. A user who used any of the terms "lesbian," "gay," "bisexual," "transgender," "trans," "queer," "LGBT," "LGBTQ," "intersex," "homosexual," or "cis" in their profile description was deemed an LGBTQ user. A non-LGBTQ user was defined as any Twitter user who did not use these terms in their profile description. The control population was comprised of the user population that did not self-identify as being LGBTQ on Twitter. The LGBTQ user was defined as a binary variable: tweets from LGBTQ users were labeled with 1, and tweets from non-LGBTQ users were labeled as 0. Tweets were labeled according to the user description at the time of the tweet; changes in LGBTQ or non-LGBTQ status over time were not considered. A conservative manual inspection of 200 user profile descriptions categorized as LGBTQ agreed 81% (n=162) of the time.

State Identification

Tweet latitude and longitude information were matched with the United States Census Bureau's American Community Survey 5-year estimates to identify the corresponding state and generate a state-level field [36].

Geographic Regions

The geographic regions Northeast, Midwest, South, and West were determined by the US Census and were matched to the latitude and longitude of user's tweets. The region of the Northeast includes Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania. The Midwest includes Ohio, Michigan, Indiana, Wisconsin, Illinois, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, and Kansas. The South includes Delaware, Maryland, Virginia, West Virginia, Kentucky, North Carolina, South Carolina, Tennessee, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana, Texas, Oklahoma, and Washington, DC. The West comprises of Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, California, Oregon, Washington, Alaska, and Hawaii. Graphical map descriptions of patient experience sentiment maps of LGBTQ and non-LGBTQ users were conducted using the plotly Python (Python Software Foundation) visualization library [37].

Political Composition

Political party affiliation at the state level was retrieved through the Gallup Daily Tracking data set [22]. Political composition was defined as percent Republican and percent Democratic for each state. Democratic advantage was defined as percent Democratic minus percent Republican by each state and was included in the analysis as a continuous variable [22]. We chose to use political affiliation data from 2015 to represent the middle point of the study time frame. There may have been changes in political affiliation, but this was not measured for this study.

Dependent Variable: Sentiment of Patient Experience Tweet

Sentiment analysis has been frequently used to determine the attitude and emotion of the user (ie, author) with respect to a topic [38,39]. For instance, a user might tweet "the doctors were so knowledgeable and kind. Thank you!" that would be deemed as positive, whereas a tweet from a user that states "how could you allow patients to be treated so horribly" would be determined to possess negative sentiment. The sentiment of a patient experience tweet was defined as the attitude of the patient toward their health care experience. Sentiment analysis determines the attitude of the user by measuring the polarity of the sentiment which lies in the range of -1 to 1, where 1 is an extremely positive attitude and -1 means an extremely negative attitude. We measured the sentiment polarity using the widely accepted and used lexicon and rule-based sentiment classifier called Valence Aware Dictionary for Sentiment Reasoning (VADER) to identify the sentiment of the patient experience tweet [40]. VADER computes sentiment for each word and generates compound scores for the sentence by summing the sentiment score of each word. For VADER, a sentiment score is positive if the mean compound score is greater than or equal to 0.5 and negative if the score is less than or equal to -0.5. Mean compound scores between -0.5 and 0.5 are considered neutral. Scores of exactly 0.0 are discarded as they indicate that there is not sufficient context. We expanded on VADER's dictionary and rules to better represent the microblogging style of platforms like Twitter. This included the incorporation of *emojis* and their respective sentiment scores.

Analysis

Data analysis was conducted using the scikit-learn [41] and statsmodels [42] packages in Python. Descriptive statistics were conducted to compare the frequency of tweets by each state-level factor between LGBTQ and non-LGBTQ Twitter users. Tweets were separated into LGBTQ user and non-LGBTQ user tweets, and the geolocation was identified. The proportion of LGBTQ or non-LGBTQ users in a region was calculated by dividing the number of either of these users in that region by the total number of users in the region. The mean democratic advantage was calculated by summing the democratic advantage of each user group and dividing it by the total number of users. Chi-square tests for proportions were conducted to compare the breakdown of LGBTQ users by geographic region compared to the breakdown of non-LGBTQ users by geographic region. An ordinary least squares regression was used to model the effect of state-level democratic advantage on patient experience sentiment while controlling for geographic region. The inclusion

XSL•FO
RenderX

of an interaction term between democratic advantage and LGBTQ status allowed for LGBTQ status to moderate the effect of democratic advantage on patient experience sentiment.

Regression Model

tweet_sentiment ~ LGBTQ + democratic_advantage + LGBTQ * democratic_advantage + NE + S + W + ε

Results

The total number of users in the patient experience data set was 1,376,084 users. Out of these users, 13,689 (1.00%) self-identified as LGBTQ, and 1,362,395 (99.00%) did not self-identify as LGBTQ. The number of LGBTQ users that had available geolocation data used in these analyses was 5545, and the number of non-LGBTQ users was 445,919. The data-cleaning process removed 171 out of 22,570 LGBTQ user tweets and 15,211 out of 1,946,795 non-LGBTQ user tweets. As shown in Table 1, the highest proportion for both LGBTQ and non-LGBTQ users came from the South, followed by the West, Midwest, and Northeast. A statistically significant lower proportion of LGBTQ users compared to non-LGBTQ users were present in the areas of the South and the Midwest. Alternatively, there were significantly higher proportions of LGBTQ users compared to non-LGBTQ users in the areas of the Northeast and the West. Mean democratic advantage was also significantly higher in areas with greater proportions of LGBTQ users relative to non-LGBTQ users.

Table 1. Descriptive analysis of LGBTQ (n=5545) and non-LGBTQ (n=445,919) tweets in the United States, 2013-2017.

Variable	LGBTQ[a] users	Non-LGBTQ users	Test statistic	*P* value
Northeast, n (%)	1053 (18.99)	75,761 (16.99)	3.78	<.001
South, n (%)	1786 (32.21)	167,635 (37.59)	–8.52	<.001
Midwest, n (%)	1166 (21.03)	98,354 (22.06)	–1.87	.06
West, n (%)	1540 (27.77)	104,169 (23.36)	7.29	<.001
Democratic advantage, mean	3.92	2.60	9.44	<.001

[a]LGBTQ: lesbian, gay, bisexual, transgender, and queer.

The multilevel regression output is presented in Table 2, which illustrates the association between LGBTQ individual-level status and state-level predictors of patient experience sentiment from February 2013 to February 2017. Users in the Northeast had on average a 0.0444 (*P*<.001) lower patient experience sentiment, and users in the South had on average a 0.0390 (*P*<.001) lower patient experience sentiment compared to users in the Midwest, controlling for individual-level LGBTQ status and the state-level political factor. Users in the West had on average a 0.0188 (*P*<.001) higher patient experience sentiment compared to that of users in the Midwest while controlling for all other variables in the model. Users that self-identified as LGBTQ had on average 0.0191 (*P*=.01) lower patient experience sentiment compared to users that did not self-identify as LGBTQ in areas with no difference in democratic or republican advantage, controlling for region- and political-level factors.

Table 2. Ordinary least squares regression results demonstrating association between LGBTQ status and state-level political leaning with patient experience sentiment in the United States, 2013-2017.

Variables	Coefficient	SE	*t* statistic	*P* value	95% CI
Intercept	–0.0407	0.001	–31.163	<.001	–0.043 to –0.038
Northeast	–0.0444	0.002	–20.199	<.001	–0.049 to –0.040
South	–0.0390	0.002	–23.217	<.001	–0.042 to –0.036
Midwest (reference group)	N/A[a]	N/A	N/A	N/A	N/A
West	0.0188	0.002	9.820	<.001	0.015 to 0.022
LGBTQ[b] user	–0.0191	0.006	–3.408	.01	–0.030 to –0.008
Democratic advantage	0.0008	<0.0001	4.467	<.001	0.000 to 0.000
Democratic advantage x LGBTQ user	0.0014	0.001	2.713	.007	0.000 to 0.002

[a]Not applicable.

[b]LGBTQ: lesbian, gay, bisexual, transgender, and queer.

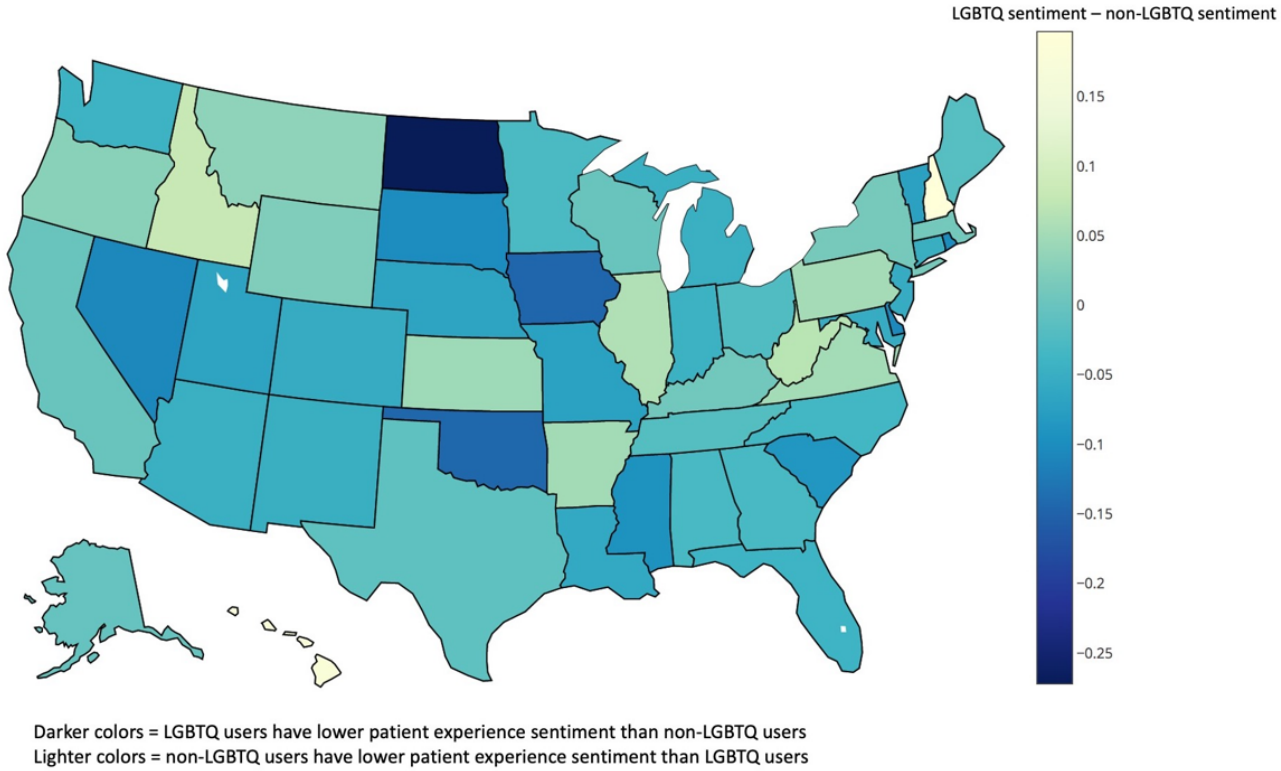
Figure 1 shows the difference in average patient experience sentiment for LGBTQ users compared to non-LGBTQ users for each US state. Darker colors represent a more negative differential in LGBTQ user patient experience sentiment compared to non-LGBTQ users, while lighter colors represent a more positive LGBTQ user sentiment compared to non-LGBTQ users. For non-LGBTQ users, each 1-point gain in state-level democratic advantage was associated with a predicted value increase of 0.0003 in patient experience sentiment, controlling for regional variation.

The effect of democratic advantage on patient experience sentiment is moderated by LGBTQ status. Democratic advantage

XSL•FO
RenderX

at the state level has a stronger effect on LGBTQ users compared to non-LGBTQ users' patient experience sentiment. For LGBTQ users, each 1-point gain in state-level democratic advantage was associated with a predicted value increase of 0.0017, controlling for all other variables in the model. The positive effect of democratic advantage on patient experience was 5.67 times greater in LGBTQ patients than non-LGBTQ patients.

Figure 1. Map of patient experience tweets. LGBTQ: lesbian, gay, bisexual, transgender, and queer. (Note: Positive sentiment scores are ≥0.5; negative sentiment score ≤-0.5; neutral sentiment scores are between ±0.5).



Discussion

Principal Findings

Our analysis highlights geographical and political factors associated with patient experience sentiment for LGBTQ persons in the United States. As documented by the online social media network Twitter, LGBTQ status was associated with worse patient experiences compared to their non-LGBTQ counterparts even after adjusting for regional variation and political stance at the state level. This is consistent with previous reports that have documented the everyday discrimination experienced by LGBTQ individuals [43]. In 2016, the Center for American Progress showed that 1 in 4 LGBTQ persons reported discrimination that compels them to make significant changes to their everyday lives to avoid that discrimination [43]. These types of changes may include health care avoidance or not sharing their LGBTQ status to medical providers or others working in the health care system. Evidence has shown strong positive associations between patient experience and clinical effectiveness for a wide range of disease areas and outcome measures [44,45]. Therefore, based on our results, poorer patient experiences among LGBTQ patients may be contributing to the disproportional burden of disease and disparity faced by LGBTQ persons [46].

Regional analysis of quality of care for the general public have indicated that areas in the South often perform more poorly on many quality measures [37]. Our results mirror such regional patterns, as we found that the South Region exhibited the worst patient experiences compared to all other regions in the United States. Findings in our study also revealed that Twitter users in the West reported the best patient experiences, which is also consistent with prior research, as this is an area that tends to perform better on the majority of quality measures [37]. However, users from the Northeast did not appear to rate their experience as well as those in the West or the Midwest. Although it has been documented that the Northeast tends to have a higher health care quality, many communities in the Northeast also have higher costs and elevated rates of avoidable hospital use, which could explain the poorer patient experience feedback reported among Twitter users in this area [36,37].

Based on our study, political values at the state level may influence the sentiment of individual-level patient experiences for LGBTQ persons more so than for non-LGBTQ persons. The political climate of a geographic area has been shown to be associated with health-related outcomes and access to care among LGBTQ persons [21,23]. However, prior studies have only focused on specific outcomes such as depression or on subsets within the LGBTQ community. For instance, one study identified that the percent of state residents voting republican was the strongest and most significant state-level predictor of health care refusal among transgender individuals [21]. Our results showed that an increasing democratic advantage at the state level contributed to better patient experiences reported at

XSL•FO
RenderX

the individual level for all users even after controlling for regional variation. Furthermore, this state-level political factor had a 5.67 times stronger positive effect on patient experience sentiment for LGBTQ users compared to non-LGBTQ users. These macrolevel political affiliations based on voting behavior can potentially serve as proxies for local attitudes and culture toward LGBTQ persons, and influence structural stigmatization in health care [21,47]. It has been demonstrated that Republican-identified voters are more likely to hold discriminatory attitudes toward LGBTQ persons compared to Democratic voters [47]. Therefore, LGBTQ persons living in states with more republican voters may be more likely to encounter and receive poorer quality care from biased providers or health care systems.

Limitations

There are limitations in this study that should be noted when interpreting our findings. First, our results examined state-level factors that were associated with patient experience sentiment and cannot be interpreted as causal. Second, the keywords methodology used to identify LGBTQ users may have incorrectly placed some LGBTQ users in the non-LGBTQ group and vice versa. The keywords list is not comprehensive and may have missed some LGBTQ-identifying descriptors such as shorthand phrases. Although we were able to manually confirm that 80-90% of LGBTQ categorized users were indeed LGBTQ users, we were not able to place a number on the accuracy of non-LGBTQ categorization. However, we believe that the percentage of LGBTQ users in the control group would not be more than the national percentage of 4.5% [48], which still provides a reasonable and valuable control population for this study. Third, there are limitations with the geo-tagged Twitter data set. Only 15% of online adults regularly use Twitter, with those 18-29 years of age being most represented. Only approximately 1% of users geo-tag the majority of the tweets they post [49,50]. Therefore, the sample population is likely not representative of the US patient population. Despite this potential limitation, many patient experience surveys have reported low response rates that range from 20-30% and have a greater proportion of older adult and female participants [51]. Our study may have a better representation of a demographic of patient participants that are normally excluded from mainstream or traditional surveys. For instance, Twitter users are generally more educated, younger, male [52], and have overrepresentation of ethnic minorities including blacks and Hispanics to a greater extent [48]. For our analysis on sentiment, previous studies have shown that the time of day, the ordering of activities within a day, and location can influence the sentiment of an individual's tweet [49,53]. We did not control for these temporal and spatial differences, which may have influenced the results of our study.

In the past, research on LGBTQ populations has been sparse due to methodological limitations or because of issues related to homophobia (fear of people who are attracted to members of the same sex) and heterosexism (discrimination against homosexuals because of the assumption that heterosexuality is the norm of sexual orientation) [18]. Most national and state surveys lack appropriate questions pertaining to gender and sexual identity, making it difficult to conduct large-scale research [46]. Even large-scale studies typically identify few LGBTQ individuals, which has meant that LGBTQ research must rely on smaller qualitative studies or that investigation is lacking altogether [18]. Our study is the first to use a novel online social media data set captured from the popular Twitter platform to investigate LGBTQ patient experiences across the United States. Research shows that LGBTQ persons are more open to coming out and defining their LGBTQ status in their online social networks when compared to offline networks [54,55]. In fact, Twitter is a highly popular social media space for LGBTQ users when compared to other social media networks such as Facebook [55]. Finally, LGBTQ persons are more likely to search for health and medical information online when compared to their non-LGBTQ counterparts (81% vs 46%) [56]. Since LGBTQ persons face significant discrimination in typical health care settings [8], LGBTQ persons are less likely to voluntarily identify themselves as LGBTQ in the health care setting or in response to health surveys. This lack of identification prevents researchers from effectively capturing LGBTQ patient experiences. However, through our online big data set, we were able to identify an extensive number of self-identified LGBTQ users on Twitter. Based on this evidence, even though not all LGBTQ persons may have identified themselves as LGBTQ on Twitter, our study was able to capture novel insights into LGBTQ patient experiences that may not have been documented in the past.

Conclusion

The Institute of Medicine recognizes that LGBTQ persons experience a disproportionate burden of disease and poorer health outcomes compared to the general population, and there is growing recognition of the need to further study these disparities to identify the factors that contribute to them [57]. This is the first study to leverage an online social media data set to characterize the patient experiences of LGBTQ persons. This study demonstrated that LGBTQ users experience worse patient experiences compared to non-LGBTQ users and that the political climate of a state determined through voting percentages is a prominent factor influencing patient experiences, especially for LGBTQ persons. This is especially relevant given recent emphasis on how different US government administrations can directly impact policy decisions regarding equitable health care for LGBTQ persons [58]. By identifying the factors that impact patient experience, researchers, health care providers, and policy makers can begin to develop targeted practices and policies that improve health equity for LGBTQ persons and other marginalized groups.

Acknowledgments

This study was funded by the Robert Wood Johnson Foundation Grant 73495 (to YH, JBH, KCS, and JSB); NIH/National Human Genome Research Institute Grant 5U54HG007963-04 (to JSB and JBH); and the Canadian Institutes of Health Research (to YH).

XSL•FO
RenderX

The funders played no role in the study design; collection, analysis, or interpretation of data; writing of the manuscript; or decision to submit the manuscript for publication

Conflicts of Interest

None declared.

Multimedia Appendix 1
Lesbian, gay, bisexual, transgender, and queer patient experiences.
[DOCX File , 1029 KB - jmir_v22i7e17087_app1.docx]

References

1. McKay B. Lesbian, gay, bisexual, and transgender health issues, disparities, and information resources. Med Ref Serv Q 2011;30(4):393-401. [doi: 10.1080/02763869.2011.608971] [Medline: 22040245]
2. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195. [doi: 10.1161/circoutcomes.109.900597]
3. Isaac T, Zaslavsky A, Cleary P, Landon B. The relationship between patients' perception of care and measures of hospital quality and safety. Health Serv Res 2010 Aug;45(4):1024-1040 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01122.x] [Medline: 20528990]
4. Meterko M, Wright S, Lin H, Lowy E, Cleary P. Mortality among patients with acute myocardial infarction: the influences of patient-centered care and evidence-based medicine. Health Serv Res 2010 Oct;45(5 Pt 1):1188-1204 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01138.x] [Medline: 20662947]
5. Willging CE, Salvador M, Kano M. Brief reports: Unequal treatment: mental health care for sexual and gender minority groups in a rural state. Psychiatr Serv 2006 Jun;57(6):867-870. [doi: 10.1176/ps.2006.57.6.867] [Medline: 16754766]
6. Grant JM, Mottet LA, Tanis J, Herman JL, Harrison J, Keisling M. National transgender discrimination survey report on health and health care. National Center for Transgender Equality and the National Gay and Lesbian Task Force 2010:1-220.
7. Lindsey H. Overcoming barriers to cancer care for LGBT patients. Oncol Times 2014;36(10):95-96. [doi: 10.1097/01.cot.0000450344.88617.87]
8. Legal L. When health care isn't caring: Lambda Legal's survey of discrimination against LGBT people and people with HIV. New York: Lambda Legal 2010:1-26. [doi: 10.4135/9781483371283.n230]
9. Koester KA, Collins SP, Fuller SM, Galindo GR, Gibson S, Steward WT. Sexual healthcare preferences among gay and bisexual men: a qualitative study in San Francisco, California. PLoS One 2013;8(8):e71546 [FREE Full text] [doi: 10.1371/journal.pone.0071546] [Medline: 23977073]
10. Platzer H, James T. Lesbians' experiences of healthcare. NT Research 2016 Aug 18;5(3):194-202. [doi: 10.1177/136140960000500305]
11. Hayman B, Wilkes L, Halcomb E, Jackson D. Marginalised mothers: lesbian women negotiating heteronormative healthcare services. Contemp Nurse 2013 Feb 25:2410-2429. [doi: 10.5172/conu.2013.2410]
12. Malebranche DJ, Peterson JL, Fullilove RE, Stackhouse RW. Race and sexual identity: perceptions about medical culture and healthcare among Black men who have sex with men. J Natl Med Assoc 2004 Jan;96(1):97-107. [Medline: 14746359]
13. Institute of Medicine. The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding. Washington, DC: The National Academies Press; 2011.
14. Rentfrow PJ, Gosling SD, Jokela M, Stillwell DJ, Kosinski M, Potter J. Divided we stand: three psychological regions of the United States and their political, economic, social, and health correlates. J Pers Soc Psychol 2013 Dec;105(6):996-1012. [doi: 10.1037/a0034434] [Medline: 24128185]
15. Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States. N Engl J Med 1993 Mar 04;328(9):621-627. [doi: 10.1056/nejm199303043280906]
16. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/nejmp1211775]
17. Baicker K, Chandra A, Skinner J. Geographic variation in health care and the problem of measuring racial disparities. Perspect Biol Med 2005;48(1):42-S53. [doi: 10.1353/pbm.2005.0020]
18. Meyer IH. Why lesbian, gay, bisexual, and transgender public health? Am J Public Health 2001 Jun;91(6):856-859. [doi: 10.2105/ajph.91.6.856] [Medline: 11392921]
19. Legislation affecting LGBT rights across the country. ACLU. 2018. URL: https://www.aclu.org/other/legislation-affecting-lgbt-rights-across-country [accessed 2020-03-20]
20. Law T. Policy Institute and National Gay and Lesbian Task Force: scope of explicitly transgender-inclusive anti-discrimination laws. Center for American Progress 2016 Dec 08:1. [doi: 10.4135/9781412956215.n602]

XSL•FO
RenderX

21. White Hughto JM, Murchison GR, Clark K, Pachankis JE, Reisner SL. Geographic and individual differences in healthcare access for U.S. transgender adults: a multilevel analysis. LGBT Health 2016 Dec;3(6):424-433 [FREE Full text] [doi: 10.1089/lgbt.2016.0044] [Medline: 27636030]
22. Jones JM. Red states outnumber blue for first time in gallup tracking. Gallup. 2016 Feb 03. URL: http://news.gallup.com/poll/188969/red-states-outnumber-blue-first-time-gallup-tracking.aspx [accessed 2018-05-02]
23. Hatzenbuehler ML, Keyes KM. Inclusive anti-bullying policies and reduced risk of suicide attempts in lesbian and gay youth. J Adolesc Health 2013 Jul;53(1 Suppl):S21-S26. [doi: 10.1016/j.jadohealth.2012.08.010] [Medline: 23790196]
24. Perez-Brumer A, Hatzenbuehler ML, Oldenburg CE, Bockting W. Individual- and structural-level risk factors for suicide attempts among transgender adults. Behav Med 2015;41(3):164-171 [FREE Full text] [doi: 10.1080/08964289.2015.1028322] [Medline: 26287284]
25. Everett BG. Changes in neighborhood characteristics and depression among sexual minority young adults. J Am Psychiatr Nurses Assoc 2014;20(1):42-52 [FREE Full text] [doi: 10.1177/1078390313510319] [Medline: 24217448]
26. Hersh ED, Goldenberg MN. Democratic and Republican physicians provide different care on politicized health issues. Proc Natl Acad Sci U S A 2016 Oct 18;113(42):11811-11816 [FREE Full text] [doi: 10.1073/pnas.1606609113] [Medline: 27698126]
27. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Jun;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]
28. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. N Engl J Med 2008 Oct 30;359(18):1921-1931. [doi: 10.1056/nejmsa0804116]
29. Zimlichman E, Rozenblum R, Millenson ML. The road to patient experience of care measurement: lessons from the United States. Isr J Health Policy Res 2013 Sep 17;2(1):35 [FREE Full text] [doi: 10.1186/2045-4015-2-35] [Medline: 24044672]
30. Hswen Y, Sewalk KC, Alsentzer E, Tuli G, Brownstein JS, Hawkins JB. Investigating inequities in hospital care among lesbian, gay, bisexual, and transgender (LGBT) individuals using social media. Soc Sci Med 2018 Oct;215:92-97. [doi: 10.1016/j.socscimed.2018.08.031] [Medline: 30219749]
31. Krueger EA, Young SD. Twitter: a novel tool for studying the health and social needs of transgender communities. JMIR Ment Health 2015;2(2):e16 [FREE Full text] [doi: 10.2196/mental.4113] [Medline: 26082941]
32. Webb F, Karami A, Kitzie V. Characterizing diseases and disorders in gay users' tweets. arXiv 2018:180309134.
33. Mozumder MK. Validation of Bengali perceived stress scale among LGBT population. BMC Psychiatry 2017 Aug 29;17(1):314 [FREE Full text] [doi: 10.1186/s12888-017-1482-0] [Medline: 28851332]
34. Whitehead J, Shaver J, Stephenson R. utness, stigma, and primary health care utilization among rural LGBT populations. PLoS One 2016;11(1):e0146139 [FREE Full text] [doi: 10.1371/journal.pone.0146139] [Medline: 26731405]
35. Sewalk KC, Tuli G, Hswen Y, Brownstein JS, Hawkins JB. Using Twitter to examine web-based patient experience sentiments in the United States: longitudinal study. J Med Internet Res 2018 Oct 12;20(10):e10043 [FREE Full text] [doi: 10.2196/10043] [Medline: 30314959]
36. Radley D, How S, Fryer A, McCarthy D, Schoen C. Rising to the Challenge: Results From a Scorecard on Local Health System Performance. New York: Commonwealth Fund; 2012.
37. Edmunds M, Sloan F, Steinwald A. Evidence of geographic variation in access, quality, and workforce distribution. Committee on Geographic Adjustment Factors in Medicare Payment; Board on Health Care Services; Institute of Medicine 2005 Oct 13:2011.
38. Pang B, Lee L. Opinion mining and sentiment analysis. FNT in Information Retrieval 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
39. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. 2005 Oct 13 Presented at: Conference on Human Language Technology and Empirical Methods in Natural Language Processing; 2005; Vancouver. [doi: 10.3115/1220575.1220619]
40. Gilbert C. Vader: a parsimonious rule-based model for sentiment analysis of social media text. 2014 May 16 Presented at: Eighth International Conference on Weblogs and Social Media (ICWSM-14); 2014; Atlanta p. A URL: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf
41. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp and Comm 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]
42. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. 2010 Jun 28 Presented at: 9th Python in Science Conference; 2010; Austin, Texas p. E. [doi: 10.25080/majora-92bf1922-011]
43. Caiazza T. CAP study shows one in four LGBT people faced discrimination last year, many make significant changes to their lives to avoid discrimination. Center for American Progress 2018 Jul 01:2016. [doi: 10.1093/acref/9780195301731.013.51090]
44. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 03;3(1):1 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
45. Tsai TC, Orav EJ, Jha AK. Patient satisfaction and quality of surgical care in US hospitals. Ann Surg 2015;261(1):2-8. [doi: 10.1097/sla.0000000000000765]

XSL•FO
RenderX

46. Hafeez H, Zeshan M, Tahir M, Jahan N, Naveed S. Health care disparities among lesbian, gay, bisexual, and transgender youth: a literature review. Cureus 2017 Apr 20;9(4):e1184 [FREE Full text] [doi: 10.7759/cureus.1184] [Medline: 28638747]
47. Kimmel MS, Mahler M. Adolescent masculinity, homophobia, and violence. Am Behav Scientist 2016 Jul 27;46(10):1439-1458. [doi: 10.1177/0002764203046010010]
48. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the demographics of Twitter users. ICWSM (5th) 2011;11:25.
49. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
50. Smith A, Brenner J. Twitter use 2012. Pew Internet & American Life Project 2012;4:1-12.
51. Boscardin CK, Gonzales R. The impact of demographic characteristics on nonresponse in an ambulatory patient satisfaction survey. Joint Commission J Quality Patient Safety 2013 Mar;39(3):123-AP4. [doi: 10.1016/s1553-7250(13)39018-7]
52. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Social media update 2014. Pew research center 2015;19:1-12.
53. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
54. Brammer J. LGBTQ and out on social media — but nowhere else. NBC News. 2018. URL: https://www.nbcnews.com/feature/nbc-out/lgbtq-out-social-media-nowhere-else-n809796 [accessed 2018-05-20]
55. Craig SL, McInroy L. You can form a part of yourself online: the influence of new media on identity development and coming out for LGBTQ youth. J Gay Lesbian Ment Health 2014 Jan 30;18(1):95-109. [doi: 10.1080/19359705.2013.777007]
56. Marra A. Out online: the experiences of LGBT youth on the internet. GLSEN. 2013 Jul 10. URL: https://www.glsen.org/press/study-finds-lgbt-youth-face-greater-harassment-online [accessed 2018-05-10]
57. Committee N. Consideration of the institute of medicine (IOM) report on the health of lesbian, gay, bisexual, and transgender (LGBT) individuals. Washington DC: National Institutes of Health 2013:1-24. [doi: 10.1037/e561602013-001]
58. Gonzales G, McKay T. What an emerging Trump administration means for lesbian, gay, bisexual, and transgender health. Health Equity 2017;1(1):83-86 [FREE Full text] [doi: 10.1089/heq.2017.0002] [Medline: 30283837]

Abbreviations

LGBTQ: lesbian, gay, bisexual, transgender, and queer
MTurk: Amazon Mechanical Turk
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by G Eysenbach; submitted 16.11.19; peer-reviewed by R Gore, A Cyr, Z Ben-Miled; comments to author 06.03.20; revised version received 25.04.20; accepted 26.04.20; published 31.07.20.

Please cite as:
Hswen Y, Zhang A, Sewalk KC, Tuli G, Brownstein JS, Hawkins JB
Investigation of Geographic and Macrolevel Variations in LGBTQ Patient Experiences: Longitudinal Social Media Analysis
J Med Internet Res 2020;22(7):e17087
URL: http://www.jmir.org/2020/7/e17087/
doi:10.2196/17087
PMID:33137713

©Yulin Hswen, Amanda Zhang, Kara C Sewalk, Gaurav Tuli, John S Brownstein, Jared B Hawkins. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effects of Telemonitoring on Patient Compliance With Self-Management Recommendations and Outcomes of the Innovative Telemonitoring Enhanced Care Program for Chronic Heart Failure: Randomized Controlled Trial

Hang Ding[1,2,3], PhD; Rajiv Jayasena[4], PhD; Sheau Huey Chen[5], PGCert; Andrew Maiorana[6,7], PhD; Alison Dowling[8], MPH; Jamie Layland[9], MBChB, MD, PhD; Norm Good[1], MSc; Mohanraj Karunanithi[1], BE, MBiomeE, PhD; Iain Edwards[8], MBA

[1]RECOVER Injury Research Centre, Faculty of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia
[2]The Australian e-Health Research Centre, Commonwealth Scientific & Industrial Research Organisation Health & Biosecurity, Brisbane, Australia
[3]Faculty of Medicine, The University of Queensland, Brisbane, Australia
[4]The Australian e-Health Research Centre, Commonwealth Scientific & Industrial Research Organisation Health & Biosecurity, Melbourne, Australia
[5]School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia
[6]School of Physiotherapy and Exercise Science, Australia Faculty of Health Sciences, Curtin University, Perth, Australia
[7]Allied Health Department and Advanced Heart Failure and Cardiac Transplant Service, Fiona Stanley Hospital, Perth, Australia
[8]Department of Community Health, Peninsula Health, Melbourne, Australia
[9]Department of Cardiology, Peninsula Health, Melbourne, Australia

Corresponding Author:
Rajiv Jayasena, PhD
The Australian e-Health Research Centre
Commonwealth Scientific & Industrial Research Organisation Health & Biosecurity
343 Royal Parade
Parkville VIC 3052
Melbourne, 3052
Australia
Phone: 61 416129871
Email: Rajiv.Jayasena@csiro.au

Abstract

Background: Telemonitoring enables care providers to remotely support outpatients in self-managing chronic heart failure (CHF), but the objective assessment of patient compliance with self-management recommendations has seldom been studied.

Objective: This study aimed to evaluate patient compliance with self-management recommendations of an innovative telemonitoring enhanced care program for CHF (ITEC-CHF).

Methods: We conducted a multicenter randomized controlled trial with a 6-month follow-up. The ITEC-CHF program comprised the provision of Bluetooth-enabled scales linked to a call center and nurse care services to assist participants with weight monitoring compliance. Compliance was defined a priori as weighing at least 4 days per week, analyzed objectively from weight recordings on the scales. The intention-to-treat principle was used to perform the analysis.

Results: A total of 184 participants (141/184, 76.6% male), with a mean age of 70.1 (SD 12.3) years, were randomized to receive either ITEC-CHF (n=91) or usual care (control; n=93), of which 67 ITEC-CHF and 81 control participants completed the intervention. For the compliance criterion of weighing at least 4 days per week, the proportion of compliant participants in the ITEC-CHF group was not significantly higher than that in the control group (ITEC-CHF: 67/91, 74% vs control: 56/91, 60%; *P*=.06). However, the proportion of ITEC-CHF participants achieving the stricter compliance standard of at least 6 days a week was significantly higher than that in the control group (ITEC-CHF: 41/91, 45% vs control: 23/93, 25%; *P*=.005).

Conclusions: ITEC-CHF improved participant compliance with weight monitoring, although the withdrawal rate was high. Telemonitoring is a promising method for supporting both patients and clinicians in the management of CHF. However, further refinements are required to optimize this model of care.

XSL•FO
RenderX

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12614000916640; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=366691

(J Med Internet Res 2020;22(7):e17559) doi:10.2196/17559

KEYWORDS

heart failure; digital health; telemonitoring; remote monitoring; patient compliance; randomized controlled trial

Introduction

Chronic heart failure (CHF) is a severe chronic disease that affects more than 26 million people worldwide [1]. It significantly reduces the health-related quality of life and increases the risk of hospitalization and mortality [1]. To improve health outcomes, it is recommended that patients with CHF undertake self-management, such as daily monitoring of body weight to assess fluid balance and seek early clinical support in the event of symptoms, which may indicate decompensation. This has been consistently outlined by evidence-based clinical guidelines for CHF [2,3] and is practically supported by CHF clinics and rehabilitation programs in standard care. Despite these clinical efforts, patient compliance with self-management recommendations is often suboptimal for activities such as body weight recording, fluid restrictions, and medication adherence [4]. Time constraints [5], limited knowledge [6], and insufficient ongoing clinical support [7] are some of the reported barriers to the self-management of CHF. Poor compliance with self-management recommendations often leads to delays in essential treatment [4] and increases the risk of mortality and hospitalization [8].

In recent years, there has been significant research interest in telemonitoring as an innovative approach to remotely assist patients with CHF in self-managing their health [9]. However, to date, only 2 studies, to our knowledge, have evaluated patient compliance with weight monitoring in a randomized controlled trial (RCT) [10,11]. Although they demonstrated a higher rate of *compliant* participants in the telemonitoring arm (telemonitoring vs usual care: 88.6% vs 70.9% [10] and 91.7% vs 67.4% [11]), the studies relied on self-report, which is known to be influenced by recall bias [12]. In addition, the definition of *compliance* was loosely defined based on terms such as *most of the time* or *all of the time* and, hence, was not sufficiently accurate to reflect the daily weight monitoring recommendation. Moreover, patient adherence to telemonitoring systems has often been found to be low, even in large, well-designed RCTs (55% [13] and 55.4% [14]). This has led to an ongoing debate about the practicality of using telemonitoring to improve CHF care [13-15]. Therefore, further rigorous research for evaluating patient compliance is needed in telemonitoring studies for CHF care.

We evaluated an innovative telemonitoring enhanced care program for CHF (ITEC-CHF) in an open multicenter RCT. The ITEC-CHF program focused on assisting patients in daily weight monitoring and engaging with nurse-supported care in the event of weight fluctuations. This study aimed to examine whether the ITEC-CHF program improved patient compliance with weight monitoring as well as other self-management behaviors and health outcomes.

Methods

Study Design

The protocol for the ITEC-CHF study has been previously published [15]. Images of the user interface and the Bluetooth-enabled scales are provided in Multimedia Appendices 1 and 2. In this study, patients with CHF were recruited from 2 trial sites in Australia: one in Victoria (VIC) and one in Western Australia (WA). The trial sites were at 2 hospitals in VIC and WA, respectively. This study complies with the Declaration of Helsinki. All participants provided written informed consent. The clinical trial protocol was approved by the Human Research Ethics Committee at Peninsula Health, VIC (HREC reference: HREC/14/PH/27), and Royal Perth Hospital, WA (reference: 15-081 and reference: HR 181/2014), Australia. Participants were enrolled from January 2015 to October 2017. The latest data collection of hospitalizations and emergency department (ED) presentations was conducted in September 2018.

Randomization and Masking

Participants in the trial were individually randomized with an allocation ratio of 1:1 to receive either ITEC-CHF or usual care (control) for 6 months. Randomization was stratified by the 2 trial sites (VIC and WA) to ensure that the allocation ratio was consistent at each site. A block method was used to achieve a balanced number of participants between the ITEC-CHF and control groups throughout the trial. The random allocation assignments were sealed in opaque envelopes. Data analysts generated the randomization sequence and were blinded to the trial because of the use of deidentified patient data.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: patients (1) with CHF with reduced ejection fraction (EF; ie, EF≤40%), (2) able to weigh themselves safely, (3) aged at least 18 years, (4) having a regular personal general practitioner (GP) or agreeing to use a designated GP, (5) with a permanent residential address, and (6) without significant cognitive impairments. The exclusion criteria were as follows: (1) patients with expected survival <12 months, (2) patients with end-stage renal failure on dialysis, (3) long-term nursing home residents, or (4) patients participating in any other clinical trial.

Interventions

At baseline, control participants were provided with a standard package of a paper-based diary and the *Living Well with Chronic Heart Failure* booklet, produced by the Heart Foundation of

XSL•FO
RenderX

Australia. They were instructed to maintain their usual CHF care, as provided by clinics specialized in CHF and primary care physicians, and to undertake CHF self-management as previously instructed. Each participant was also provided with an electronic weight scale (FORA TN'G W550; ForaCare) and asked to use the scale to measure their body weight daily, immediately after waking, following voiding, without shoes, in light clothing, and before taking medication. Approximately every 3 months, project nurses visited the participants to download the weight entries from the weight scale.

Participants in the ITEC-CHF group received the same resources as those in the control group, in addition to the ITEC-CHF program. The ITEC-CHF consists of 3 major components: remote body weight monitoring, structured telephone support, and nurse-led collaborative care. The service was integrated with a telephone call center (MEPACS, VIC, Australia) and community nurse care services at the trial sites in VIC and WA. Each participant was provided with an electronic weight scale and a computer tablet (Galaxy Tab A, Samsung). Participants were asked to weigh themselves using the procedure described for the control group. After the measurement, the weight entry was automatically transmitted from the weight scale to the tablet via a wireless Bluetooth function. The tablet was preloaded with an Android app (Medtech Global). This app received the weight entry and uploaded the entry to a software package called Manage My Health (MMH; Medtech Global). A rule-based decision support system (a web app) in MMH automatically monitored the uploaded weight entries in real time. In response to the weight data generated by telemonitoring, 6 types of alerts were possible: (1) *rapid weight fluctuation* (an increase or decrease of 2 kg over 2 days), (2) *slow weight fluctuation* (an increase or decrease of 5 kg over 28 days), (3) *low-risk weight fluctuation* (an increase or decrease of 1 kg over 24 hours), (4) missed weight measurement, (5) low level of tablet battery, and (6) tablet connection lost. In the event of a *rapid weight fluctuation*, project nurses were alerted, and they called the participant to assist him or her in assessing symptoms and activating their CHF action plan, such as attending their GP, visiting a CHF clinic, or presenting to an ED as indicated. For an alert of *slow weight fluctuations*, the project nurses assisted the participants in assessing CHF symptoms and arranging clinical reviews at the participants' GP or CHF clinics as indicated. For *low-risk fluctuations*, a questionnaire was automatically triggered and sent to the participant's computer tablet to help him or her determine the need for further clinical follow-up. Finally, the generated alerts were distributed to project nurses and/or the MEPACS call center in the ITEC-CHF program. Call operators at the center responded to the alerts in real time (24 hours, 7 days a week), focusing on reminding participants to weigh themselves if they had not done so before 10 AM, helping assess CHF symptoms and manage diet, and arranging a nurse follow-up if needed. The project nurses reviewed their alert requests on weekdays and followed up with the participants via a telephone call. Some participants were unable to monitor their body weight for a short period, such as when they were hospitalized for medical treatment, traveled away from home, or experienced unresolved technical issues. Under such conditions, they were required to notify the call center, and weight monitoring was skipped. If a participant notified the call center to skip the monitoring for a period, the telemonitoring intervention was then *switched off* during the skipped period, and the call center did not receive any alerts from the participant and provide intervention until the skipped period ended. Monitoring days that were *skipped* were still included in the per-protocol analysis for the ITEC-CHF group (described in the Primary and Secondary Outcomes section).

Primary and Secondary Outcomes

The primary outcome was in compliance with weight monitoring. The monitoring frequency was calculated as average weight monitoring days per week during the 6-month assessment period (monitoring frequency=weight monitoring days/180 days×7 days/week [for 6 months: 6 months×30 days/month=180 days]). A weight monitoring day was determined if at least one weight entry was practically recorded on the weight scale on that day, irrespective of time. In total, 2 frequencies were employed in the examination. One was that the participant monitored their weight on at least four days per week. This frequency reflects the compliance threshold of *most of the time*, as previously applied in questionnaire-based assessments [4,12]. The other frequency was at least 6 days per week, which more closely aligns with the advice for patients to monitor their weight *daily*.

Secondary outcomes included patient compliance with weight monitoring based on a per-protocol analysis (only undertaken in participants who completed the trial) and an analysis of other guideline recommendations assessed by the Heart Failure Compliance Questionnaire [12], health-related quality of life (five-dimension EuroQol, EQ-5D [16]), 6-min walk test distance [17], psychological state (cardiac depression scale short form 2 [18]), frailty (clinical frailty index [19]), and clinical outcomes of CHF-related and all-cause hospitalizations and ED presentations. CHF-related events were determined by using the International Classification of Diseases, Tenth Revision, Clinical Modification, diagnosis codes (Multimedia Appendix 3) [20]. Furthermore, we reported the alerts provided in ITEC-CHF and days when ITEC-CHF participants requested to *skip* weight monitoring.

Statistical Analysis

In the evaluation, a chi-square test was applied to analyze continuous variables such as age, and a Fisher exact test was used to compare categorical variables such as sex and subgroups of participants under a given weight monitoring frequency (participants who achieved a given monitoring frequency vs participants who did not achieve the monitoring frequency). The Wilcoxon signed-rank test was used to compare compliance scores from questionnaire-based assessments. An analysis of covariance model [21] was used to evaluate the improvement or change in the outcome variables between the 2 groups with an adjustment for baseline. The Andersen-Gill model [22] with an adjustment for sex and age was used to analyze the hazard of hospitalizations and ED presentations. The 95% CI was estimated for the hazard function in each group. A P value<.05 was considered statistically significant. The analysis was conducted using RStudio version 1.1.383 (RStudio Inc) [23] with the R package of *survival* version 2.43-3. The intention-to-treat principle was applied to the analysis of the

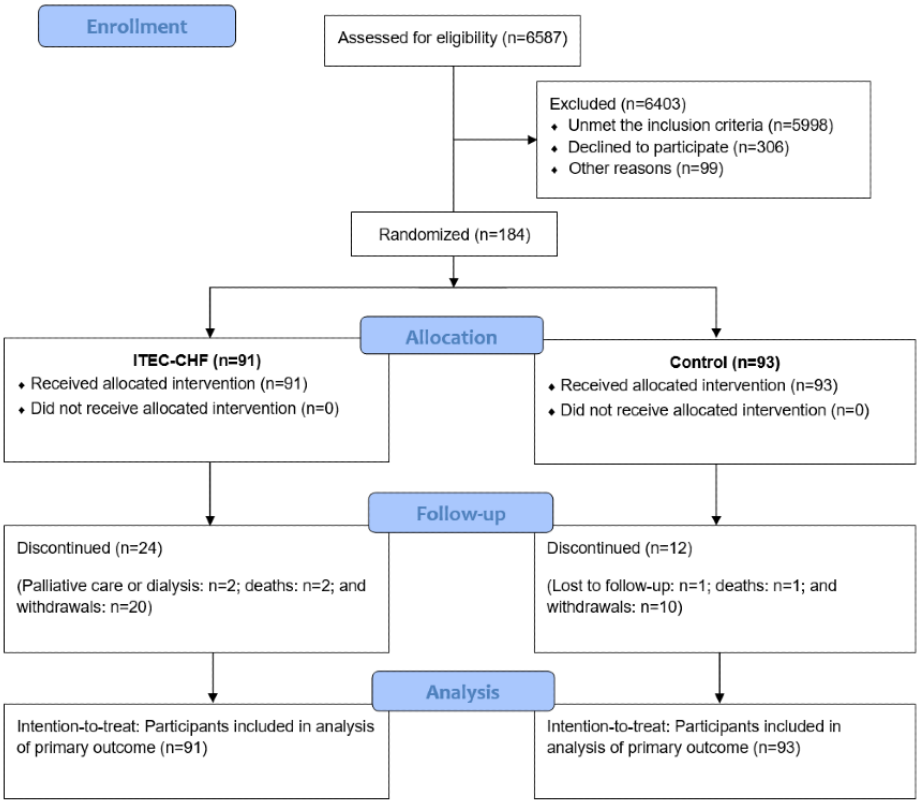
XSL•FO
RenderX

primary outcome of patient compliance with weight monitoring. It was also applied to the analysis of the hazards of hospitalization and ED. In the intention-to-treat analysis, all participants in the RCT were included. A per-protocol analysis was also applied to weight monitoring as a secondary outcome, only including participants who did not discontinue in the trial. A complete case analysis, which restricts the analysis to individuals with complete data, was used to analyze improvements in questionnaire-based assessments and 6-min walk distances.

Results

A total of 6587 patients were screened for eligibility. Among them, 6403 patients were excluded because of failure to meet the inclusion criteria (n=5998), declined (n=306), or for other reasons (n=99), such as losing contact with the patient (Figure 1). Finally, 184 patients were randomized to the ITEC-CHF (n=91) and control (n=93) groups. During the 6-month intervention period, 24 participants in the ITEC-CHF group discontinued (palliative care or dialysis: n=2; deaths: n=2; and withdrawals: n=20), and 12 participants in the control group discontinued (lost to follow-up: n=1; deaths: n=1; and withdrawals: n=11). According to the intention-to-treat principle, all randomized participants (ITEC-CHF: n=91; control: n=93) were included in the analysis of the primary outcomes of patient compliance with weighing. They were also included in the analysis of the hazards of hospitalization and/or ED presentation.

Figure 1. Patient enrollment and disposition. ITEC-CHF: innovative telemonitoring enhanced care program for chronic heart failure.



Baseline Characteristics of the Participants

There were no significant differences between the characteristics of the ITEC-CHF and control groups at baseline (Table 1). The mean ages of the participants in the ITEC-CHF and control groups were 69.5 (SD 12.3) years and 70.8 (SD 12.4) years, respectively. Participants were predominantly male, and a high proportion of participants were diagnosed with type 2 diabetes, chronic obstructive pulmonary disease, or asthma. Common medications included angiotensin-converting enzyme inhibitors, beta-blockers, loop diuretics, and/or aldosterone receptor antagonists.

XSL•FO
RenderX

Table 1. Patient baseline characteristics.

Characteristic[a]	Victoria		Western Australia		Total	
	ITEC-CHF[b] (n=42)	Control (n=42)	ITEC-CHF (n=49)	Control (n=51)	ITEC-CHF (n=91)	Control (n=93)
Age (years), mean (SD)	69.8 (13.4)	69.6 (11.7)	69.2 (11.5)	71.8 (13.0)	69.5 (12.3)	70.8 (12.4)
Gender, n (%)						
Male	28 (67)	36 (86)	38 (78)	39 (77)	66 (73)	75 (81)
Weight (kg), mean (SD)	87.7 (24.1)	86.9 (19.1)	88.6 (18.3)	83.3 (19.4)	88.2 (21.0)	84.9 (19.3)
BMI (kg/m^2), mean (SD)	31.2 (8.9)	30.2 (9.5)	31.6 (10.2)	29.1 (6.6)	31.4 (9.6)	29.6 (8.0)
NYHA[c] class, mean (SD)	1.9 (0.6)	2.0 (0.6)	2.1 (0.4)	2.3 (0.5)	2.0 (0.5)	2.2 (0.6)
LVEF[d] (%), mean (SD)	29.4 (6.5)	25.5 (23.7)	29.0 (7.5)	28.6 (7.9)	29.1 (7.1)	27.4 (15.9)
Chronic condition, n (%)						
Type 1 diabetes	1 (2)	3 (7)	0 (0)	1 (2)	1 (1)	4 (4)
Type 2 diabetes	9 (21)	19 (45)	19 (39)	16 (31)	28 (31)	35 (38)
COPD[e] or asthma	11 (26)	5 (12)	12 (25)	15 (29)	23 (25)	20 (22)
Chronic renal disease	4 (10)	10 (24)	6 (12)	10 (20)	10 (11)	20 (22)
Medical treatment, n (%)						
ACEI[f]	24 (57)	24 (57)	45 (92)	49 (96)	69 (76)	73 (79)
Beta-blockers	30 (71)	38 (91)	46 (94)	47 (92)	76 (84)	85 (91)
Digoxin	5 (12)	6 (14)	10 (20)	8 (16)	15 (17)	14 (15)
Loop diuretic	28 (67)	26 (62)	42 (86)	48 (94)	70 (77)	74 (80)
Aldosterone receptor antagonist	18 (43)	17 (41)	32 (65)	39 (77)	50 (55)	56 (60)

[a]There were no statistical differences between the characteristics of ITEC-CHF and control.

[b]ITEC-CHF: innovative telemonitoring enhanced care program for chronic heart failure.

[c]NYHA: New York Heart Association Functional Classification.

[d]LVEF: left ventricular ejection fraction.

[e]COPD: chronic obstructive pulmonary disease.

[f]ACEI: angiotensin-converting enzyme inhibitor.

Skipped Monitoring Days in the Innovative Telemonitoring Enhanced Care Program for Chronic Heart Failure

There were 1312 participant monitoring days when weight monitoring was *skipped*. Being away from home or traveling was the major reason for skipped monitoring, occurring on 515 participant monitoring days (39.3% of total skipped days) in 41 participants. Technical issues (Table 2) resulted in skipped monitoring on 390 participant monitoring days (29.7%) in 27 participants. Hospitalizations and ED presentations caused skipped monitoring on 232 participant monitoring days (17.7%) in 18 participants, whereas health conditions such as being unwell, falling, surgery, and chemotherapy led to 136 skipped participant monitoring days (10.4%) in 11 participants. The skipped days were regarded as noncompliant with daily weighing in the analysis of the primary outcome.

XSL•FO
RenderX

Table 2. The percentage of technical issues in the innovative telemonitoring enhanced care program for chronic heart failure group.

Technical issues	Value, n (%)
Bluetooth connectivity	203 (52.1)
Network connectivity	61 (15.6)
Weighing scale connectivity	49 (12.6)
CHF[a] app	34 (8.8)
Call center system/support	23 (5.9)
Weighing scale battery	18 (4.6)
Tablet battery	2 (0.4)

[a]CHF: chronic heart failure.

Primary Outcome and Related Analysis Results

Applying the intention-to-treat analysis to the primary outcome of weight monitoring at least four days a week on average over the duration of the trial, the proportion of compliant participants in the ITEC-CHF group did not achieve statistical significance compared with that of the control group (ITEC-CHF: 67/91, 74% vs control: 56/93, 60%; *P*=.06). However, ITEC-CHF was associated with significantly more participants who monitored their body weight on average for at least 6 days per week over the duration of the trial than the control (ITEC-CHF: 41/91, 45% vs control: 23/93, 25%; *P*≤.005; Table 3).

Table 3. Comparison of participant compliance with daily weight monitoring between the innovative telemonitoring enhanced care program for chronic heart failure (ITEC-CHF) and the control groups. Under the conventional weight monitoring standard of at least 4 days per week, there was no significant difference between the ITEC-CHF and the control groups (*P*=.06). Under a stricter weight monitoring criterion of at least 6 days per week, more participants in the ITEC-CHF group were found to achieve this criterion than those in the control group (*P*=.005).

Compliance[a]	ITEC-CHF[b], n (%)	Usual care, n (%)	*P* value
Participants who monitored body weight at least 6 days per week	41 (45)	23 (25)	.005
Participants who monitored body weight at least 4 days per week	67 (74)	56 (60)	.06

[a]Compliance with daily weight monitoring.

[b]ITEC-CHF: innovative telemonitoring enhanced care program for chronic heart failure.

Secondary Outcomes

Applying a *per-protocol* analysis by excluding participants who discontinued the study (ITEC-CHF: 24/91 and control: 12/93), the difference in weight monitoring compliance was significant for weight monitoring at least 4 days a week (ITEC-CHF: 65/67, 97% vs control: 56/81, 69%; *P*<.01) and at least 6 days a week (ITEC-CHF: 41/67, 61% vs control: 23/81, 28%; *P*<.01).

In the complete case analysis, 147 participants (ITEC-CHF: 66 and control: 81) completed the Heart Failure Compliance Questionnaire at baseline and 6-month assessments (Table 4). ITEC-CHF was associated with a significantly improved score in the domains of health maintenance (*P*<.01), medication adherence (*P*<.01), and diet (*P*<.01). No significant differences were found in the category of exercise (*P*=.10), smoking (*P*=.48), or alcohol use (*P*=.32).

In the EQ-5D assessment, no significant differences were found in the change in category of mobility (*P*=.44), self-care (*P*=.26), usual activities (*P*=.59), discomfort (*P*=.46), and anxiety or depression (*P*=.38). The mean change in the overall score of EQ-5D was also not significantly different (ITEC-CHF: 4.05, SD 15.95 vs control: 1.10, SD 14.24; *P*=.13).

No significant effects were found for the 6-min walk test distance, frailty, and depression.

No significant differences were found in all-cause hospitalizations (ITEC-CHF: 73 vs control: 58; hazard ratio [HR] 1.18; *P*=.49) or emergency department (ED) presentations (ITEC-CHF: 36 vs control: 45; HR 0.83; *P*=.55), chronic heart failure (CHF)–related hospitalizations (ITEC-CHF: 15 vs control: 8; HR 1.98; *P*=.24), CHF-related ED presentations (ITEC-CHF: 4 vs control: 5; HR 0.98; *P*=.98), or unplanned hospitalizations (ITEC-CHF: 41 vs control: 39; HR 1.06; *P*=.86).

XSL•FO
RenderX

Table 4. Secondary outcomes of self-management behaviors, quality of life, 6-min walk test, frailty, and depression.

Compliance[a]	Baseline				6 months				Difference from baseline				*P* value
	ITEC-CHF[b]		Usual care		ITEC-CHF		Usual care		ITEC-CHF		Usual care		
	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	
Health maintenance score	90 (99)	3.83 (0.48)	93 (100)	3.77 (0.61)	67 (74)	3.88 (0.33)	81 (87)	3.60 (0.96)	66 (73)	0.06 (0.49)	81 (87)	−0.20 (1.03)	.04
Medications score	90 (99)	3.68 (0.79)	93 (100)	3.73 (0.81)	67 (74)	3.70 (0.78)	81 (87)	3.69 (0.74)	66 (73)	0.08 (0.54)	81 (87)	−0.04 (0.87)	.05
Diet score	90 (99)	2.56 (1.11)	93 (100)	2.81 (1.10)	67 (74)	2.97 (1.02)	81 (87)	2.81 (1.05)	66 (73)	0.35 (1.05)	81 (87)	−0.01 (1.07)	.008
Exercise score	90 (99)	1.83 (1.19)	93 (100)	2.04 (1.30)	67 (74)	1.85 (1.33)	81 (87)	1.83 (1.34)	66 (73)	−0.03 (1.02)	81 (87)	−0.21 (1.06)	.1
Smoking score	90 (99)	0.53 (1.43)	93 (100)	0.45 (1.26)	67 (74)	0.39 (1.18)	81 (87)	0.37 (1.13)	66 (73)	−0.14 (1.51)	81 (87)	−0.11 (1.14)	.48
Alcohol use score	90 (99)	0.57 (1.27)	93 (100)	0.70 (1.47)	67 (74)	0.45 (1.07)	81 (87)	0.68 (1.44)	66 (73)	−0.06 (1.19)	81 (87)	0.02 (1.15)	.32
EQ-5D[c]													
Mobility score	90 (99)	1.63 (0.84)	93 (100)	1.86 (0.90)	66 (73)	1.71 (0.95)	81 (87)	1.98 (1.03)	65 (71)	0.12 (0.72)	81 (87)	0.11 (0.81)	.44
Self-care score	90 (99)	1.21 (0.53)	93 (100)	1.33 (0.66)	66 (73)	1.22 (0.54)	81 (87)	1.30 (0.64)	65 (71)	0.03 (0.39)	81 (87)	−0.04 (0.62)	.26
Usual activities score	90 (99)	1.74 (0.94)	93 (100)	1.74 (1.01)	66 (73)	1.65 (1.01)	81 (87)	1.68 (0.92)	66 (73)	−0.06 (0.73)	81 (87)	−0.07 (0.79)	.59
Discomfort score	90 (99)	1.86 (0.94)	93 (100)	2.00 (0.98)	66 (73)	1.86 (1.01)	81 (87)	2.02 (0.97)	65 (71)	0.00 (0.77)	81 (87)	0.02 (0.85)	.46
Anxiety/depression score	90 (99)	1.87 (0.91)	93 (100)	1.82 (0.91)	66 (73)	1.85 (0.92)	81 (87)	1.79 (0.94)	65 (71)	−0.05 (0.74)	81 (87)	−0.01 (0.68)	.38
Your health today score	90 (99)	70.22 (18.54)	93 (100)	68.74 (17.95)	66 (73)	75.97 (20.95)	81 (87)	70.02 (18.62)	65 (71)	4.05 (15.95)	81 (87)	1.10 (14.28)	.13
6-min walk test													
Walked distance (m)	89 (98)	367.24 (122.93)	93 (100)	350.45 (108.96)	62 (68)	396.77 (131.68)	77 (87)	367.19 (130.88)	61 (67)	17.16 (54.23)	77 (83)	10.23 (70.04)	.4
Frailty													
Frailty score	90 (99)	3.07 (1.01)	93 (100)	3.42 (1.13)	63 (69)	2.89 (1.29)	81 (87)	3.22 (1.31)	64 (70)	−0.19 (0.94)	81 (87)	−0.23 (0.90)	.54
Depression													
Sleep score	89 (98)	3.30 (2.10)	91 (98)	3.74 (1.93)	64 (70)	3.40 (2.04)	81 (87)	3.80 (2.00)	65 (71)	0.12 (1.59)	80 (86)	0.01 (1.72)	.5
Spirits score	89 (98)	2.26 (1.39)	91 (98)	2.48 (1.58)	64 (70)	2.28 (1.63)	81 (87)	2.53 (1.54)	65 (71)	0.18 (1.27)	80 (86)	0.02 (1.10)	.48
Tearful score	89 (98)	2.12 (1.80)	91 (98)	2.36 (1.89)	64 (70)	2.08 (1.77)	81 (87)	2.47 (1.84)	66 (73)	−0.03 (1.30)	80 (86)	0.09 (1.22)	.3
Frustrated score	89 (98)	3.04 (2.02)	91 (98)	3.23 (1.78)	64 (70)	2.89 (1.95)	81 (87)	2.89 (1.76)	66 (73)	−0.03 (1.48)	80 (86)	−0.21 (1.26)	.42
Pleasure score	89 (98)	2.39 (1.42)	91 (98)	2.59 (1.59)	64 (70)	2.23 (1.60)	81 (87)	2.46 (1.57)	66 (73)	−0.02 (1.33)	80 (86)	−0.06 (1.04)	.4

[a]Compliance with self-management behaviors was assessed using the Heart Failure Compliance Questionnaire.

[b]The innovative telemonitoring enhanced care program for chronic heart failure (ITEC-CHF) was associated with significant improvements in the subcategories of health maintenance, medication adherence, and diet in the compliance assessment.

[c]EQ-5D: five-dimension EuroQol.

XSL•FO
RenderX

Discussion

Principal Findings

In this study of an innovative telemonitoring program (ITEC-CHF), facilitated by community nurses with call center support, we observed no significant differences in the weight monitoring frequency of at least 4 days a week but observed a significantly higher proportion of the intervention group achieving a weight monitoring frequency of at least 6 days per week compared with the control group receiving usual care. To our knowledge, this is the first study to use objective measures of weight monitoring and the intention-to-treat principle to comprehensively evaluate patient compliance with daily weighing in patients with CHF.

The higher weight monitoring frequency of at least 6 days per week reflects better compliance with the recommendation in contemporary clinical guidelines for patients with CHF to weigh themselves daily as a self-management strategy to maintain fluid balance and identify signs of edema [2,3]. This criterion for compliance is stricter than that applied to weight monitoring in previous studies of *most of the time* (or at least 4 days per week) and was limited by a questionnaire-based assessment of compliance, which is prone to bias [4,12].

In the intention-to-treat analysis, 45% of participants randomized to ITEC-CHF achieved a monitoring frequency of at least 6 days per week over the 6-month follow-up period of the trial. This figure was influenced by the relatively high proportion of participants who discontinued the trial from the intervention group and provides valuable insight into factors that are pertinent to telemonitoring in clinical practice. In the early stages of the trial, there were relatively frequent technical issues with the telemonitoring system, which may have led to the withdrawal of some participants. It has previously been reported that learning how to use telemonitoring technology is perceived as burdensome and creates anxiety in some patients, especially those who are older [24]. This may be further exacerbated in the event of technical issues. Technical issues also resulted in increased reliance on technical support, which would have increased the cost of telemonitoring, although this was not assessed in this study, highlighting the need for future telemonitoring studies with a health economics component. These issues highlight the importance of telemonitoring systems being seamless and reliable to not create an unnecessary burden on patients and their carers or health service providers [25]. Therefore, they underscore the need to improve the reliability and user experience of telemonitoring systems [26] for use in clinical CHF care.

In participants who completed the trial (the cohort in which the per-protocol analysis was conducted), compliance with weight monitoring in ITEC-CHF was high; 97% of participants monitored themselves for at least 4 days a week over the 6-month duration of the trial, and 61% of participants monitored themselves at least 6 days a week. In the ITEC-CHF group, 390 participant monitoring days were *skipped* because of technical issues, meaning that the difference in weight monitoring compliance between the ITEC-CHF and control groups is likely to be underestimated. These positive findings regarding weight monitoring compliance in participants who adhered to the program are likely to be underpinned by the multifactorial support mechanisms provided. This is further supported by a substantial number of alerts reported in this study, which included the automated generation of reminder alerts on 715 patient days when a weight recording was not received by 10 AM as well as contact made by the call center and project nurses. These findings not only demonstrated the effectiveness of ITEC-CHF in supporting weight management but also indicated a strong need for such support in ongoing CHF care.

The ITEC-CHF group experienced a significant improvement in health maintenance compared with usual care, as measured by the Heart Failure Compliance Questionnaire. This positive result was consistent with a finding of improved self-care maintenance reported in 2 other RCTs of telemonitoring in CHF [27,28]. Similarly, there was a significant improvement in adherence to medication and diet recommendations in the ITEC-CHF group, but not in the control group. These findings imply increased engagement with the heart failure nurses that occurred following the alerts generated through the telemonitoring intervention. These interactions created the opportunity for *teachable moments*, enabling nurses to provide informal education to reinforce self-management practices. It has previously been acknowledged that patients often benefit from ongoing support in CHF care to effectively manage their health conditions through reinforcement of self-management strategies [29]. It is also possible that merely being monitored was sufficient to enhance adherence to more desirable patterns of care because of a surveillance effect [30]. Nevertheless, the combination of telemonitoring with nursing support resulted in improved self-management activities, although it is often difficult for researchers to identify which component of the ITEC-CHF program drove these improvements.

There were no significant effects of ITEC-CHF on hospitalizations and ED presentations, although it should be noted that the study was underpowered for this analysis. We observed that CHF-related events were not the major cause of hospitalizations (23/131, 17.5%) or ED presentations (9/81, 11%). CHF is a complex condition, more prevalent in older people, and associated with a range of comorbidities. Telemonitoring in the context of this study focused exclusively on daily weight recordings. This finding implies a need to extend telemonitoring intervention for comorbidity and critical non-CHF–related health conditions to more comprehensively address the range of health issues faced by patients with CHF. To date, the effectiveness of telemonitoring in improving hospitalization and mortality remains inconclusive [13,14]. However, it has been shown that reduction in health services utilization, including unscheduled hospitalizations and length of stay, can be achieved for broader chronic disease management by monitoring a range of vital signs using telemonitoring enhanced care coordination [31,32]. Further research to understand the underlying principles that impact hospitalizations and ED presentations related to a specific primary diagnosis such as CHF remains essential in future studies.

There are several limitations of this study that warrant discussion. First, the study was limited to a 6-month intervention, which may have been insufficient to translate to

XSL•FO
RenderX

meaningful changes in clinical characteristics of patients. In addition, a relatively high number of participants discontinued their involvement in the trial, which has the potential to bias the analysis of some secondary outcomes where the intention-to-treat principle could not be applied. High rates of discontinuation in the ITEC-CHF group suggest a potential bias in the per-protocol analysis. Participants who discontinued in the ITEC-CHF group were unlikely to be random because some were because of deteriorating health (palliative care or dialysis) and because some were deceased, as shown in Figure 1. Therefore, we did not use the multiple imputation approach outlined in the trial protocol [15]. The study did not achieve the target sample size (n=300), proposed in the trial protocol [15]; this might have compromised the power to detect significant effects in the analysis of the primary outcome. There are several reasons for the smaller than proposed sample size. We experienced technical issues with the telemonitoring software that were not apparent during an extensive testing phase before study commencement, which delayed recruitment. This highlights the challenges that arise in a real-world environment that may not be present in a testing scenario. We also experienced slower than anticipated recruitment; some patients were reluctant to engage in a model of care involving technology, whereas others were concerned about their ability to weigh themselves safely because of frailty and were therefore excluded from the trial. A substantial number of patients reported as having CHF did not have an echocardiogram documented in their medical records and were therefore excluded because they failed to meet the inclusion criteria of an EF of <40%. Finally, a high proportion of patients who were residents of nursing homes were excluded on this basis. These issues highlight that the clinical complexity of many patients with CHF, who are often older with multiple comorbidities, may complicate their ability to engage with telemonitoring and underpins the importance of telemonitoring models being developed that are reliable, easy to use, and accessible for patients across the clinical spectrum of CHF. Finally, the patient compliance rate in usual care (25% for at least 6 days per week and 60% for at least 4 days per week) was likely to be influenced by the provision of weight scales to participants, which kept a record of their weight recordings and, accordingly, might have resulted in a Hawthorne effect [33], reducing the difference between the ITEC-CHF and control groups.

Conclusions

The ITEC-CHF study is the first to report the effects of telemonitoring on weight monitoring compliance using an objective measure of weight recordings in patients with CHF. The proportion of participants in the ITEC-CHF program achieving a weight monitoring frequency of at least 6 days per week was higher than that in usual care controls. Furthermore, ITEC-CHF resulted in significant improvements in CHF self-management related to health maintenance, medication adherence, and diet. Among participants who completed the study, there was a high level of compliance with weight monitoring, underscoring the importance of telemonitoring platforms that are seamless to reduce the risk of patients disengaging with the technology. Although telemonitoring and digital health more broadly offer significant potential for supporting patients in self-managing chronic conditions such as heart failure, further research is required to refine these evolving strategies to achieve effective care outcomes.

Acknowledgments

The authors gratefully acknowledge (1) the members of the Project Control Board, including Mr David Anderson, Dr Phil Carrillo, Dr Fergus McGee, and Mr Mark Gravell for their management and guidance in the clinical trial; (2) the members of MEPACS, including Mr Dean Richardson and Mr Dhruv Singh, for implementing the telemonitoring system and managing the call service in the clinical trial; (3) the members of Medtech Global, including Mr Rama Kumble and GK Shridhar, for implementing and managing the telemonitoring system in the clinical trial; and (4) nursing and medical support provided by Julie Smith and Dr Matthew Best from the trial site in WA. This project is fully funded by the health market validation program by the Department of State Development, Business and Innovation, Victorian government (121 Exhibition Street, Melbourne, Vic, 3001, Australia. The contract was signed in May 2014), and the Western Australia Targeted Research Fund Grant by the Department of Health, WA (notification, February 4, 2014). These funding sponsors will not have any role in the design of this study, execution of the trial, analysis and interpretation of the trial data, and research publications.

Authors' Contributions

This outcome study was mainly developed by HD, AD, IE, and RJ. The first draft of the manuscript was written by HD, SC, and RJ. The statistical analysis was mainly conducted by HD and NG (a statistician at Commonwealth Scientific and Industrial Research Organization). All authors contributed to the design of the clinical study and/or critical revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
User interface.
[PNG File , 46 KB - jmir_v22i7e17559_app1.PNG]

XSL•FO
RenderX

Multimedia Appendix 2
Bluetooth-enabled scales.
[PNG File , 3865 KB - jmir_v22i7e17559_app2.png]

Multimedia Appendix 3
International Classification of Diseases, Tenth Revision, Clinical Modification, diagnosis codes used to determine heart failure–related hospitalizations and emergency department presentations.
[DOCX File , 14 KB - jmir_v22i7e17559_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 7603 KB - jmir_v22i7e17559_app4.pdf]

References

1. Savarese G, Lund LH. Global public health burden of heart failure. Card Fail Rev 2017 Apr;3(1):7-11 [FREE Full text] [doi: 10.15420/cfr.2016:25:2] [Medline: 28785469]
2. NHFA CSANZ Heart Failure Guidelines Working Group, Atherton JJ, Sindone A, de Pasquale CG, Driscoll A, MacDonald PS, et al. National heart foundation of Australia and cardiac society of Australia and New Zealand: guidelines for the prevention, detection, and management of heart failure in Australia 2018. Heart Lung Circ 2018 Oct;27(10):1123-1208 [FREE Full text] [doi: 10.1016/j.hlc.2018.06.1042] [Medline: 30077227]
3. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, ESC Scientific Document Group. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC)developed with the special contribution of the heart failure association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
4. van der Wal MH, van Veldhuisen DJ, Veeger NJ, Rutten FH, Jaarsma T. Compliance with non-pharmacological recommendations and outcome in heart failure patients. Eur Heart J 2010 Jun;31(12):1486-1493. [doi: 10.1093/eurheartj/ehq091] [Medline: 20436049]
5. Jeon Y, Kraus SG, Jowsey T, Glasgow NJ. The experience of living with chronic heart failure: a narrative review of qualitative studies. BMC Health Serv Res 2010 Mar 24;10:77 [FREE Full text] [doi: 10.1186/1472-6963-10-77] [Medline: 20331904]
6. Ni H, Nauman D, Burgess D, Wise K, Crispell K, Hershberger RE. Factors influencing knowledge of and adherence to self-care among patients with heart failure. Arch Intern Med 1999 Jul 26;159(14):1613-1619. [doi: 10.1001/archinte.159.14.1613] [Medline: 10421285]
7. Toh CT, Jackson B, Gascard DJ, Manning AR, Tuck EJ. Barriers to medication adherence in chronic heart failure patients during home visits. J Pharm Pract Res 2015 Apr 13;40(1):27-30. [doi: 10.1002/j.2055-2335.2010.tb00721.x]
8. Jones CD, Holmes GM, Dewalt DA, Erman B, Broucksou K, Hawk V, et al. Is adherence to weight monitoring or weight-based diuretic self-adjustment associated with fewer heart failure-related emergency department visits or hospitalizations? J Card Fail 2012 Jul;18(7):576-584 [FREE Full text] [doi: 10.1016/j.cardfail.2012.05.004] [Medline: 22748492]
9. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
10. Boyne JJ, Vrijhoef HJ, Spreeuwenberg M, de Weerd G, Kragten J, Gorgels AP, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
11. Ramaekers BL, Janssen-Boyne JJ, Gorgels AP, Vrijhoef HJ. Adherence among telemonitored patients with heart failure to pharmacological and nonpharmacological recommendations. Telemed J E Health 2009;15(6):517-524. [doi: 10.1089/tmj.2009.0160] [Medline: 19566401]
12. Evangelista LS, Berg J, Dracup K. Relationship between psychosocial variables and compliance in patients with heart failure. Heart Lung 2001;30(4):294-301. [doi: 10.1067/mhl.2001.116011] [Medline: 11449216]
13. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
14. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]

XSL•FO
RenderX

15. Ding H, Jayasena R, Maiorana A, Dowling A, Chen SH, Karunanithi M, et al. Innovative telemonitoring enhanced care programme for chronic heart failure (ITEC-CHF) to improve guideline compliance and collaborative care: protocol of a multicentre randomised controlled trial. BMJ Open 2017 Oct 8;7(10):e017550 [FREE Full text] [doi: 10.1136/bmjopen-2017-017550] [Medline: 28993389]
16. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
17. Enright PL. The six-minute walk test. Respir Care 2003;48(8):783-785. [doi: 10.1007/springerreference_184598]
18. Hare DL, Selvadurai LP, Wang JQ, Yau HH, Stone MJ, Raman B, et al. Psychometric assessment of the cardiac depression scale short form in cardiac outpatients. Eur J Cardiovasc Nurs 2017 Mar;16(3):249-255. [doi: 10.1177/1474515116652759] [Medline: 27231395]
19. Cheung A, Haas B, Ringer TJ, McFarlan A, Wong CL. Canadian study of health and aging clinical frailty scale: does it predict adverse outcomes among geriatric trauma patients? J Am Coll Surg 2017 Nov;225(5):658-65.e3. [doi: 10.1016/j.jamcollsurg.2017.08.008] [Medline: 28888692]
20. World Health Organization. 2016. ICD-10 Online Versions URL: https://www.who.int/classifications/icd/icdonlineversions/en/ [accessed 2020-07-08]
21. Senn S. Change from baseline and analysis of covariance revisited. Stat Med 2006 Dec 30;25(24):4334-4344. [doi: 10.1002/sim.2682] [Medline: 16921578]
22. Ozga A, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. BMC Med Res Methodol 2018 Jan 4;18(1):2 [FREE Full text] [doi: 10.1186/s12874-017-0462-x] [Medline: 29301487]
23. RStudio. 2015. URL: http://www.rstudio.com/ [accessed 2020-05-12]
24. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
25. Ding H, Fatehi F, Russell AW, Karunanithi M, Menon A, Bird D, et al. User experience of an innovative mobile health program to assist in insulin dose adjustment: outcomes of a proof-of-concept trial. Telemed J E Health 2018 Jul;24(7):536-543. [doi: 10.1089/tmj.2017.0190] [Medline: 29261476]
26. Woo K, Dowding D. Factors affecting the acceptance of telehealth services by heart failure patients: an integrative review. Telemed J E Health 2018 Apr;24(4):292-300 [FREE Full text] [doi: 10.1089/tmj.2017.0080] [Medline: 28767315]
27. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
28. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a tablet computer on self-care, quality of life, and knowledge: a randomized clinical trial. J Cardiovasc Nurs 2018;33(4):336-343. [doi: 10.1097/JCN.0000000000000462] [Medline: 29369123]
29. Hocking A, Laurence C, Lorimer M. Patients' knowledge of their chronic disease - the influence of socio-demographic characteristics. Aust Fam Physician 2013 Jun;42(6):411-416 [FREE Full text] [Medline: 23781550]
30. Madigan E, Schmotzer BJ, Struk CJ, DiCarlo CM, Kikano G, Piña IL, et al. Home health care with telemonitoring improves health status for older adults with heart failure. Home Health Care Serv Q 2013;32(1):57-74 [FREE Full text] [doi: 10.1080/01621424.2012.755144] [Medline: 23438509]
31. Celler B, Varnfield M, Nepal S, Sparks R, Li J, Jayasena R. Impact of at-home telemonitoring on health services expenditure and hospital admissions in patients with chronic conditions: before and after control intervention analysis. JMIR Med Inform 2017 Sep 8;5(3):e29 [FREE Full text] [doi: 10.2196/medinform.7308] [Medline: 28887294]
32. Ding H, Fatehi F, Maiorana A, Bashi N, Hu W, Edwards I. Digital health for COPD care: the current state of play. J Thorac Dis 2019 Oct;11(Suppl 17):S2210-S2220 [FREE Full text] [doi: 10.21037/jtd.2019.10.17] [Medline: 31737348]
33. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

CHF: chronic heart failure
ED: emergency department
EF: ejection fraction
EQ-5D: five-dimension EuroQol
GP: general practitioner
ITEC-CHF: innovative telemonitoring enhanced care program for chronic heart failure
MMH: Manage My Health
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by F Fatehi, E van der Velde; comments to author 09.03.20; revised version received 02.04.20; accepted 22.04.20; published 08.07.20.

Please cite as:
Ding H, Jayasena R, Chen SH, Maiorana A, Dowling A, Layland J, Good N, Karunanithi M, Edwards I
The Effects of Telemonitoring on Patient Compliance With Self-Management Recommendations and Outcomes of the Innovative Telemonitoring Enhanced Care Program for Chronic Heart Failure: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17559
URL: https://www.jmir.org/2020/7/e17559
doi:10.2196/17559
PMID:32673222

©Hang Ding, Rajiv Jayasena, Sheau Huey Chen, Andrew Maiorana, Alison Dowling, Jamie Layland, Norm Good, Mohanraj Karunanithi, Iain Edwards. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Wearable Motion Sensor Device to Facilitate Rehabilitation in Patients With Shoulder Adhesive Capsulitis: Pilot Study to Assess Feasibility

Yu-Pin Chen[1,2,3], MD; Chung-Ying Lin[4], PhD; Ming-Jr Tsai[5], MD; Tai-Yuan Chuang[2,3], MD; Oscar Kuang-Sheng Lee[1,6,7], PhD, MD

[1]Institute of Clinical Medicine, National Yang Ming University, Taipei, Taiwan

[2]Department of Orthopedic Surgery, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[3]Department of Orthopedic Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[4]Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, China (Hong Kong)

[5]Department of Orthopedic Surgery, Puli Christian Hospital, Nantou, Taiwan

[6]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan

[7]Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan

Corresponding Author:
Oscar Kuang-Sheng Lee, PhD, MD
Institute of Clinical Medicine
National Yang Ming University
No.155, Sec.2, Linong Street
Taipei, 112
Taiwan
Phone: 886 2 28200183
Email: oscarlee9203@gmail.com

Abstract

Background: Adhesive capsulitis (AC) of the shoulder is a common disorder that painfully reduces the shoulder range of motion (ROM) among middle-aged individuals. Although physical therapy with home-based exercises is widely advised to restore ROM in the treatment of AC, clinical results vary owing to inconsistent patient compliance.

Objective: In this study, we aimed to verify the feasibility of a treatment model that involves applying a wearable motion sensor device to assist patients conduct home-based exercises to improve training compliance and the accuracy of exercises, with the ultimate goal of improving the functional recovery of patients with AC.

Methods: The motion sensor device was comprised of inertial measurement unit–based sensors and mobile apps for patients and physicians, offering shoulder mobility tracing, home-based exercise support, and progress monitoring. The interrater reliability of shoulder mobility measurement using the motion sensor device on 10 healthy participants and 15 patients with AC was obtained using an intraclass correlation coefficient analysis and compared with the assessments performed by two highly experienced physicians. A pilot prospective control trial was then carried out to allocate the 15 patients with AC to two groups: home-based exercise group and motion sensor–assisted rehabilitation group. Changes in active and passive shoulder ROM, pain and functional scores, and exercise completion rates were compared between the groups during a treatment period of 3 months.

Results: Shoulder ROM, as measured using the motion sensor device, exhibited good to excellent reliability based on the comparison with the measurements of two physicians (intraclass correlation coefficient range, 0.771 to 0.979). Compared with patients with AC in the home-based exercise group, those in the motion sensor–assisted rehabilitation group exhibited better shoulder mobility and functional recovery and a higher exercise completion rate during and after 3 months of rehabilitation.

Conclusions: Motion sensor device–assisted home-based rehabilitation for the treatment of AC is a useful treatment model for telerehabilitation that enhances the compliance of patients through training, thus improving functional recovery. This helps overcome important obstacles in physiotherapy at home by providing comprehensible and easily accessible exercise instructions, enhancing compliance, ensuring the correctness of exercise, and monitoring the progress of patients.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e17032) doi:10.2196/17032

KEYWORDS

motion sensor; adhesive capsulitis; rehabilitation; home-based exercise; telerehabilitation; telehealth, telemonitoring

Introduction

Adhesive capsulitis (AC) of the shoulder, which occurs in approximately 2% to 5% of the general population, is an idiopathic, progressive, and painful restriction of the active range of motion (aROM) and passive range of motion (pROM) [1,2]. AC typically affects patients older than 50 years, and involvement of both shoulders is noted in 20% to 30% of cases [1,2]. Although symptomatic improvement tends to occur naturally within years even with minimal treatment [3], approximately 50% of patients experience pain or some mild restriction of movement, and 11% experience some residual disability even several years after the resolution of their other symptoms [4,5]. Although AC is a common condition, high-quality evidence of successful treatment methods for AC has not yet been obtained [6]. The well-accepted standard treatment for AC mostly involves physical therapy and home exercises to restore ROM [7,8]. Evidence suggests that, compared with less frequent self-exercise in a home setting or joint mobilization sessions in a hospital setting, regular and daily self-exercise in a home setting could contribute to greater improvement in shoulder ROM and a shorter duration of symptoms [9].

Despite realizing the importance of daily physiotherapy, including mobilization and strength exercises, clinicians have achieved varied outcomes in patients who are trained in a home-based exercise program [10]. Challenges that affect the effectiveness of home-based exercises for AC may include training compliance and exercise correctness. Patients with AC do not maintain their training frequency and duration at home because the prescribed exercise programs are typically not followed without the constant supervision of a physiotherapist. Failure to incorporate exercises into daily life is the main form of noncompliance, as reported in up to 60% of patients whose treatment plan included home-based exercise [11]. Whether patients can correctly perform exercises at home after initial instruction by physiotherapists is also a concern [12]. Therefore, the development of methods to improve compliance and exercise correctness for patients with AC is worthwhile to maximize the effectiveness of home-based physiotherapy.

Techniques for detecting bodily motions are extensively used in health care to monitor and rehabilitate disabled patients. With the evolution of sensing and body area network technologies, wearable rehabilitation technology has opened up the possibility of independent training, which has advantages over traditional rehabilitation services [13]. Inertial measurement units (IMUs), including accelerometers and gyroscopes, have been extensively used in technology-assisted rehabilitation, with sufficient efficacy [14-16]. However, despite the potential of using IMU-based sensors in neurorehabilitation and for treating musculoskeletal impairments [13,17,18], few such sensors have been used in clinical studies, especially those involving physiotherapy for AC.

Through this study, we tested wearable IMU-based sensors that integrated with interactive mobile apps using wireless telecommunication technology to enable physiotherapists to monitor the progress of patients with AC who were conducting home-based self-exercise and improve their training compliance. We hypothesized that the use of a wearable IMU-based motion sensor device could help patients perform home-based exercises correctly and increase their motivation and compliance regarding training, thereby improving functional outcomes.

Methods

Motion Sensor Device

A wearable motion sensor device (BoostFix wearable self-training kit, COMPAL Electronics Inc, Taipei, Taiwan) was used (see Multimedia Appendix 1). The motion sensor device was comprised of wearable IMU-based sensors that record the angular shoulder motion of a patient, a mobile app termed Patient App for use by the patient, and a mobile app termed Doctor App for use by a qualified health care professional.

Wearable IMU–Based Sensors

The IMU-based sensors were comprised of 6-axis microelectromechanical systems that included accelerometers and gyroscopes. These devices collect information about the angular motion of the shoulder of interest. Three sensors are required for angular measurements of the shoulder. In this study, they were strapped to the sternum, distal third of the lateral upper arm approximately 5 cm above the lateral epicondylar of the humerus bone, and dorsal wrist (Figure 1). Three sensors were calibrated in position A once the measurement began (Figure 1A). The initial calibration process involved placing the sensor on a horizontal fixture to measure the offset for each axis, to align the initial difference for each sensor. The calibration process involved obtaining the initial value of the sensor for relative offset correction instead of correcting the deviation caused by incorrect sensor placement. In addition, the process entailed identifying the original offset of the sensors to each other and applying the zero setting using an algorithm.

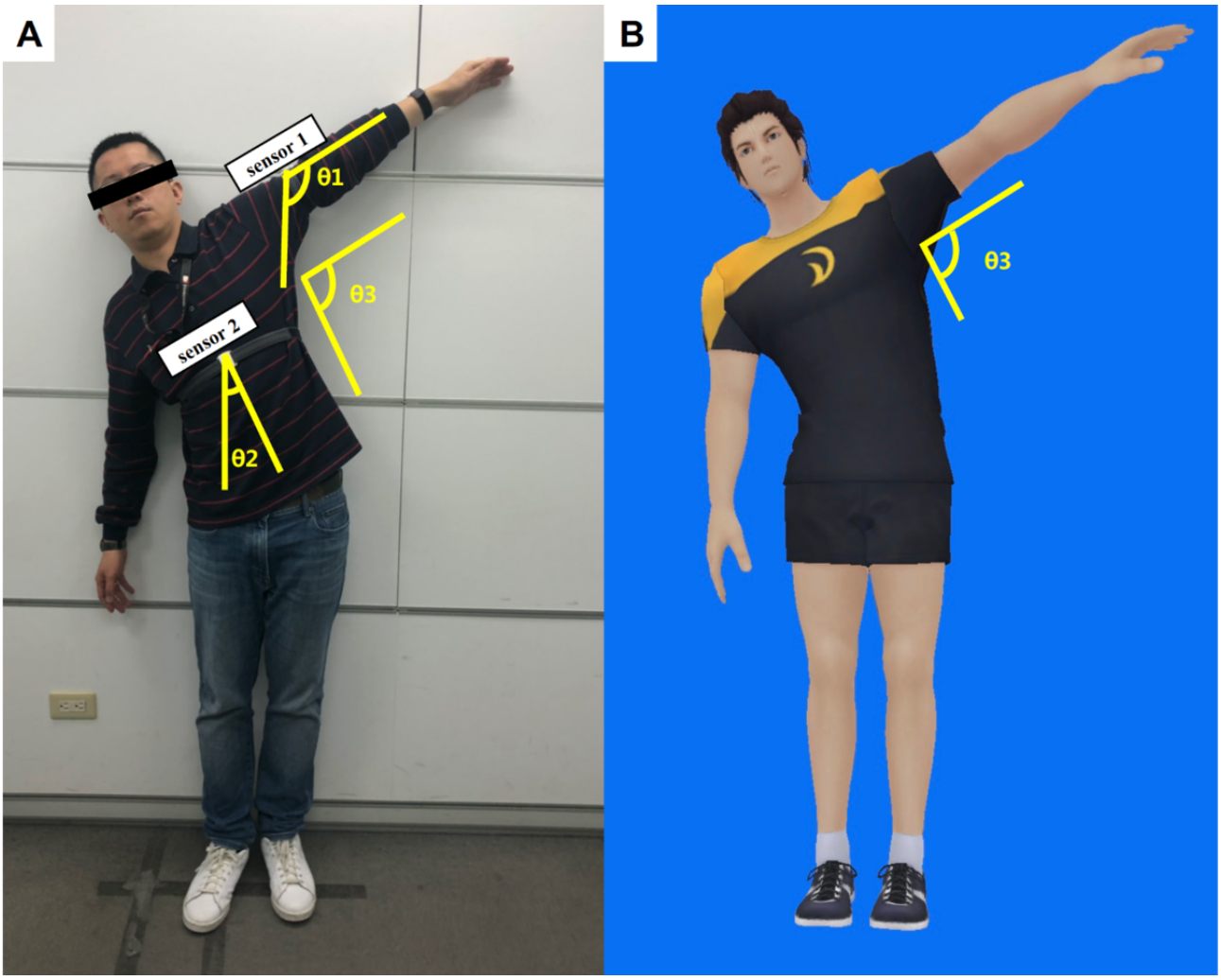
XSL•FO
RenderX

Figure 1. Angular measurement of the shoulder with motion sensors worn on the body for (A) calibration, (B) shoulder abduction, (C) shoulder flexion, (D) shoulder extension, (E) shoulder external rotation, (F) shoulder internal rotation.



Dual-sensor systems were used in the motion sensor device to accurately record the shoulder ROM angle. Using the sensors on the sternum as a reference datum point, the raw data from each sensor can be converted into a quaternion algorithm to convert the relative angle changes of the sensors on the upper arm and wrist into a 3-dimensional (3D) motion of the shoulder structure. Although one sensor can provide the angular change with the gyroscopes on one plane of motion, dual-sensor systems can provide relative angular changes in shoulder movements between the arm and axial body (by compensating for the angle of the body tilting relative to the ground), which was superior for eliminating the error in the shoulder ROM measurement attributed to body tilting from a one-sensor system, especially for patients with AC and painful shoulder movement (Figure 2). Additionally, repeated angular measurements with the dual-sensor system on simulated shoulder motion was validated as highly accurate. This is referred to in Multimedia Appendix 2.

Figure 2. Dual-sensor system simulates the motion of (A) patients with adhesive capsulitis performing shoulder abduction using a (B) 3-dimensional avatar. θ1: angular motion reported by the sensor on the upper arm; θ2: axial body tilting angle reported by the sensor on the sternum; θ3: shoulder abduction angle (ie, θ1 - θ2).



XSL•FO
RenderX

Mobile Phone App for Patients

To design the mobile app for patients with AC for use on mobile phones, a cocreation process was used. The app has modes for measuring shoulder mobility, generating historical records of angular measurements and exercise completion rates, and providing daily shoulder exercises with detailed instructions (Figure 3). The mobile app determines shoulder ROM in all directions by calculating the relevant information regarding the angle measurement reported by the 3 wearable sensors (Figures 2B and 3B). The Patient App provides 7 sets of home-based shoulder exercises for training, including a forward wall walking stretch, lateral wall walking stretch, and cane stretch for shoulder flexion, extension, abduction, internal rotation, and external rotation (Figure 4). The mobile app assigns each set of exercises as a "daily task" assigned by the supervising physiotherapist or physicians and provides a demonstration using a 3D avatar (Figure 3E). The app mirrors the user's shoulder movement during the exercise using angular information collected from the worn motion sensors. Each user has an account in the Patient App to access his or her records, which includes daily progress and the completion rate of the daily exercises (Figures 3C and 3D).

Figure 3. Examples of screens on the mobile app for patients, including the (A) functional introduction, (B) shoulder mobility measurement, (C) historical records of shoulder range of motion, (D) historical records of daily exercise completion rate, (E) daily home-based exercise tasks.

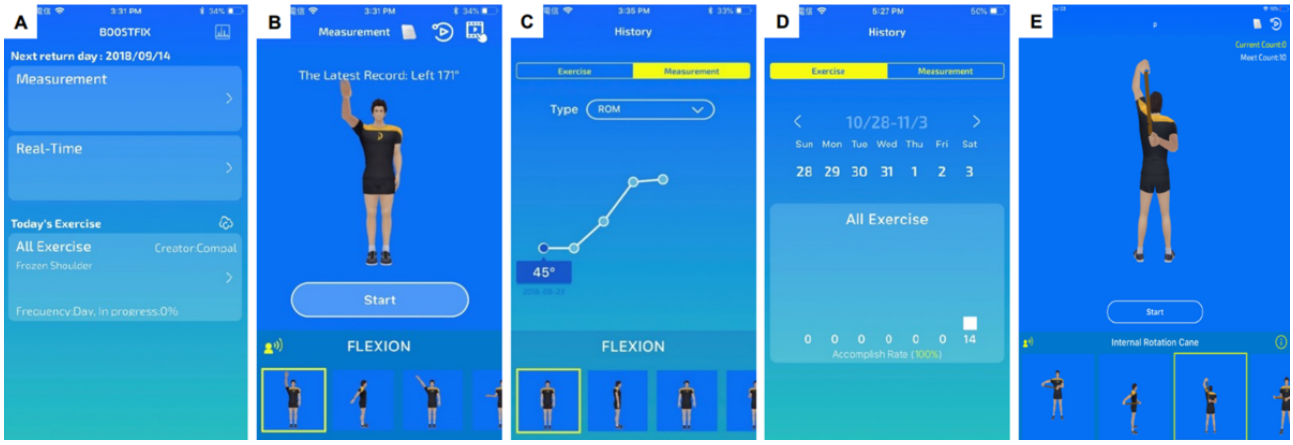
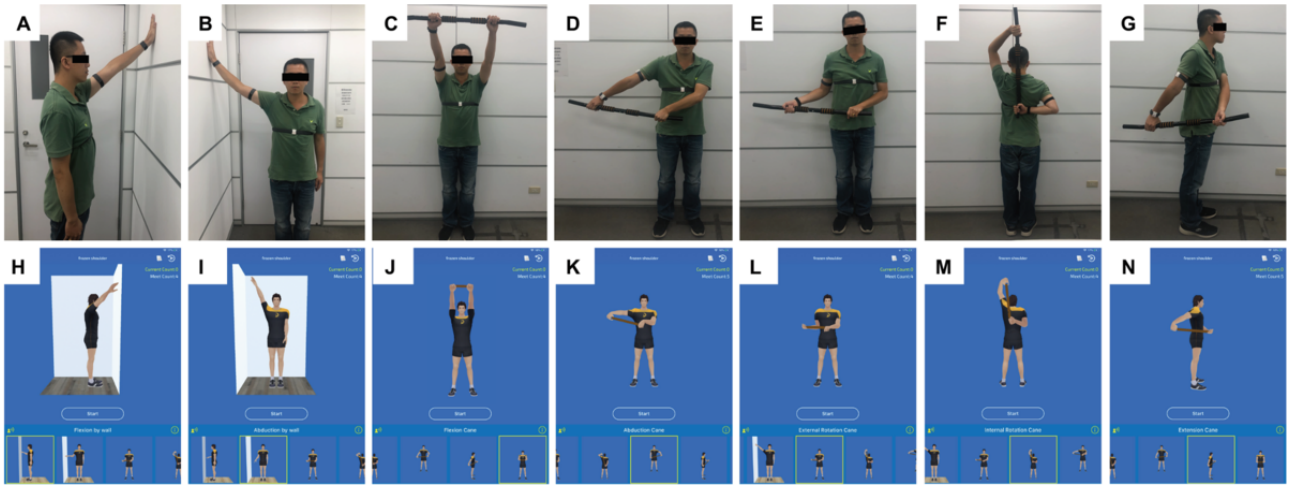


Figure 4. Sets of home-based shoulder exercises displayed on the mobile app and demonstrated by participants: (A, H) forward wall walking stretch; (B, I) lateral wall walking stretch; (C, J) cane stretch for shoulder flexion; (D, K) cane stretch for shoulder abduction; (E, L) cane stretch for shoulder external rotation; (F, M) cane stretch for shoulder internal rotation; (G, N) cane stretch for shoulder extension.

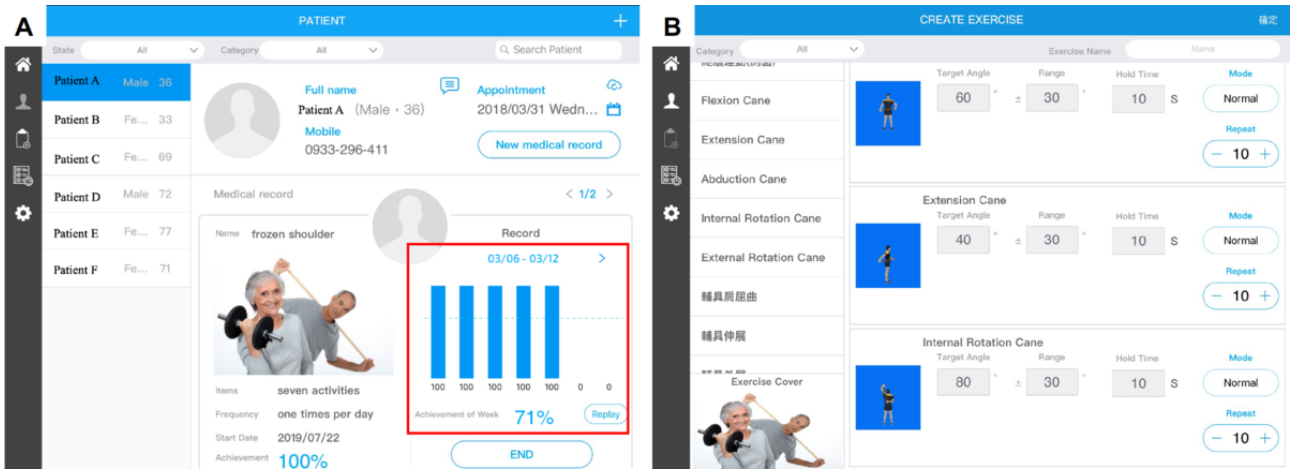


Mobile App for Physiotherapists and Physicians

The Doctor App was designed for physiotherapists and physicians for use on a mobile tablet. This app provides each patient's information, including the latest shoulder ROM measurements and exercise completion rates for the previous week (Figure 5A). Physiotherapists and physicians can assign personalized daily home-based exercises with adjustable targeted angles, numbers of repetitions, and holding times for each patient based on the angular status of the affected shoulder (Figure 5B). Physiotherapists/physicians can directly communicate with the patients by sending text messages through the Doctor App.

XSL•FO
RenderX

Figure 5. Examples of screens on the mobile app for physiotherapists and physicians, including the (A) daily and weekly exercise completion rate of each patient displayed with patient information and (B) doctor-assigned daily exercises with specified targeted angles, holding times, and numbers of repetitions.



Study Protocol

This study included two investigations. The first evaluated the reliability of measurements of shoulder ROM using the motion sensor device, and the second confirmed the effectiveness of using the motion sensor device for the rehabilitation training of patients with AC. The Ethical Committee of Taipei Medical University approved the entire protocol and instrumentation (TMU-JIRB N201708048). All participants consented to participation in the study and the publication of data.

Reliability of Shoulder ROM Measurements Using a Motion Sensor Device

From January 2017 to August 2017, two groups of participants — 10 healthy participants and 15 patients with a confirmed diagnosis of AC — were prospectively enrolled for measurement of shoulder angular motion.

Volunteers aged 20 to 70 years were included in the group of healthy participants. Participants who reported discomfort or limited ROM of the shoulder within the preceding year were excluded. Patients were included in the AC group who (1) were aged between 40 and 70 years, (2) had spontaneous onset of a painful stiff shoulder and marked loss of active and passive global shoulder motion, (3) had reported local shoulder pain that was frequently present either over the anteromedial aspect of the shoulder extending distally into the biceps region or over the lateral aspect of the shoulder extending into the lateral deltoid region, and (4) had symptoms for at least 3 months, with normal findings on anteroposterior and axillary radiographs of the glenohumeral joint. Patients were excluded if they (1) had pathological findings of rotator cuff tear on physical examination (eg, abduction weakness and positive dropped arm test); (2) had glenohumeral osteoarthritis upon radiographic evaluation; (3) had clinical evidence of severe cervical spine disease; (4) had a history of severe trauma to the shoulder; (5) received a local corticosteroid injection or any physiotherapeutic intervention to the affected shoulder in the 3 months preceding the study start date; (6) had a history of inflammatory joint disease, infection, stroke, or thyroid diseases that affected the shoulder; or (7) were unwilling to undergo an intervention or participate in the trial.

For angle measurements, aROM and pROM for flexion, extension, abduction, and internal and external rotation at a neutral position for both shoulders of healthy participants and for the affected shoulder of the patients with AC were assessed. The shoulder ROM in all directions for each participant was first assessed using the motion sensor device. Two examiners (one well-trained physiotherapist and one highly experienced orthopedic surgeon) who were blinded to the angle measurements by the motion sensor device measured shoulder ROM in all directions using a universal goniometer.

Each participant stood straight, and scapular rotation was permitted to measure abduction, flexion, and extension (Figures 1B, 1C, 1D). External rotation was measured in a neutral position with the shoulder abducted, the elbow flexed at 90°, and the forearm in a neutral position; the angle between the long axis of the forearm and the sagittal plane of the trunk was measured as external rotation (Figure 1E). Internal rotation was determined by having the participants reach up the back to the highest point along the midline (Figure 1F). Six levels of internal rotation were defined according to the areas that the participant's hand reached, as follows: interscapular region (T7), T12 vertebra, waist (L3), lumbosacral junction, buttock, and lateral thigh. A high reach corresponded to better internal rotation. The aROM was measured by instructing participants to move their arm as far as they could, and the pROM was measured as the examiner moved the subject's arm to its mechanical limit or a limit imposed by pain.

Effectiveness of Motion Sensor Device for Rehabilitation Training of Patients with AC

The 15 patients with confirmed diagnoses of AC were allocated to 2 intervention groups: 7 patients to the home-based exercise group and 8 patients to the motion sensor–assisted rehabilitation group. The patients were allocated according to their willingness to use, and familiarity with, mobile apps. The patients in both groups received 3 months of rehabilitation.

Patients with AC in the home-based exercise group received comprehensive instructions on the daily shoulder exercises from an experienced orthopedic surgeon before beginning the exercise program. The information included instructions for home-based

XSL•FO
RenderX

exercises; a description of frozen shoulder; and advice on sleep, posture, and pain relief. Home-based exercises included Codman's or pendulum exercises (circumduction) and passive stretching exercises, including "wall walking stretch exercises" and "cane stretch exercises" (Figure 4). Further information regarding the standard protocol of the home-based exercises is available in Multimedia Appendix 3. The patients were instructed to perform these exercises within the painless range, maintain the maximal achievable angle for at least 10 seconds in each exercise, and repeat each exercise at least 10 times daily. All patients were instructed to follow the standard home-based exercise protocol, but individualized modification was permitted based on health professionals' judgment. In practice, an orthopedic surgeon reviewed the progress of each patient every month. They gave advice and modified the exercise program based on the patient's compliance and AC status. Therefore, if a patient had poor performance on a specific direction of shoulder ROM after the exercise program of the previous month (eg, on shoulder abduction), the orthopedic surgeon could make adjustments to the next month's exercise protocol by increasing the daily exercise frequency to 15-20 times and the holding time to 15-20 seconds per session for the "lateral wall walking stretch" and "cane stretch for shoulder abduction" exercises.

All patients in the motion sensor–assisted rehabilitation group received detailed instructions on operating the motion sensor device and Patient App on their own mobile phones before they began the rehabilitation program. The aforementioned orthopedic surgeon assigned the standard protocol of daily home-based shoulder exercises to each patient via the Patient App with individualized modifications according to the patient's performance. The completion rate of these daily tasks was calculated using the app for each patient and sent to the Doctor App over the internet, enabling the supervisor to immediately discuss the progress of each patient with them. In addition, patients in the motion sensor–assisted rehabilitation group were individually assigned each set of stretching exercises with the target angle on the mobile app by the orthopedic surgeon based on the patients' previous shoulder angular performance (Figure 5B). For example, if the patient had 100° shoulder abduction before starting the shoulder exercise program, the target angle for the "lateral wall walking stretch" and "cane stretch for shoulder abduction" would be assigned as 120° with an acceptable range of 30°. Achievement of the target angle of the assigned exercise could be recorded as a successful count for calculating the daily exercise completion rate on the mobile app (Figure 3D). In this manner, the patient could therefore stretch the involved shoulder to the maximal limit of the angle on each attempt, thereby effectively maximizing the training effect of stretching exercises.

Outcome Measurement

Basic demographic data, including age, gender, and educational status, were collected for each patient with AC through face-to-face interviews. For all patients, shoulder aROM and pROM in all directions of the affected shoulder were assessed and recorded by the motion sensor device at baseline and each month for 3 months following the initiation of the rehabilitation program. All patients received and completed questionnaires that allowed relevant metrics to be evaluated at baseline and at the 1-month, 2-month, and 3-month follow-ups by an independent physiotherapist who was not involved in treating any patients and was blinded to the treatment allocation. This questionnaire was based on a pain severity visual analog scale (VAS) and the Quick Disabilities of the Arm, Shoulder, and Hand (qDASH) questionnaire. The monthly exercise completion rate, as reported by all participants and by the motion sensor devices in the motion sensor–assisted rehabilitation group, was recorded every month.

Instruments

Measurement of Shoulder Function

The self-reported qDASH, consisting of 11 items, was developed to measure the relevant physical function and symptoms of patients with upper limb disorders [19]. The Chinese version of qDASH has excellent reliability (Cronbach α=.818, intraclass correlation coefficient [ICC]=0.907) [20]. A higher qDASH score indicates poorer shoulder function.

Measurement of Pain

A VAS is an instrument that is regularly used to measure the intensity of pain [21]. A researcher asked the patients, "On a scale of 0 to 10, how severe has been the worst pain that you have experienced within the last week?" This question was repeated in the follow-up sessions.

Measurement of Exercise Completion Rate

The exercise completion rate reported by the motion sensor device was calculated by dividing the number of exercises completed daily by the number of assigned exercise tasks daily. The monthly exercise completion rate was the average of the daily exercise completion rates over an entire month. The patient-reported exercise completion rate in each month was obtained by asking each patient, "On a scale of 0%-100%, what was your average exercise completion rate in the previous month?"

Statistical Analysis

The ICC was used to calculate the interrater reliability between the shoulder aROM and pROM outcomes measured by the motion sensor device and the two experienced examiners. The ICC was obtained by conducting a 2-way analysis of variance in a random effects model. ICCs >0.8 and >0.9 were considered to indicate good and excellent reliability, respectively [22]. The ICC was further used to determine the minimal clinically important difference (MCID) in a distribution-based method. Specifically, the standard error of the measurement (SEM) was first computed using the following equation: SEM = SD of the baseline measures × $0\sqrt{1 - \text{ICC}}$; then, the MCID was computed as follows: SEM × 1.96 × $\sqrt{2}$ [23,24].

In the second part of the study, in which the home-based exercise and motion sensor–assisted rehabilitation groups were compared, the Mann-Whitney *U* test was used to compare two independent, ordinal groups, regardless of the normal distribution of the metric of interest owing to small sample sizes. The Wilcoxon signed-rank test was used for within-group comparisons of the dependent variables at follow-up. The MCID computed from the first part of the study was used to illustrate whether the changes in aROM and pROM were meaningful.

XSL•FO
RenderX

Along with descriptive statistics concerning participant characteristics, generalized estimating equations were used to compare the two groups with respect to improvements in several outcomes (including aROM, pROM, VAS, and qDASH). All generalized estimating equations were analyzed using the restricted maximum likelihood estimation. They all controlled for the effects of time during follow-up. SPSS 23.0 (IBM Corp, Armonk, NY) was used for all data analyses. $P \leq .05$ was considered to indicate statistical significance.

Results

Regarding the first part of the study on the reliability of measurement by the motion sensor device, Table 1 presents the results for shoulder aROM and pROM in all directions for healthy participants and for patients with AC, as obtained by the motion sensor device and the two examiners.

Table 1. Active and passive ranges of motion of the shoulder measured by different examiners.

Examiner	Abduction (°), mean (SD)		Flexion (°), mean (SD)		Extension (°), mean (SD)		External rotation (°), mean (SD)		Internal rotation (level)[a]	
	Active	Passive	Active	Passive	Active	Passive	Active	Passive	Active	Passive
Healthy participants (10 patients, 20 shoulders)										
Examiner 1	167.4 (9.1)	173.2 (5.2)	161 (8.4)	164.9 (5.6)	61.5 (7.9)	74.7 (8.7)	74.7 (6.6)	85.6 (4.6)	13[b], 6[c], 1[d]	18[b], 2[c]
Examiner 2	167.1 (8.5)	171.6 (6.7)	161.8 (9.9)	166 (7.1)	59.3 (5.4)	71.3 (8.0)	74.3 (8.7)	84.6 (4.8)	11[b], 9[c]	17[b], 3[c]
Motion sensor device	162.4 (9.1)	171.7 (7.3)	156.6 (9.4)	162.9 (6.4)	56.2 (7.9)	69.1 (10.8)	77.6 (9.9)	86.2 (6.2)	15[b], 5[c]	19[b], 1[c]
Patients with adhesive capsulitis (15 patients, 15 shoulders)										
Examiner 1	117.6 (18.1)	128.9 (18.1)	130.7 (13.8)	141.8 (10.5)	56.7 (11.7)	65.7 (9.5)	49.3 (11.6)	61.3 (12.1)	1[b], 2[c], 12[d]	3[b], 7[c], 5[d]
Examiner 2	118.8 (17.4)	129.5 (18.3)	132 (14.2)	144.9 (9.4)	56.3 (9.1)	64.7 (9.3)	51.4 (10.8)	60.1 (11.0)	1[b], 3[c], 11[d]	3[b], 7[c], 5[d]
Motion sensor device	118.7 (17.8)	131.7 (18.7)	133.7 (13.2)	146.7 (11.4)	55.7 (9.8)	63.9 (8.5)	49.5 (11.1)	59.6 (12.2)	1[b], 2[c], 11[d], 1[e]	3[b], 7[c], 5[d]

[a]Reported at the internal rotation level.

[b]Interscapular (T7).

[c]T12 vertebra.

[d]Waist (L3).

[e]Lumbosacral junction.

Table 2 shows whether the interrater reliability for the measurement of shoulder aROM and pROM in all directions, except for shoulder extension, as assessed by the motion sensor device and the two examiners was good or excellent (ICC range 0.899-0.979). The interrater reliability of the measurement of active and passive shoulder extension was fair to good (aROM: 0.771; pROM: 0.799).

Table 2. Interobserver reliability between angle measurements obtained by different examiners.

Values	Abduction		Flexion		Extension		External rotation		Internal rotation	
	Active	Passive	Active	Passive	Active	Passive	Active	Passive	Active	Passive
ICC[a] among examiners	0.971	0.979	0.952	0.899	0.771	0.799	0.951	0.960	0.914	0.966
95% lower limit	0.950	0.964	0.918	0.832	0.640	0.680	0.916	0.931	0.855	0.941
95% upper limit	0.984	0.989	0.974	0.944	0.867	0.885	0.973	0.978	0.952	0.981

[a]ICC: intraclass correlation coefficient.

In the second part of the study on the effectiveness of the motion sensor device in rehabilitation training, one patient in the motion sensor–assisted rehabilitation group was excluded from the analysis owing to progressive shoulder pain and newly developed shoulder abduction weakness 1 month after rehabilitation. The patient received a subsequent diagnosis of a full-thickness rotator cuff tear based on shoulder magnetic resonance imaging. Therefore, the comparison was of 7 patients in the motion sensor–assisted rehabilitation group and 7 patients in the home-based exercise group. Table 3 presents the patient demographics, which did not differ significantly at baseline between the two groups.

XSL•FO
RenderX

Table 3. Comparison of patient demographics between groups.

Demographic characteristics	Motion sensor–assisted rehabilitation group (n=7)	Home-based exercise group (n=7)	*P* value
Lesion side, n (%)			
Left	3 (42.9)	4 (57.1)	.59
Right	4 (57.1)	3 (42.9)	
Age (years), mean (SD)	53.0 (6.2)	56.1 (13.3)	.35
Gender, n (%)			
Male	4 (57.1)	5 (71.4)	.58
Female	3 (42.9)	2 (28.6)	
Education, n (%)			
Senior high	2 (28.6)	4 (57.1)	.28
Bachelor's degree and higher	5 (71.4)	3 (42.9)	
Symptom duration (months), n (%)			
3-6	4 (57.1)	3 (42.9)	.59
6-12	3 (42.9)	4 (57.1)	
Range of motion, mean (SD)			
Active abduction (°)	122.4 (15.7)	113.1 (20.3)	.48
Passive abduction (°)	126.9 (16.9)	131.6 (17.1)	.65
Active flexion (°)	138.4 (14.4)	127.9 (11.0)	.28
Passive flexion (°)	146.4 (15.7)	146.7 (7.3)	.75
Active extension (°)	56.4 (12.6)	54.1 (5.0)	.95
Passive extension (°)	63.9 (11.0)	63.1 (6.6)	.95
Active external rotation (°)	52.4 (11.6)	49.0 (9.7)	.48
Passive external rotation (°)	61.7 (14.0)	58.0 (11.9)	.52
Active internal rotation (°)[a]	1[b], 4[c], 2[d]	1[b], 2[c], 4[d]	.16
Passive internal rotation (°)[a]	2[b], 4[c], 1[d]	1[b], 4[c], 2[d]	.67
VAS[e] score, mean (SD)	5.3 (1.3)	6.1 (1.8)	.37
qDASH[f], mean (SD)	30.6 (18.1)	23.3 (7.2)	.90

[a]Reported at the internal rotation level.

[b]Interscapular (T7).

[c]T12 vertebra.

[d]Waist (L3).

[e]VAS: visual analogue scale.

[f]qDASH: Quick Disabilities of the Arm, Shoulder, and Hand questionnaire.

After 3 months of follow-up, patients in the motion sensor–assisted rehabilitation group exhibited a significant improvement from baseline in terms of shoulder aROM and pROM in all directions, pain score, and qDASH (Table 4). By contrast, patients in the home-based exercise group exhibited significant improvements only in aROM in shoulder abduction, aROM and pROM in shoulder extension, and VAS score (Table 5). Additionally, when the improvements in the motion sensor–assisted rehabilitation and home-based exercise groups were compared, the motion sensor–assisted rehabilitation group had achieved better and faster meaningful changes than the home-based exercise group.

XSL•FO
RenderX

Table 4. Improvement in parameters compared with baseline in the motion sensor–assisted rehabilitation group.

Parameters	Baseline	1 month	*P* value	2 months	*P* value	3 months	*P* value	MCID[a]
Abduction (°), mean (SD)								
Active	122.4 (15.7)	148.1 (17.5)	.04	154.1 (11.9)	.02	156.7 (9.9)	.02	5.99
Passive	126.9 (16.9)	152.9 (17.1)	.03	157.6 (10.7)	.02	161.6 (10.3)	.02	4.63
Flexion (°), mean (SD)								
Active	138.4 (14.4)	149.1 (11.4)	.02	151.0 (5.4)	.04	159.6 (8.5)	.02	6.78
Passive	146.4 (15.7)	162.0 (11.9)	.02	164.7 (5.6)	.03	170.1 (8.1)	.02	7.14
Extension (°), mean (SD)								
Active	56.4 (12.6)	62.7 (12.6)	.06	65.9 (10.8)	.03	73.6 (11.9)	.03	11.15
Passive	63.9 (11.0)	69.3 (12.2)	.11	78.1 (11.4)	.03	84.3 (6.4)	.02	11.36
External rotation (°), mean (SD)								
Active	52.4 (11.6)	66.4 (15.4)	.02	64.7 (14.8)	.02	64.6 (18.8)	.04	5.88
Passive	61.7 (14.0)	67.6 (18.6)	.20	71.9 (15.2)	.03	76.7 (15.1)	.02	4.44
Internal rotation (level)[b], mean (SD)								
Active	1^{c}, 5^{d}, 1^{e}	1^{f}, 4^{c}, 2^{d}	.01	3^{f}, 2^{c}, 2^{d}	.04	3^{f}, 2^{c}, 2^{d}	.04	—[g]
Passive	1^{f}, 3^{c}, 3^{d}	2^{f}, 4^{c}, 1^{d}	.08	3^{f}, 4^{c}	.06	3^{f}, 4^{c}	.06	—[g]
VAS[h] score, mean (SD)	5.3 (1.3)	4.1 (1.1)	.12	2.7 (1.6)	.03	2.0 (0.6)$^{+}$	.02	—[g]
qDASH[i] score, mean (SD)	30.6 (18.1)	15.5 (7.5)	.03	11.4 (9.5)	.02	9.8 (12.4)	.02	—[g]

[a]MCID: minimal clinically important difference. Computed using a distribution-based method using the standard error of measurement (SEM). SEM = SD of the baseline measurements * √1-reliability, where the reliability used was retrieved from the intraclass correlation coefficient, and MCID = SEM * 1.96 * √2.

[b]Internal rotation level.

[c]T12 vertebra.

[d]Waist (L3).

[e]Lumbosacral junction.

[f]Interscapular (T7).

[g]Not applicable.

[h]VAS: visual analogue scale.

[i]qDASH: Quick Disabilities of the Arm, Shoulder, and Hand questionnaire.

XSL•FO
RenderX

Table 5. Improvement in parameters compared with baseline and with time in the home-based exercise group.

Parameters	Baseline	1 month	*P* value	2 months	*P* value	3 months	*P* value	MCID[a]
Abduction (°), mean (SD)								
Active	113.1 (20.3)	115.9 (26.4)	.35	127.9 (21.6)	.04	134.3 (21.8)	.051	5.99
Passive	131.6 (17.1)	133.9 (15.9)	.40	135.3 (15.6)	.18	142.7 (18.2)	.13	4.63
Flexion (°), mean (SD)								
Active	127.9 (11.0)	132.3 (15.1)	.13	138.0 (13.7)	.03	140.0 (11.2)	.03	6.78
Passive	146.7 (7.3)	148.6 (9.5)	.35	152.7 (8.3)	.07	154.3 (8.1)	.13	7.14
Extension (°), mean (SD)								
Active	54.1 (5.0)	55.7 (6.9)	.50	58.3 (4.6)	.26	56.0 (5.9)	.35	11.15
Passive	63.1 (6.6)	63.3 (8.0)	.93	65.9 (6.4)	.50	68.3 (6.3)	.27	11.36
External rotation (°), mean (SD)								
Active	49.0 (9.7)	50.9 (8.7)	.46	51.1 (8.4)	.45	53.1 (12.9)	.13	5.88
Passive	58.0 (11.9)	58.7 (10.2)	.67	59.1 (10.2)	.40	60.9 (11.5)	.13	4.44
Internal rotation (level)[b], mean (SD)								
Active	1[c], 2[d], 4[e]	1[c], 2[d], 3[e],1d	.56	1[c], 2[d], 4[e]	>.99	1[c], 2[d], 3[e], 1[f]	.56	—[g]
Passive	1[c], 4[d], 2[e]	1[c], 3[d], 3[e]	.56	2[c], 3[d], 2[e]	.66	3[c], 2[d], 2[e]	.41	—[g]
VAS[h] score, mean (SD)	6.1 (1.8)	5.6 (1.7)	.46	4.1 (1.1)	.04	3.3 (1.1)	.02	—[g]
qDASH[i] score, mean (SD)	23.3 (7.2)	24.4 (18.7)	.80	19.8 (12.1)	.35	19.1 (13.7)	.25	—[g]

[a]MCID: minimal clinically important difference. Computed using a distribution-based method using the standard error of measurement (SEM). SEM = SD of the baseline measurements * √1-reliability, where the reliability used was retrieved from the intraclass correlation coefficient, and MCID = SEM * 1.96 * √2.

[b]Internal rotation level.

[c]Interscapular (T7).

[d]T12 vertebra.

[e]Waist (L3).

[f]Lumbosacral junction.

[g]Not applicable.

[h]VAS: visual analogue scale.

[i]qDASH: Quick Disabilities of the Arm, Shoulder, and Hand questionnaire.

A comparison of the improvements in the motion sensor–assisted rehabilitation and home-based exercise groups over time revealed that, compared with patients in the home-based exercise group, those in the motion sensor–assisted rehabilitation group had significantly better pROM in shoulder abduction, flexion, extension, and external rotation; better aROM in shoulder extension and internal rotation; and a better qDASH score at the 3-month follow-up (Table 6). In addition, the changes in shoulder ROM in most of the directions at different follow-up time points from baseline were correlated with improvements in qDASH score, inferring that the motion sensor device is a reliable tool for evaluating treatment efficacy in patients with AC (Multimedia Appendix 4). Table 7 shows that, compared with patients in the home-based exercise group, those in the motion sensor–assisted rehabilitation group had a significantly higher patient-reported exercise completion rate.

XSL•FO
RenderX

Table 6. Generalized estimating equation analysis for improvements between groups with time (reference: home-based exercise group).

Dependent variables	1 month, beta (SE)	*P* value	2 months, beta (SE)	*P* value	3 months, beta (SE)	*P* value
Abduction (°)						
Active	23.00 (10.49)	.03	17.00 (8.09)	.04	13.14 (9.83)	.19
Passive	23.71 (8.37)	.006	27.00 (5.49)	<.001	23.57 (8.38)	.007
Flexion (°)						
Active	6.29 (3.29)	.06	2.43 (6.00)	.69	9.00 (5.33)	.10
Passive	13.71 (2.78)	<.001	12.29 (6.10)	.049	16.14 (4.87)	.002
Extension (°)						
Active	4.71 (3.52)	.19	5.29 (3.86)	.18	15.29 (6.03)	.01
Passive	5.29 (3.57)	.14	11.57 (5.98)	.06	15.29 (5.72)	.01
External rotation (°)						
Active	12.14 (3.31)	<.001	10.14 (4.11)	.02	8.00 (4.50)	.08
Passive	5.14 (3.90)	.19	9.00 (3.12)	.006	12.14 (3.74)	.002
Internal rotation (level)[a]						
Active	1.00 (0.28)	<.001	1.14 (0.37)	.003	1.29 (0.40)	.002
Passive	0.57 (0.31)	.07	0.57 (0.41)	.17	0.43 (0.43)	.32
VAS[b] score	–0.64 (1.15)	.58	–0.57 (0.94)	.55	–0.43 (0.70)	.54
qDASH[c]	–16.19 (8.28)	.06	–15.70 (6.11)	.01	–16.60 (5.97)	.007

[a]Improvement at the internal rotation level.

[b]VAS: visual analogue scale.

[c]qDASH: Quick Disabilities of the Arm, Shoulder, and Hand questionnaire.

Table 7. Exercise completion rate reported by the participants and motion sensor device.

Time point	Motion sensor–assisted rehabilitation group		Home-based exercise group	*P* value[a]
	Recorded by motion sensor device	Reported by participants	Reported by participants	
1 month, n (%)	80.0 (21.2)	82.1 (19.1)	35.7 (18.1)	.007
2 months, n (%)	79.5 (23.2)	85.6 (19.8)	36.4 (19.1)	.007
3 months, n (%)	78.5 (26.9)	89.9 (5.5)	35.0 (17.6)	.002
Overall mean (SD)	79.3 (22.7)	86.0 (13.5)	35.7 (13.2)	.002

[a]Comparison of the participant-reported exercise completion rates between the motion sensor–assisted rehabilitation and home-based exercise groups.

Discussion

Principal Findings

This study revealed that wearable IMU-based sensors can measure shoulder ROM with acceptable reliability. When the wearable motion sensors were integrated with interactive mobile apps and telecommunication technology for the treatment of AC, patients could correctly perform daily home-based shoulder exercises under the remote supervision of a physician, as revealed by a 3-month ambulatory assessment. Consequently, compliance with the daily shoulder exercise regime was enhanced, thus improving the recovery of shoulder ROM and function among patients with shoulder AC using motion sensor device–assisted rehabilitation.

Measurement of shoulder ROM is an important component of related clinical assessments, providing information that is useful for guiding and evaluating interventions for patients with AC. Many methods can be used to measure ROM including a goniometer, 3-dimensional motion analysis (3DMA) systems, and IMU-based sensors. The goniometer is widely used but has variable reliability and tends to fail in assessing joint angles during dynamic movement [25]. The 3DMA system, which is regarded as the gold standard for evaluating lower limb kinematics [26], measures joint angles during dynamic movement with multiple degrees of freedom. However, the use of the system to measure shoulder ROM is technically challenging owing to the complexity of shoulder movement [27]. Moreover, the 3DMA system is typically immobile, difficult to access, and requires expertise to operate, preventing

XSL•FO
RenderX

its general clinical use. Wearable IMU-based sensors are becoming popular and have the potential to measure the joint angles of upper limbs with acceptable reliability [28]. Studies have already revealed that wearable IMU-based sensors can track shoulder movements with high accuracy [29]. In the present study, shoulder ROM measurement by IMU sensors yielded good to excellent interrater reliability with reports from well-experienced physicians who were using goniometers; it can therefore provide reliable angular information that physicians can use to monitor the progress of shoulder rehabilitation and facilitate the design of upper limb exercises that promote the rehabilitation of patients with AC.

An important aspect of the motion sensor device used in this study is the integration of a mobile phone–based system. Mobile phone apps for medical and rehabilitation purposes that are adaptable and easily accessible are being intensively researched [29,30]. Such apps may help in the development of a platform for delivering self-management interventions, thus improving the delivery of health care and outcomes [28]. Evidence has also revealed that app-based exercise instructions and tools can help patients to monitor their training compliance and progress with high acceptance and usability [29]. In this study, the mobile apps with the motion sensor device system provided an informative patient interface with comprehensible exercise instructions and simple progress monitoring; they support self-managed health care anywhere through the monitoring of rehabilitation exercise performance. Wireless telecommunication is used to synchronize with the Doctor App, send information regarding each patient's progress with home-based exercises, enable physicians to remotely supervise rehabilitation, and provide instant feedback to patients. Studies have established that the delivery of simple text message reminders to mobile phones increases the compliance of AC patients to a shoulder exercise regimen [31]. With the assistance of a mobile app that is used in conjunction with a motion sensor device for shoulder rehabilitation, physiotherapists or physicians can not only send text messages to remind patients to perform their daily exercises but also assign personal home-based exercises based on their training performance, thus increasing their motivation.

Overall, the positive results justify further work on motion sensor devices for treating AC. Further investigations of the usability of mobile apps and the development of auto-adjusted rehabilitation programs that are based on a user's personal performance are warranted before well-controlled research is conducted with large patient cohorts to demonstrate the viability of motion sensor devices in real-world environments. The motion sensor device–assisted rehabilitation model in this study can be applied in conjunction with telerehabilitation. Telerehabilitation is the provision of rehabilitation services at a distance; it can be image-based or sensor-based and can be delivered using virtual environments or virtual reality [32]. A motion sensor device can support image-based and sensor-based telerehabilitation using an activity recognition model, an interactive 3D avatar in mobile phone apps, and a wireless telecommunicated network, supporting medical treatment for patients with AC. The future use of motion sensor devices in telerehabilitation for various muscle skeletal disorders is expected after its development reaches maturation.

Limitations

The main limitations of this study were the relatively small number of patients, short follow-up period, and consequent lack of information on the long-term effects of the intervention. The allocation of patients was not randomized, owing to their varying degrees of familiarity with mobile apps and motion sensor devices. Studies have found that new rehabilitation technology may be unlikely to be accepted by many elderly patients [33]. In this study, the targeted patients with AC were 40-70 years old. Within this population, younger and highly educated patients are generally accepting of motion sensor device–assisted rehabilitation and so were likely to be allocated to the motion sensor–assisted rehabilitation group. Researchers of future studies should consider a training program for motion sensor device usage before enrollment to prevent any potential selection bias. Finally, although this study revealed that the IMU-based sensors that were used to measure shoulder ROM were as reliable as goniometers used by two well-experienced physicians, the true accuracy of IMU-based sensors for shoulder ROM measurement is still unclear. Evidence has confirmed that the reliability of shoulder and elbow ROM measurement using the goniometer varies (interrater ICC, 0.36-0.91) [25]. Further assessment may be required to compare IMU-based sensors and the 3DMA system in the measurement of shoulder ROM with regard to accuracy and reliability.

Conclusions

Wearable IMU-based sensors can reliably record the angular motion of shoulders. Motion sensor device–assisted home-based rehabilitation can increase patient compliance with a daily self-exercise regime and facilitate the early stage of functional recovery for patients with shoulder AC. Collectively, motion sensor device–assisted home-based rehabilitation is a useful treatment model for AC with telerehabilitation to overcome obstacles to physiotherapy at home; it provides comprehensive and easily accessible exercise instructions, increases compliance, and enhances exercise correctness through progress monitoring.

Acknowledgments

The authors acknowledge financial support from the Ministry of Science and Technology (MOST 109-2926-I-010-501, MOST 107-2314-B-010-015-MY3, MOST 109-2926-I-010-502, MOST 109-2321-B-010-005 and MOST 108-2923-B-010-002-MY3). This work was specifically supported by the "Development and Construction Plan" of the School of Medicine, National Yang-Ming University (107F-M01-0504) and Aiming for the Top University Plan, a grant from the Ministry of Education.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
BoostFix Quick Guide.
[DOCX File , 7159 KB - jmir_v22i7e17032_app1.docx]

Multimedia Appendix 2
Accuracy of Angular Measurement obtained with Single-Sensor and Dual-Sensor Systems.
[DOCX File , 2261 KB - jmir_v22i7e17032_app2.docx]

Multimedia Appendix 3
Home-based Exercise Protocol for Shoulder Adhesive Capsulitis.
[DOCX File , 839 KB - jmir_v22i7e17032_app3.docx]

Multimedia Appendix 4
Correlation between changes in qDASH score and shoulder ROM in the different directions at different time points of follow-up from baseline.
[DOCX File , 14 KB - jmir_v22i7e17032_app4.docx]

References

1. Neviaser AS, Hannafin JA. Adhesive capsulitis: a review of current treatment. Am J Sports Med 2010 Nov;38(11):2346-2356. [doi: 10.1177/0363546509348048] [Medline: 20110457]
2. Robinson CM, Seah KTM, Chee YH, Hindle P, Murray IR. Frozen shoulder. J Bone Joint Surg Br 2012 Jan;94(1):1-9. [doi: 10.1302/0301-620X.94B1.27093] [Medline: 22219239]
3. Hsu JE, Anakwenze OA, Warrender WJ, Abboud JA. Current review of adhesive capsulitis. J Shoulder Elbow Surg 2011 Apr;20(3):502-514. [doi: 10.1016/j.jse.2010.08.023] [Medline: 21167743]
4. Shaffer B, Tibone JE, Kerlan RK. Frozen shoulder. A long-term follow-up. J Bone Joint Surg Am 1992 Jun;74(5):738-746. [Medline: 1624489]
5. Neviaser AS, Neviaser RJ. Adhesive capsulitis of the shoulder. J Am Acad Orthop Surg 2011 Sep;19(9):536-542. [doi: 10.5435/00124635-201109000-00004] [Medline: 21885699]
6. Rookmoneea M, Dennis L, Brealey S, Rangan A, White B, McDaid C, et al. The effectiveness of interventions in the management of patients with primary frozen shoulder. J Bone Joint Surg Br 2010 Sep;92(9):1267-1272. [doi: 10.1302/0301-620X.92B9.24282] [Medline: 20798446]
7. Maund E, Craig D, Suekarran S, Neilson A, Wright K, Brealey S, et al. Management of frozen shoulder: a systematic review and cost-effectiveness analysis. Health Technol Assess 2012;16(11):1-264 [FREE Full text] [doi: 10.3310/hta16110] [Medline: 22405512]
8. Hanchard NCA, Goodchild L, Thompson J, O'Brien T, Davison D, Richardson C. Evidence-based clinical guidelines for the diagnosis, assessment and physiotherapy management of contracted (frozen) shoulder: quick reference summary. Physiotherapy 2012 Jun;98(2):117-120. [doi: 10.1016/j.physio.2012.01.001] [Medline: 22507361]
9. Tanaka K, Saura R, Takahashi N, Hiura Y, Hashimoto R. Joint mobilization versus self-exercises for limited glenohumeral joint mobility: randomized controlled study of management of rehabilitation. Clin Rheumatol 2010 Dec;29(12):1439-1444. [doi: 10.1007/s10067-010-1525-0] [Medline: 20585816]
10. Russell S, Jariwala A, Conlon R, Selfe J, Richards J, Walton M. A blinded, randomized, controlled trial assessing conservative management strategies for frozen shoulder. J Shoulder Elbow Surg 2014 Apr;23(4):500-507. [doi: 10.1016/j.jse.2013.12.026] [Medline: 24630545]
11. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan JL. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138 [FREE Full text] [Medline: 11154253]
12. Faber M, Andersen MH, Sevel C, Thorborg K, Bandholm T, Rathleff M. The majority are not performing home-exercises correctly two weeks after their initial instruction-an assessor-blinded study. PeerJ 2015;3:e1102 [FREE Full text] [doi: 10.7717/peerj.1102] [Medline: 26244112]
13. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil 2017 Dec 11;14(1):20 [FREE Full text] [doi: 10.1186/s12984-017-0229-y] [Medline: 28284228]
14. Ongvisatepaiboon K, Vanijja V, Chignell M, Mekhora K, Chan J. Smartphone-Based Audio-Biofeedback System for Shoulder Joint Tele-Rehabilitation. J Med Imaging Hlth Inform 2016 Aug 01;6(4):1127-1134. [doi: 10.1166/jmihi.2016.1810]

XSL•FO
RenderX

15. Timmermans AAA, Seelen HAM, Geers RPJ, Saini PK, Winter S, te Vrugt J, et al. Sensor-based arm skill training in chronic stroke patients: results on treatment outcome, patient motivation, and system usability. IEEE Trans Neural Syst Rehabil Eng 2010 Jun;18(3):284-292. [doi: 10.1109/TNSRE.2010.2047608] [Medline: 20388603]
16. Lemmens RJM, Janssen-Potten YJM, Timmermans AAA, Smeets RJEM, Seelen HAM. Recognizing complex upper extremity activities using body worn sensors. PLoS One 2015;10(3):e0118642 [FREE Full text] [doi: 10.1371/journal.pone.0118642] [Medline: 25734641]
17. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Technology in S&C: Assessing Bodyweight Squat Technique with Wearable Sensors. J Strength Cond Res 2017 Apr 15:2303-2312. [doi: 10.1519/JSC.0000000000001957] [Medline: 28426512]
18. Whelan DF, O'Reilly MA, Ward TE, Delahunt E, Caulfield B. Technology in Rehabilitation: Evaluating the Single Leg Squat Exercise with Wearable Inertial Measurement Units. Methods Inf Med 2016 Oct 26;55(6):88-94. [doi: 10.3414/ME16-02-0002] [Medline: 27782290]
19. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005 May;87(5):1038-1046. [doi: 10.2106/JBJS.D.02060] [Medline: 15866967]
20. Cao S, Zhou R, Zhou H, Chen Y, Cui H, Lu Z, et al. Reliability and validity of Simplified Chinese version of Quick Disabilities of the Arm, Shoulder, and Hand (QuickDASH) questionnaire: cross-cultural adaptation and validation. Clin Rheumatol 2019 Jul 03:3281-3287. [doi: 10.1007/s10067-019-04661-8] [Medline: 31270698]
21. Flandry F, Hunt JP, Terry GC, Hughston JC. Analysis of subjective knee complaints using visual analog scales. Am J Sports Med 1991;19(2):112-118. [doi: 10.1177/036354659101900204] [Medline: 2039061]
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
23. Su C, Ng H, Yang A, Lin C. Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. Psychol Assess 2014 Sep;26(3):980-989. [doi: 10.1037/a0036764] [Medline: 24796341]
24. Cheng C, Luh W, Yang A, Su C, Lin C. Agreement of Children and Parents Scores on Chinese Version of Pediatric Quality of Life Inventory Version 4.0: Further Psychometric Development. Applied Research Quality Life 2015 Apr 24;11(3):891-906. [doi: 10.1007/s11482-015-9405-z]
25. Muir SW, Corea CL, Beaupre L. Evaluating change in clinical status: reliability and measures of agreement for the assessment of glenohumeral range of motion. N Am J Sports Phys Ther 2010 Sep;5(3):98-110 [FREE Full text] [Medline: 21589666]
26. McGinley JL, Baker R, Wolfe R, Morris ME. The reliability of three-dimensional kinematic gait measurements: a systematic review. Gait Posture 2009 Apr;29(3):360-369. [doi: 10.1016/j.gaitpost.2008.09.003] [Medline: 19013070]
27. Rau G, Disselhorst-Klug C, Schmidt R. Movement biomechanics goes upwards: from the leg to the arm. J Biomech 2000 Oct;33(10):1207-1216. [Medline: 10899329]
28. Walmsley CP, Williams SA, Grisbrook T, Elliott C, Imms C, Campbell A. Measurement of Upper Limb Range of Motion Using Wearable Sensors: A Systematic Review. Sports Med Open 2018 Nov 29;4(1):53 [FREE Full text] [doi: 10.1186/s40798-018-0167-7] [Medline: 30499058]
29. Lin H, Chiang S, Lee K, Kan Y. An activity recognition model using inertial sensor nodes in a wireless sensor network for frozen shoulder rehabilitation exercises. Sensors (Basel) 2015 Jan 19;15(1):2181-2204 [FREE Full text] [doi: 10.3390/s150102181] [Medline: 25608218]
30. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
31. Chen H, Chuang T, Lin P, Lin Y, Chuang Y. Effects of Messages Delivered by Mobile Phone on Increasing Compliance With Shoulder Exercises Among Patients With a Frozen Shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
32. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
33. Axelrod L, Fitzpatrick G, Burridge J, Mawson S, Smith P, Rodden T, et al. The reality of homes fit for heroes: design challenges for rehabilitation technology at home. Jnl of Assistive Technologies 2009 Jul 10;3(2):35-43. [doi: 10.1108/17549450200900014]

Abbreviations

AC: adhesive capsulitis
aROM: active range of motion
3D: 3-dimensional
3DMA: 3-dimensional motion analysis
ICC: intraclass correlation coefficient

XSL•FO
RenderX

IMU: inertial measurement unit
MCID: minimal clinically important difference
pROM: passive range of motion
qDASH: Quick Disabilities of the Arm, Shoulder, and Hand
ROM: range of motion
SEM: standard error of the measurement
VAS: visual analogue scale

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by J Richards, H Lin; comments to author 13.01.20; revised version received 04.03.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Chen YP, Lin CY, Tsai MJ, Chuang TY, Lee OKS
Wearable Motion Sensor Device to Facilitate Rehabilitation in Patients With Shoulder Adhesive Capsulitis: Pilot Study to Assess Feasibility
J Med Internet Res 2020;22(7):e17032
URL: http://www.jmir.org/2020/7/e17032/
doi:10.2196/17032
PMID:32457026

©Yu-Pin Chen, Chung-Ying Lin, Ming-Jr Tsai, Tai-Yuan Chuang, Oscar Kuang-Sheng Lee. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Care Professionals' Views on Using Remote Measurement Technology in Managing Central Nervous System Disorders: Qualitative Interview Study

Jacob A Andrews[1,2], PhD; Michael P Craven[2,3], PhD; Jennifer Jamnadas-Khoda[2], MSc; Alexandra R Lang[2,4], PhD; Richard Morriss[1,2,5], MBChB, MMedSci, MD, FRCPsych; Chris Hollis[1,2,5], PhD, FRCPsych; The RADAR-CNS Consortium[6]

[1]Division of Psychiatry and Applied Psychology, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[2]NIHR MindTech Medtech Co-operative, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[3]Bioengineering Research Group, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

[4]Human Factors Research Group, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

[5]NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom

[6]The RADAR-CNS Consortium, London, United Kingdom

Corresponding Author:
Jacob A Andrews, PhD
Division of Psychiatry and Applied Psychology, Institute of Mental Health
School of Medicine
University of Nottingham
Mindtech, Institute of Mental Health
Jubilee Campus
Nottingham, NG7 2TU
United Kingdom
Phone: 44 01157484218
Email: jacob.andrews@nottingham.ac.uk

Abstract

Background: Remote measurement technologies (RMT) can be used to collect data on a variety of bio-behavioral variables, which may improve the care of patients with central nervous system disorders. Although various studies have explored their potential, prior work has highlighted a knowledge gap in health care professionals' (HCPs) perceptions of the value of RMT in clinical practice.

Objective: This study aims to understand HCPs' perspectives on using RMT in health care practice for the care of patients with depression, epilepsy, or multiple sclerosis (MS).

Methods: Semistructured interviews were conducted with 26 multidisciplinary primary and secondary care HCPs who care for patients with epilepsy, depression, or MS. Interviews were transcribed verbatim and analyzed using thematic analysis.

Results: A total of 8 main themes emerged from the analysis: (1) potential clinical value of RMT data; (2) when to use RMT in care pathways; (3) roles of health care staff who may use RMT data; (4) presentation and accessibility of data; (5) obstacles to successful use of RMT; (6) limits to the role of RMT; (7) empowering patients; and (8) considerations around alert-based systems.

Conclusions: RMT could add value to the system of care for patients with central nervous system disorders by providing clinicians with graphic summaries of data in the patient record. Barriers of both technical and human nature should be considered when using these technologies, as should the limits to the benefits they can offer.

(J Med Internet Res 2020;22(7):e17414) doi:10.2196/17414

KEYWORDS

epilepsy; multiple sclerosis; depression; medical devices; barriers; health personnel; qualitative; mobile phones; mHealth; eHealth

XSL•FO
RenderX

Introduction

Background

In a health care context, remote measurement technologies (RMT) can be used by health care professionals (HCPs) and clinical teams to collect data on a patient's health or behavior and use this to inform clinical decision making. The benefits of RMT have been explored in the management of patients with cardiac conditions [1-3], early stage dementia [4], neurological disease [5], and attention deficit hyperactivity disorder [6] as well as in behavior change [7] and monitoring for indicators of sepsis [8] among others.

The European Union H2020 Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) project [9] explores the use of RMT in the care of patients with epilepsy, depression, or multiple sclerosis (MS). These conditions were chosen for this project as exemplars of central nervous system disorders that are under-researched in relation to RMT. As part of this project, RADAR-base, a cloud-based platform, is being developed to explore the potential to receive data from patients' RMT and to provide these data to HCPs with a view to informing clinical decision making [10]. This study is part of this project, aiming to understand the clinical utility of RMT in the care of patients with epilepsy, depression, or MS.

Previous literature has demonstrated the benefits of using RMT in health care practice. A 2013 literature review of the use of RMT in cardiology identified a number of studies demonstrating reduced hospital visits in terms of both emergency and routine appointments, as well as higher survival rates, in patients who were monitored using RMT [11]. Benefits were also found in patients' relations with the care team, quality of life, and compliance with treatments. However, no quality appraisal was conducted in this narrative review.

Some studies have challenged these findings, suggesting little or no evidence of an effect of RMT on key outcomes [12,13]. Other work has highlighted a number of barriers to implementing RMT in health care practice. Erdmier et al [14] described a lack of regulatory control over wearables as well as a number of barriers to progress in implementation, including technical capability, erratic user (patient) behavior, and a lack of transparency from manufacturers. A patient and HCP-led priority setting exercise in the field of digital mental health highlighted the need to explore the impact of removing face-to-face human interaction in care pathways for mental health conditions, and raised issues of safety, effectiveness, evaluation, and inequalities [15]. These issues apply equally to the use of RMT.

Objectives

Authors of prior work in the area of RMT highlighted a need for research to investigate the value to HCPs of implementing RMT [13]. Davis et al [16] conducted a systematic review of health care staffs' views on the use of RMT in clinical practice and included 15 relevant studies. They concluded that "there is a critical need to engage end-users in the development and implementation of RMT" and highlighted that the evidence base in this area is small. This paper seeks to address these points by exploring HCPs' perspectives on the implementation of RMT in three central nervous system disorders.

Methods

Aim

This study aims to understand the perspectives of HCPs on the use of RMT in health care practice for the care of patients with depression, epilepsy, or MS.

Recruitment

We purposively recruited a sample of 26 HCPs, with the intention of covering multiple clinical roles (with representation from medical, nursing, and allied health professionals). Participants were all working in the National Health Service (NHS) in England in the care of patients with epilepsy, depression, or MS or a combination of these conditions. Participants were contacted through the professional networks of the research team members.

Procedure

A semistructured interview approach was used, with interviews lasting from 16 to 56 min (mean 30 min). An interview schedule was used to guide questioning, with ad hoc follow-up questions used to further explore salient points. Participants gave informed consent and were incentivized with a £15 (US $18) charity donation.

A total of 23 interviews were conducted one to one, whereas one interview was conducted with three participants together. Moreover, 13 interviews were conducted by phone, whereas 11 were conducted in person. All interviews were recorded from start to finish using a voice recorder. The study was approved by the University of Nottingham School of Medicine research ethics committee (ref 277-1802).

Analysis

Interviews were transcribed verbatim and analyzed using thematic analysis [17]. Data were coded and themed (by JA) using NVivo 12 (QSR International). Initial codes and themes were discussed within the research team and were iteratively renamed and reformed throughout the analysis process. No new themes emerged when the last interview was coded, and so, it is considered that the sample reached data saturation.

Results

Participants

Participants were HCPs (medical doctors, nurses, clinical psychologists, physiotherapists, and dietitians) from 13 NHS trusts (health care organizations) within England. Of the 26 participants, 8 (31%) were female. A total of 12 participants specialized in the care of patients with epilepsy, 6 in depression, 6 in MS, and 2 were general practitioners working across all 3 conditions. Participants included both primary and secondary care clinicians. Moreover, 13 of the 26 interviewees had used RMT with their patients, and 14 of the 26 said that their patients had brought data to appointments. The health care roles of the participants are presented in Table 1 along with their specialization and gender.

XSL•FO
RenderX

Table 1. Job roles, genders, and specializations of interview participants.

Clinical specialization	Depression (n=6)	Epilepsy (n=12)	Multiple sclerosis (n=6)	Generalist (n=2)	Total (N=26)
Gender (female), n (%)	1 (17)	4 (33)	2 (33)	1 (50)	8 (30)
Job role					
Psychiatrist	4 (67)	0 (0)	0 (0)	0 (0)	4 (15)
Psychologist	1 (17)	0 (0)	2 (33)	0 (0)	3 (12)
Neurologist	0 (0)	6 (50)	3 (50)	0 (0)	9 (35)
Dietician	0 (0)	1 (8)	0 (0)	0 (0)	1 (4)
Specialist nurse	1 (17)	4 (33)	0 (0)	0 (0)	5 (19)
Physiotherapist	0 (0)	0 (0)	1 (17)	0 (0)	1 (4)
General practitioner	0 (0)	1 (8)	0 (0)	2 (100)	3 (12)

Thematic Analysis

Our analysis generated 8 main themes, each of which also featured a number of subthemes.

Theme 1: Potential Clinical Value of the Remote Measurement Data

The interviews explored the types of physiological, psychosocial, and lifestyle variables that could be targets for measurement using RMT. The HCPs described uses for certain variables they considered to be potentially useful in the care of patients with 1 of the 3 conditions. Variables considered by participants to hold potential are summarized in Table 2.

Table 2. Target variables considered potentially useful to measure using remote measurement technology.

Condition	Variable
Epilepsy	Activity, anxiety, cognition, diet[a], heart rate[b], mood, quality of life, seizures (or proxies thereof), and sleep
Depression	Activity, anxiety, diet, mood, relapse signatures, sleep, and weight
Multiple sclerosis	Activity, anxiety, cognition, fatigue, mood, pain, quality of life, and visual acuity

[a]Opposing views on the value of measuring diet were offered by different participants.

[b]It was noted that heart rate would be worth measuring in epilepsy only if proven to be a proxy measure of seizures.

Epilepsy

Participants stated that it would be useful to collect data from RMT that could indicate the occurrence of a seizure or number of seizures, especially in those who have many. However, participants thought that the current approaches to seizure detection (eg, Empatica and Embrace) were limited by lack of sensitivity to detect the full range of seizure types:

> *The limitations of this particular device is that it is designed for detecting repetitive movements which is of use for tonic clonic seizures, however, there are different types of seizures which definitely don't all involve movement but are still epileptic so they wouldn't be able to detect that [...] It would be ideal to have something where all types of seizures would be recorded.* [P02]

Several types of RMT data were mentioned by participants as possible proxy measures of seizures. These were skin conductance, heart rate, accelerometry, pressure sensor readings, and electroencephalography. There were differing views on the usefulness of measuring heart rate and bed pressure as proxy measures for seizures, given the possibility of false positive signals (ie, lack of specificity).

Video and audio may be utilized remotely to assist with diagnosis and seizure identification, particularly through measurement of the sound and duration of a seizure:

> *I will sometimes tell people who struggle to video their loved one in a seizure for whatever reason, just to start the video going, put the phone down, and then it will both record the sound, and the duration of the seizure, which are probably the two most useful things we need, outside of seeing one.* [P09]

Depression

Activity data, including movement and GPS, were thought to be useful in detecting whether a patient was "leaving the house" [P21] or "getting out and about" [P06], which in turn could be considered a potential "proxy marker of depression severity" [P16]. This was also thought to have the potential to indicate a patient's global level of functioning. RMT data were also thought to offer a level of objectivity in the measurement of depression, which was otherwise lacking. Prospective mood monitoring using an electronic diary and measurement of sleep using wearables were also mentioned as potentially helpful in managing depression, where currently, systems such as Fitbits and mood diary apps could not collate this information together automatically.

XSL•FO
RenderX

Multiple Sclerosis

Participants saw potential in the use of RMT to measure fatigue, via a self-report app, and also cognition, given its association with relapse. However, some considered it difficult to pick one particular aspect of MS to measure using RMT because the symptoms experienced by patients vary:

> *I don't think you can have a particular tool that you would need to use for everyone. I think it is largely going to be dependent on the symptom profile.* [P20]

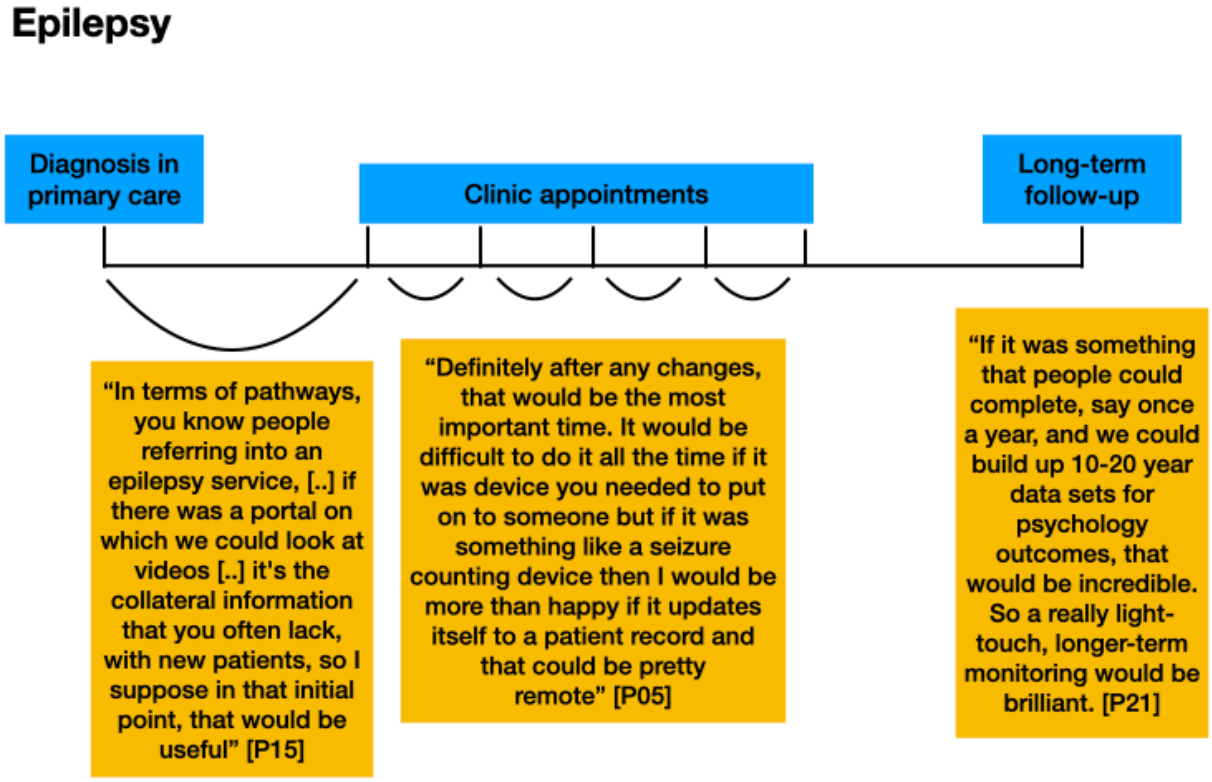
Multiple participants mentioned visual acuity as an important indicator of relapse, which could be measured using RMT, although they were unsure if technology would be able to measure it when even well-trained humans struggled:

> *It would have to be really well designed to pick up those intricacies [...] sometimes it is really difficult even for the neurologist to say this person is having a relapse or they are not.* [P19]

Theme 2: When to Use Remote Measurement Technology in the Care Pathway

Participants described different points in care pathways when data from RMT could usefully inform the care of their patients. Figures 1-3 demonstrate where participants indicated that it could be useful to receive data from RMT. Clinicians indicated that they would like data collected via RMT to be readily accessible in their electronic patient records (EPRs) when patients attended for appointments.

Figure 1. Participants' comments about the timing of the use of remote measurement technology in the epilepsy care pathway, including monitoring on a yearly basis once stable to allow assessment of follow-up and to create data for future research.



XSL•FO
RenderX

Figure 2. Participants' comments about the timing of the use of remote measurement technology in the depression care pathway, including monitoring during a change in treatment.

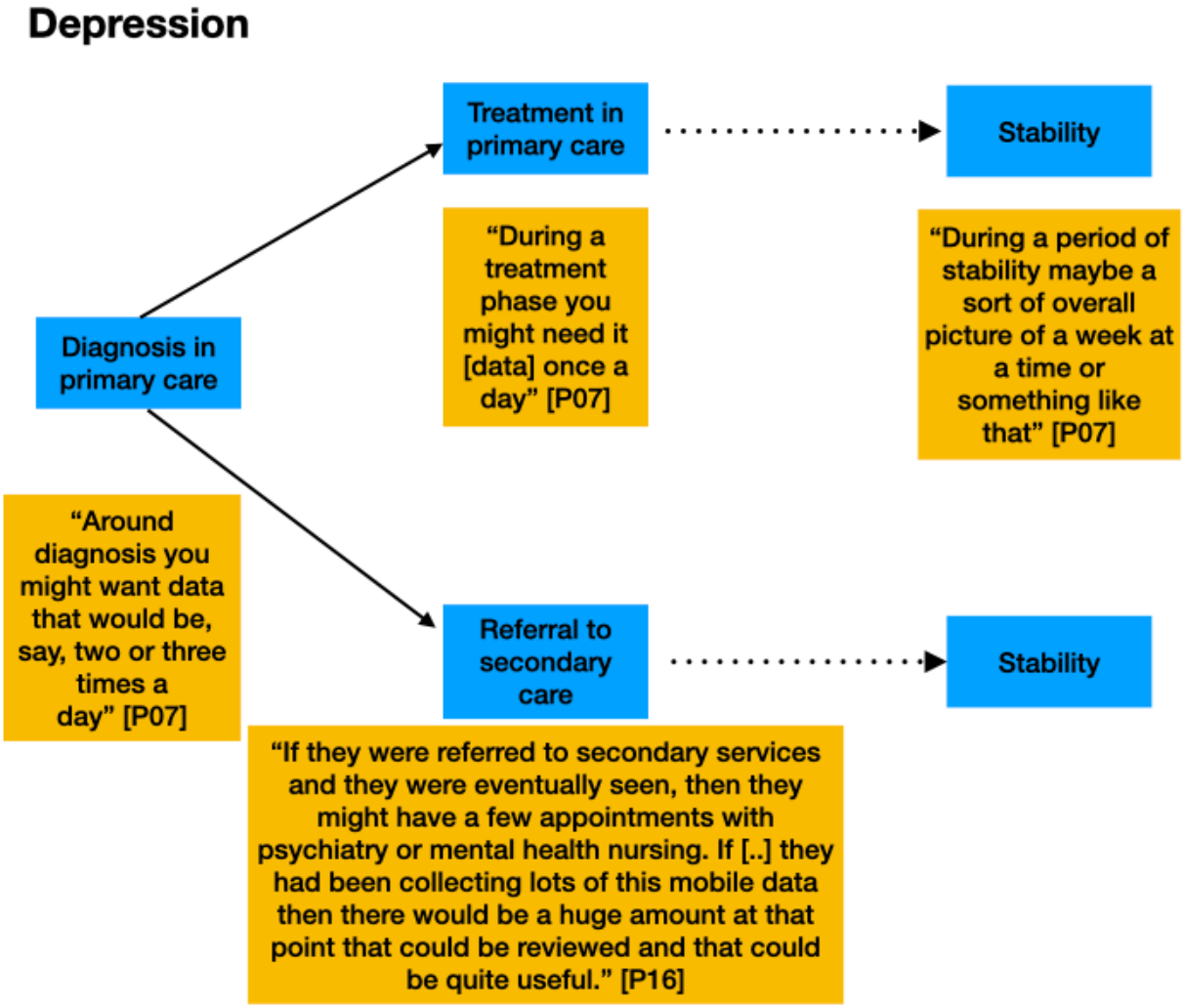
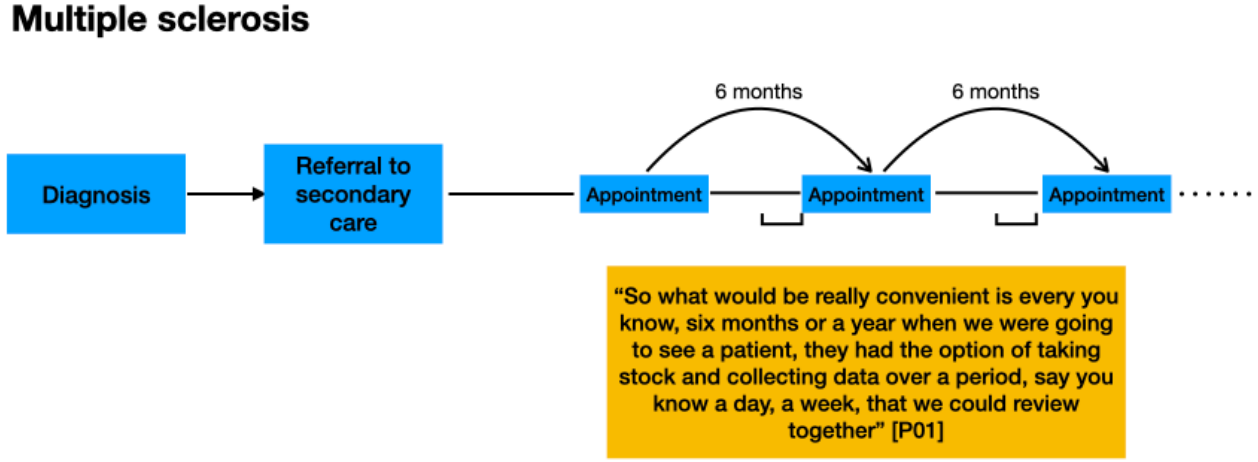


Figure 3. Participants' comments about the timing of the use of remote measurement technology in the multiple sclerosis care pathway, indicating the benefits of monitoring for a short period before an appointment.



It is acknowledged that not every patient's journey along a care pathway follows the same trajectory. These figures are for illustrative purposes only.

Theme 3: Roles of Health Care Staff Who May Use Remote Measurement Technology Data.

Participants discussed the roles of health care staff who should be involved in using any RMT data. Across all 3 conditions, participants suggested that primary care was a good place for data to be managed, given the systems available:

XSL•FO
RenderX

> *Primary care is certainly quite well set up with systems in place to action on things based through the electronic patient record.* [P06]

In secondary care, specialist nurses were considered to have the closest relationships with patients among members of the care team, and thus, they were suggested to be the team member most likely to review data from RMT. For example, in epilepsy:

> *If it's sort of data that is being in some way downloaded in between clinics, then there would certainly be a role for something like an epilepsy nurse to look at that data.* [P09]

Participants considered it important that all clinical, but not administrative, members of a team involved in the treatment of a patient should have access to RMT data when it is collected:

> *The discussions we've had are about, is it appropriate for admin staff to review that or actually does it have to be a clinician from a risk perspective [...] Your economic arguments come in about experienced clinicians are too expensive but I think making the wrong triage decision is also too expensive.* [P21]

Theme 4: Presentation and Accessibility of the Remote Measurement Technology Data

Participants had a variety of views on the best way for data to be presented to them. Particular emphasis was placed on the importance of interoperability and the ability for any data from RMT to be accessible within an existing EPR system rather than requiring the opening of another window or program:

> *If [...] you want to have the information available to you at times other than when your patient is there, then it would be good if it could feed directly into your electronic patient record.* [P07]

Across all 3 conditions, HCPs were keen to have data aggregated in a visual or graphical format. Some also recalled instances where they had used devices or software that presented information in graphs automatically and commented that this was useful for the patient as well as for the clinician:

> *For most of them [mood-reporting apps] you can do a graph function so they can show you the whole three months, in a fairly small chart, which helps us to think about if there have been any stresses or life events that have changed their mood or whether there's a pattern to the time of day, and so on.* [P07]

The speed of access to information was also considered to be a priority. Participants spoke of particular situations in which the use of digital records could increase efficiency through time saving:

> *I certainly think if you can access the information quickly then it could be a focus point for the whole consultation and it could speed things up.* [P16]

Uploading data to the EPR was considered preferable to reviewing data that patients brought to the clinic on their mobile phones:

> *It could be sent in and loaded up in the patients notes or some other big screen device otherwise you are kind of stuck with little handheld mobiles and it's not really that helpful.* [P13]

Theme 5: Obstacles to the Successful Use of Remote Measurement Technologies

Participants mentioned several aspects of the use of RMT, which they considered to be obstacles to their successful use and adoption in practice. These fitted into 2 broad categories of *technical issues* and *human issues*.

The most frequently mentioned technical issue was data accuracy ("I'm not sure they're accurate." [P03]). However, participants indicated that they would be happy to use devices even when they provided data that were not 100% accurate, so long as the clinician was aware of the margins of error that the data may contain:

> *So being as clear as possible what the potential pitfalls might be about all the data that we get back [...] I think as long as you know, kind of where it might go wrong, or how to be careful which bits to not over-interpret, then I think it's fine* [P09]

There was also concern about the interoperability of any new system with existing ones, as clinicians are already required to use several different software packages to manage patients:

> *The main trouble currently is a lack of integration.* [P01]

Data security was also discussed as an important issue to consider, although participants had differing views on the level of risk that providing data remotely may entail. Some had concerns:

> *Who has access to this data? Including if they have it on their phone, what if their phone goes missing, where does this data go? There's a big, data protection bit, there's a big big, patient safety bit.* [P03]

Some advocated taking a pragmatic approach to reduce risk while continuing to use technology where it provided a benefit:

> *I think as long as appropriate safeguards are taken then that's fine, and I think sometimes this can be a barrier, an unnecessary barrier to introducing things that can be helpful.* [P06]

Human issues considered obstacles to the use of RMT largely focused on 2 areas: patient anxiety and patient motivation. Participants discussed ways in which the use of RMT may cause patient anxiety through over focusing on their symptoms and how this in turn could be problematic for a health service:

> *We have had quite a few patients coming in that have used these monitoring devices and say my heart rate is really fast. For them it's another layer of education so it actually creates us more work.* [P03]

However, others were less concerned, believing that RMT would not induce anxiety in patients who were not prone to it in the first place or that any anxiety would be manageable:

XSL•FO
RenderX

> *We've always found ways to react to that anxiety, this is just what it looks like in the current generation.* [P21]

Participants suggested that patients with depression would have less motivation to engage with RMT as a result of their condition:

> *A lot of our patients may, especially if they're more severely depressed, not be very motivated to interact with the app* [P16]

However, the use of RMT to generate more objective evidence of a patient's health state was considered by some to be useful in motivating engagement with their care:

> *So if you do your usual interview and you've got objective evidence to say, I think your depression is coming back or you haven't been exercising enough or you have way more seizures than you think, then of course, that might help motivating them to do certain things.* [P22]

Theme 6: Limits to the Role of Remote Measurement Technology

Several of our participants mentioned elements of care in their specialism, which, in their view, should not be replaced by an RMT-enabled approach. In MS, the importance of face-to-face appointments was highlighted as essential for HCPs to identify subtle signs of worsening condition:

> *You really need to be physically examining the patient as well as hearing their perspective because there's subtle deficits that you can pick up on at examination that people won't notice day to day [...] you can pick up things like subtle signs like nystagmus, or problems with the balance or things like that, that people often won't notice.* [P01]

In epilepsy too, HCPs reiterated the importance of seeing their patients face to face:

> *If their seizure frequency has increased you're there thinking I probably need to see you, what else is going on? Have they got a cold, a water infection? Is there something else going on in their life? Are they not taking their tablets? Sometimes some of those conversations need to be had.* [P03]

A general practitioner mentioned the relational side of their work as important in the care of patients with depression:

> *The relationship element of it is very important, and obviously in primary care a lot of what I'm thinking is around depression [...] the human to human contact with someone who's struggling with mood, and the fact that you've got someone who can be empathic and rapport rather than just crunching data.* [P06]

Theme 7: Empowering Patients

HCPs believed that RMT may benefit patients because it might empower them to take a leading role in their care. Some clinicians believe that patients should lead the use of technology and, therefore, have more control over their own care:

> *The way I see it is it's more about the patient using the data for themselves, the clinician is almost the passive recipient of the data who is working with the patient to try and interpret it and help them develop techniques to use the data themselves.* [P07]

HCPs also spoke of how patients could be given full control of their own data collected using RMT and allow those data to be shared with a chosen clinician when they deemed it necessary:

> *Within the patient held database, [...] presumably a sort of secure log in, and that is, it's patient-controlled [...] and they could give out the ability to share.* [P15]

Participants provided examples of how a patient could be empowered through the use of RMT, by determining when to arrange an appointment based on the outputs from the technology:

> *Through prospective mood monitoring you could capture periods where there had been a persistent lowering of mood over two weeks or more with associated other features or even shorter periods than that, that you'd agreed as part of a relapse signature. What people could do in those instances is potentially bring appointments forward.* [P13]

However, in the case of depression, participants saw difficulties with patient motivation and thus thought it would be unlikely that patients would be able to take control of their own care:

> *[That] involves them taking a lot of responsibility for their own healthcare and I guess that may work better in some conditions, more than depression.* [P16]

For some, it was a case of providing care on an individual basis:

> *I would tailor it to what they wanted, so you will have those who are very tech savvy who don't have any time and think this will really suit me, others are very much I really want to see you doctor [...] the key is to listen to them and individualise care rather than doing tick box medicine which we sometimes do.* [P03]

Theme 8: Alert-Based Systems

There was debate across all 3 conditions in the interviews about the potential to use RMT to alert clinicians when a monitoring variable fell outside normative parameters, for example, if seizures increased in severity or frequency, if mood or activity were found to be particularly low, or if fatigue increased. The majority of participants considered such a system to be beneficial, so that interventions could be put in place as soon as possible:

> *It would be a system that had parameters set and triggered active alerts when those parameters were exceeded, I think would be the only way that I could see a lot of our consultant body engaging in it.* [P01]

However, a small number of participants thought that such a system would be problematic, principally because alerts may create excessive demand for immediate processing, interpretation, and response (eg, outside of normal working

XSL•FO
RenderX

hours), and there would not be enough health care staff available to respond to the alerts produced:

> *Outside of fixed appointments the question would be who would actually have time and headspace to actually look at what was being flagged up. You would need to really carefully think about the staffing in the NHS and mental health services.* [P13]

HCPs also perceived there to be a risk that alerts would go unnoticed in the system:

> *My worry is this data arrives and nobody looks at it for weeks, it's sitting somewhere in the ether.* [P03]

Several participants suggested that it would be more useful if the technology alerted the patient to take action through their regular treatment pathways, rather than putting the onus on the clinician:

> *I would want it to prompt the patient to make contact with me.* [P14]

Some saw a need for further research to determine the benefits of an alert-based system:

> *Unless you could do a good study and demonstrate that sending me alerts from an automated app would be helpful, then I would just want information that I could look back on when I next met with a patient face to face.* [P14]

Discussion

Principal Findings

A total of 8 themes emerged from the analysis of our interviews. The first theme covered the potential clinical value of the remote measurement data. Where RMT are currently used in health care practice, HCPs find them to be largely inaccurate, particularly in the case of epilepsy, although efforts to develop more effective ways of monitoring epilepsy are welcomed. Participants were optimistic about the future use of activity data to monitor symptoms of depression and considered that using RMT to collect measures of fatigue and cognition in patients with MS would be useful.

In theme 2, key points in care pathways for the 3 conditions were identified as times where RMT data could provide most value. These included monitoring a short period before an appointment (MS), monitoring during a change in treatment (epilepsy), and monitoring on a regular basis once a patient was in a stable condition, to allow assessment of follow-up and to create data for future research (depression, epilepsy, and MS).

The third theme considered staff roles in the management of RMT use by patients. Participants suggested that all staff involved in a patient's care should have easy access to data generated by RMT via the patient record. Participants also made it clear that triage using data from RMT should be conducted by qualified HCPs rather than by administration staff. Primary care staff and specialist nurses in secondary care were thought to be well placed to manage incoming data from patients.

With regard to the presentation of data, in theme 4, HCPs described ease and speed of access to RMT data to be important for their successful use and emphasized the importance of interoperability with the patient record. Presentation of data in graphs was mentioned as helpful for interpretation.

Theme 5 discussed obstacles to the successful use of RMT, and these included both *technical issues* such as data accuracy and data security (where views differed on the risks involved) and *human issues* such as anxiety created by monitoring (although not all participants agreed that this was an issue).

In theme 6 on the limits of RMT benefit, participants emphasized that RMT would never completely replace face-to-face appointments, particularly in depression where relationships were considered important.

The seventh theme concerned patient empowerment. HCPs expressed the value in providing patients access to their own data, enabling them to take an active role in their own care, for example, by advancing appointments where RMT data indicated it was necessary. However, there was some concern about patients with depression having the motivation to take responsibility for their own care.

Theme 8 was related to alert-based systems. Participants debated the value of such systems and highlighted the requirements for their successful use. Some thought alerts should be used to invite the patient to take action rather than alerting a clinician, due to workload concerns. The need for further research to determine the benefits of alert-based systems was also highlighted.

Comparison With Prior Work

Although prior work exploring RMT in health care has principally identified benefits and barriers to its implementation [11,14], this study has investigated HCPs' perceptions of the clinical value of implementing RMT, helping to address the knowledge gap identified by Vegesna et al [13] and Davis et al [16].

The themes emerging in this study add to findings from prior work in this area. Our findings support the work of Bruno et al [18], who highlighted that HCPs may view the management of data from digital devices as a burden. Goodrich et al [19], among others, have highlighted the importance of interoperability and a preference for data from mobile technologies to be automatically integrated into clinical records, similar to the views of our participants. Clinicians' concern about the need to respond to alert-based systems has also been raised previously [16,20].

Prior work has also emphasized the importance of face-to-face contact in the context of digital technology and mental health care [21]. A priority setting exercise for digital mental health [15] identified the need to explore the impact of removing such interactions from care pathways. Our data have shown that HCPs view face-to-face appointments as essential in the care of patients with these 3 conditions, even where RMT could provide them with detailed recent data on a patient's status. Our data show that HCPs imagine patients could be empowered to determine their own need for a clinical appointment based on data from RMT, helping to address questions around the impact of technology on access to services, which has also been identified as a research priority [15].

XSL•FO
RenderX

Davis' Technology Acceptance Model [22] describes perceived usefulness and perceived ease of use as key mediators to the successful uptake of a new technology. Our analysis highlights ways in which clinicians perceive RMT data could be useful (theme 1) as well as where there are limits or obstacles to this usefulness (themes 5 and 6). We have also identified how speed and ease of access to data are desirable for HCPs (theme 4), evidencing how perceived ease of use is applicable to this area. The analysis also raises the tailoring of care for patients using RMT (theme 7), where it was discussed that patients' perception of RMT should be that it is both useful and easy to use to motivate continued use.

Beyond the findings presented in previous work, our findings specify the types of RMT data that clinicians would value in the management of epilepsy, depression, and MS as well as the points in patient care at which these data would be of most use and the health care roles that would be best placed to manage these data.

Implications for Researchers and Developers

Findings from these themes will help to inform the development of the RADAR-CNS approach in the application of RMT for better care for epilepsy, MS, and depression. Researchers and companies developing monitoring technologies should ensure that the boundaries of accuracy of any new solution are well defined, such that clinicians can understand the level of confidence they should place in readings from such devices. As HCPs believe patients may benefit from the option to move or advance appointments based on their data, it would also be worthwhile for any mobile health solutions to link with appointment planning services, so that these can be easily accessed. In the United Kingdom, the NHS app is an example of such a system.

Strengths and Limitations

This study has several strengths. We recruited a multidisciplinary group of HCPs working in a variety of clinical staff roles in primary and secondary care. Therefore, the use of RMT was considered from multiple perspectives. The study was limited in its consideration of only three specific central nervous system disorders in one national health care system. However, the analysis has considered how insights gained from staff working in these three conditions might generalize and has permitted a deeper analysis of the three conditions mentioned. The ratio of male to female participants was high, with only 8 of 26 participants being female. Epilepsy staff were over-represented in comparison with depression and MS staff due to the convenience sampling method. However, staff were represented across most roles in the care team for each of the three conditions represented, with the exception of MS, where an MS nurse could not be recruited in the time available.

Future Directions

Although we have focused our consideration of the use of patient RMT data on an individual basis, further work could usefully explore the use of combined RMT data from groups of patients to assess risk or identify trends. The 2019 Topol Review highlighted the potential of integrating predictive analytics into diagnosis and care pathways [23], and data from RMT could feed into these approaches.

Future work should also explore the views of health service managers, commissioners, and public health representatives to understand the value that the implementation of RMT could provide from a health care system payer and management perspective, for example, in its potential to increase efficiencies and improve outcomes for different patient populations.

Given the participants' views on the importance of nurses' roles in the management of patients' use of RMT, it would be useful to conduct further research to better understand nurses' views on subsuming associated responsibilities into their roles. Although some work has explored nurses' views on their roles in the use of technology in intensive care situations [24] and telehealth for diabetes [25], to our knowledge, no studies have explored views specifically relating to RMT in central nervous system disorders.

Further work should also be completed to understand how RMT might best facilitate increased patient autonomy (as advocated in the NHS Long Term Plan [26]) and situations where this may be less appropriate or successful. The remote assessment of disease and relapse-major depressive disorder study is recruiting 600 people with a major depressive disorder to use RMT over a period of 2 years, and this study may shed light in this area [27].

Conclusions

This paper has explored the views of HCPs on using RMT in managing central nervous system disorders, specifically epilepsy, MS, and depression. The results are as follows:

- target physiological variables for measurement that clinicians believe would be useful
- points in care pathways at which clinicians perceive benefit to patients using RMT
- roles of health care staff best placed to manage incoming data
- HCPs' preferred presentation of data
- obstacles to the successful implementation of RMT
- limits to the benefits that the RMT can provide
- ways in which patients may be empowered through the use of RMT
- considerations around alert-based systems.

Our findings show the importance of early engagement and co-design with HCPs when considering user requirements and potential use cases before using RMT in clinical care pathways. HCPs believe that RMT data can add value to the care of patients with these three conditions but are not sufficient for decisions about care to be made exclusively on the basis of these data. We have demonstrated that clinicians are pragmatic about the data security risks of using RMT data with patients. Further research is required to establish how RMT data could be used on a population level to benefit patients with central nervous system disorders.

XSL•FO
RenderX

Acknowledgments

The RADAR-CNS project [28] has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant agreement no 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA; [29]). This communication reflects the views of the RADAR-CNS consortium, and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. The funding body has not been involved in the design of the study, the collection or analysis of data, or the interpretation of data. This research was supported by the National Institute for Health Research (NIHR) MindTech MedTech Co-operative, NIHR Nottingham Biomedical Research Centre, and NIHR Applied Research Collaboration East Midlands. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. Guédon-Moreau L, Mabo P, Kacet S. Current clinical evidence for remote patient management. Europace 2013 Jun;15(Suppl 1):i6-10. [doi: 10.1093/europace/eut119] [Medline: 23737234]
2. Martirosyan M, Caliskan K, Theuns DA, Szili-Torok T. Remote monitoring of heart failure: benefits for therapeutic decision making. Expert Rev Cardiovasc Ther 2017 Jul;15(7):503-515. [doi: 10.1080/14779072.2017.1348229] [Medline: 28656875]
3. Varma N, Ricci RP. Impact of remote monitoring on clinical outcomes. J Cardiovasc Electrophysiol 2015 Dec;26(12):1388-1395. [doi: 10.1111/jce.12829] [Medline: 26337400]
4. Holbø K, Bøthun S, Dahl Y. Safe Walking Technology for People With Dementia: What Do They Want? In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility. 2013 Presented at: ASSETS'13; October 21-23, 2013; Bellevue, USA. [doi: 10.1145/2513383.2513434]
5. Block VA, Pitsch E, Tahir P, Cree BA, Allen DD, Gelfand JM. Remote physical activity monitoring in neurological disease: a systematic review. PLoS One 2016;11(4):e0154335 [FREE Full text] [doi: 10.1371/journal.pone.0154335] [Medline: 27124611]
6. Simons L, Valentine AZ, Falconer CJ, Groom M, Daley D, Craven MP, et al. Developing mhealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. JMIR Mhealth Uhealth 2016 Mar 23;4(1):e31 [FREE Full text] [doi: 10.2196/mhealth.5009] [Medline: 27009498]
7. Kurti AN, Davis D, Redner R, Jarvis B, Zvorsky I, Keith DR, et al. A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change. Transl Issues Psychol Sci 2016 Jun;2(2):128-152 [FREE Full text] [doi: 10.1037/tps0000067] [Medline: 27777964]
8. Bui T, Grayson M, Hofer K, McGuire K, Morrow M, Rodammer N, et al. Remote Patient Monitoring for Improving Outpatient Care of Patients at Risk for Sepsis. In: Proceedings of the IEEE Systems and Information Engineering Design Symposium. 2016 Presented at: SIEDS'16; April 29, 2016; Charlottesville, VA, USA p. 136-141. [doi: 10.1109/SIEDS.2016.7489286]
9. Glauser T, Shinnar S, Gloss D, Alldredge B, Arya R, Bainbridge J, et al. Evidence-based guideline: treatment of convulsive status epilepticus in children and adults: report of the guideline committee of the American epilepsy society. Epilepsy Curr 2016;16(1):48-61 [FREE Full text] [doi: 10.5698/1535-7597-16.1.48] [Medline: 26900382]
10. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Hyve, RADAR-CNS Consortium. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 1;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
11. Chronaki CE, Vardas P. Remote monitoring costs, benefits, and reimbursement: a European perspective. Europace 2013 Jun;15(Suppl 1):i59-i64. [doi: 10.1093/europace/eut110] [Medline: 23737233]
12. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med 2018 Jan 15;1(1):20172. [doi: 10.1038/s41746-017-0002-4] [Medline: 31304346]
13. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
14. Erdmier C, Hatcher J, Lee M. Wearable device implications in the healthcare industry. J Med Eng Technol 2016;40(4):141-148. [doi: 10.3109/03091902.2016.1153738] [Medline: 27010250]
15. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance priority setting partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/s2215-0366(18)30296-7] [Medline: 30170964]
16. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]

XSL•FO
RenderX

17. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
18. Bruno E, Simblett S, Lang A, Biondi A, Odoi C, Schulze-Bonhage A, RADAR-CNS Consortium. Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. Epilepsy Behav 2018 Aug;85:141-149. [doi: 10.1016/j.yebeh.2018.05.044] [Medline: 29940377]
19. Goodrich DE, Buis LR, Janney AW, Ditty MD, Krause CW, Zheng K, et al. Integrating an internet-mediated walking program into family medicine clinical practice: a pilot feasibility study. BMC Med Inform Decis Mak 2011 Jun 24;11:47 [FREE Full text] [doi: 10.1186/1472-6947-11-47] [Medline: 21702957]
20. Lamothe L, Fortin J, Labbé F, Gagnon M, Messikh D. Impacts of telehomecare on patients, providers, and organizations. Telemed J E Health 2006 Jun;12(3):363-369. [doi: 10.1089/tmj.2006.12.363] [Medline: 16796505]
21. Brooks E, Turvey C, Augusterfer EF. Provider barriers to telemental health: obstacles overcome, obstacles remaining. Telemed J E Health 2013 Jun;19(6):433-437. [doi: 10.1089/tmj.2013.0068] [Medline: 23590176]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
23. NHS Health Education England. 2019. The Topol Review: Preparing the Healthcare Workforce To Deliver the Digital Future URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2019-11-05]
24. Alasad J. Managing technology in the intensive care unit: the nurses' experience. Int J Nurs Stud 2002 May;39(4):407-413. [doi: 10.1016/s0020-7489(01)00041-4] [Medline: 11909617]
25. Starren J, Tsai C, Bakken S, Aidala A, Morin PC, Hilliman C, IDEATel Consortium. The role of nurses in installing telehealth technology in the home. Comput Inform Nurs 2005;23(4):181-189. [doi: 10.1097/00024665-200507000-00004] [Medline: 16027532]
26. NHS Long Term Plan. 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2019-11-05]
27. Matcham F, di San Pietro CB, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS Consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
28. RADAR-CNS. URL: https://www.radar-cns.org/ [accessed 200-05-18]
29. Innovative medicines initiative. URL: https://www.imi.europa.eu/ [accessed 2020-05-18]

Abbreviations

EFPIA: European Federation of Pharmaceutical Industries and Associations
EPR: electronic patient record
HCP: health care professional
MS: multiple sclerosis
NHS: National Health Service
NIHR: National Institute for Health Research
RADAR-CNS: Remote Assessment of Disease and Relapse-Central Nervous System
RMT: remote measurement technology

Edited by G Eysenbach; submitted 11.12.19; peer-reviewed by C Fincham, B Eapen, T Martin; comments to author 03.03.20; revised version received 17.03.20; accepted 22.03.20; published 24.07.20.

Please cite as:
Andrews JA, Craven MP, Jamnadas-Khoda J, Lang AR, Morriss R, Hollis C, The RADAR-CNS Consortium
Health Care Professionals' Views on Using Remote Measurement Technology in Managing Central Nervous System Disorders: Qualitative Interview Study
J Med Internet Res 2020;22(7):e17414
URL: https://www.jmir.org/2020/7/e17414
doi:10.2196/17414
PMID:32706664

©Jacob A Andrews, Michael P Craven, Jennifer Jamnadas-Khoda, Alexandra R Lang, Richard Morriss, Chris Hollis, The RADAR-CNS Consortium. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Comprehensive Telestroke Network to Optimize Health Care Delivery for Cerebrovascular Diseases: Algorithm Development

Hernán Bayona[1,2*], MD, MSc; Brenda Ropero[3*], MD; Antonio José Salazar[4*], PhD; Juan Camilo Pérez[4*], MSc; Manuel Felipe Granja[3,5*], MD; Carlos Fernando Martínez[1*], MD; Juan Nicolás Useche[3*], MD, MSc

[1]Primary Stroke Center, Neurology Department, University Hospital Fundación Santa Fe de Bogotá, Bogotá DC, Colombia

[2]College of Medicine, University of Los Andes, Bogotá DC, Colombia

[3]Department of Diagnostic Imaging, University Hospital Fundación Santa Fe de Bogotá, Bogotá DC, Colombia

[4]Electrophysiology and Telemedicine Laboratory, University of Los Andes, Bogotá DC, Colombia

[5]Lyerly Neurosurgery, Baptist Health, Jacksonville, FL, United States

[*]all authors contributed equally

Corresponding Author:
Hernán Bayona, MD, MSc
Primary Stroke Center
Neurology Department
University Hospital Fundación Santa Fe de Bogotá
Calle 119 No. 7-75
Bogotá DC
Colombia
Phone: 57 16030303 ext 5370
Email: hernanbayonao@gmail.com

Abstract

Background: Health care delivery for cerebrovascular diseases is a complex process, which may be improved using telestroke networks.

Objective: The purpose of this work was to establish and implement a protocol for the management of patients with acute stroke symptoms according to the available treatment alternatives at the initial point of care and the transfer possibilities.

Methods: The review board of our institutions approved this work. The protocol was based on the latest guidelines of the American Heart Association and American Stroke Association. Stroke care requires human and technological resources, which may differ according to the patient's point of entry into the health care system. Three health care settings were identified to define the appropriate protocols: primary health care setting, intermediate health care setting, and advanced health care setting.

Results: A user-friendly web-based telestroke solution was developed. The predictors, scales, and scores implemented in this system allowed the assessment of the vascular insult severity and neurological status of the patient. The total number of possible pathways implemented was as follows: 10 in the primary health care setting, 39 in the intermediate health care setting, and 1162 in the advanced health care setting.

Conclusions: The developed comprehensive telestroke platform is the first stage in optimizing health care delivery for patients with stroke symptoms, regardless of the entry point into the emergency network, in both urban and rural regions. This system supports health care personnel by providing adequate inpatient stroke care and facilitating the prompt transfer of patients to a more appropriate health care setting if necessary, especially for patients with acute ischemic stroke within the therapeutic window who are candidates for reperfusion therapies, ultimately contributing to mitigating the mortality and morbidity associated with stroke.

(J Med Internet Res 2020;22(7):e18058) doi:10.2196/18058

KEYWORDS

stroke; telestroke; cerebrovascular disease; software; algorithms; emergency medicine; clinical pathways

XSL•FO
RenderX

Introduction

Stroke is a major source of disability and death in both developed and developing countries [1-4]. The adequate delivery of care for patients with acute stroke symptoms requires the expertise of neurologists and radiologists for timely diagnosis and treatment. Whereas hemorrhagic stroke often requires urgent surgical intervention, ischemic stroke is managed with reperfusion therapies such as thrombolysis with intravenous recombinant tissue plasminogen activator (IV rtPA) as well as early endovascular thrombectomy in the case of large vessel occlusions. These approaches have significantly improved the long-term outcomes of patients with ischemic stroke [5]. Nevertheless, patients frequently do not receive the appropriate treatment either due to the lack of available specialists to perform appropriate clinical assessments or the long distances and prolonged transfer times to stroke care centers [5]. This situation can also occur for patients located in urban areas due to delays in the referral process to health care facilities with the required stroke handling capabilities.

We have developed several strategies to improve the quality of care and speed up the transfer of patients with acute stroke symptoms from urban and rural areas to our hospital, a certified primary stroke center with thrombectomy capabilities. Our experience has shown that, by necessity, robust stroke systems should be able to assist health care providers in real-time scenarios, thus resulting in adequate transfer processes between any level of complexity in a specific health care setting.

The purpose of this work was to establish a protocol for the management of patients with acute stroke symptoms according to the available treatment alternatives at the initial point of care. This protocol was implemented as a web-based telestroke solution and is based on the guidelines from the American Heart Association and American Stroke Association (AHA/ASA) [6] for any patient with acute stroke symptoms (ie, hemorrhagic stroke, acute or chronic ischemic stroke, transient ischemic attack [TIA], stroke mimics, and large vessel occlusions). The authors considered that a detailed description of the workflow protocols, pathways, and clinical and radiological scales used in the design of our telestroke network will help other maturing countries in the development of early-stage health systems assisting patients with acute stroke symptoms.

Methods

The work presented here is part of a larger project with the following objectives: evaluation of mobile systems for head computed tomography (CT) interpretation in acute stroke patients, evaluation of the quality of stroke care in our country from a public health standpoint, and development of a telestroke network system (the subject of this article). This initiative was approved by the Institutional Review Board of our hospital and university.

The latest diagnostic and therapeutic recommendation guidelines from the American Heart Association [6], in addition to several predictors, scales, and scores that allow the assessment of the vascular insult severity and neurological status of the patient (Table 1), were evaluated in order to define the protocols of this system.

Stroke care requires human resources (eg, neurologists, radiologists, or neuroradiologists) to evaluate the risk and eligibility of patients to receive reperfusion therapies (intravenous thrombolysis or endovascular thrombectomy) and to perform invasive treatments when indicated. In addition, technological resources, such as CT, computed tomography angiography (CTA), magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), and the necessary medical supplies and equipment, are needed to administer reperfusion therapies. Different human and technological resources may be available according to a patient's point of entry into the health care system. Therefore, several possible health care settings were evaluated to define the protocols and algorithms of this system. The algorithms for the clinical workflow of the three health care settings were defined and reviewed by a group of experts in our hospital: a stroke neurologist, a general neurologist, a neuroradiologist, and two physicians from our stroke center.

XSL•FO
RenderX

Table 1. Neurological and radiological scales used in the stroke treatment processes.

Matrix or scale	Description
Glasgow Coma Scale [7,8]	Assesses the level of consciousness
National Institutes of Health Stroke Scale (NIHSS) [9-12]	Quantifies the clinical severity of ischemic stroke
Posterior circulation predictor [13]	Predicts posterior circulation involvement
ABCD2 score [14,15]	Predicts subsequent risk of stroke in patients with TIA[a] diagnosed by emergency physicians
Field Assessment Stroke Triage for Emergency Destination (FAST-ED) [16]	Determines the probability of large-vessel occlusion
Intracerebral hemorrhage (ICH) score [17-19]	Grades early hemorrhage growth in patients with intracerebral hemorrhage
Fisher scale [20]	Grades the severity of the subarachnoid hemorrhage
Modified World Federation of Neurosurgical Societies (WFNS) [21]	Grades the severity of subarachnoid hemorrhage based on the Glasgow Coma Scale
BE-FAST score [22]	Evaluate potential stroke before physician evaluation to activate the stroke code
Alberta Stroke Program Early CT Score (ASPECTS) [23]	Can estimate the infarction size of the middle cerebral artery territory
Reperfusion Therapy Risk Mitigation [24]	Assesses the absolute and relative risks of thrombolysis or thrombectomy
Thrombolysis in Cerebral Infarction (TICI) score [25]	Addresses the extent of tissue reperfusion

[a]TIA: transient ischemic attack.

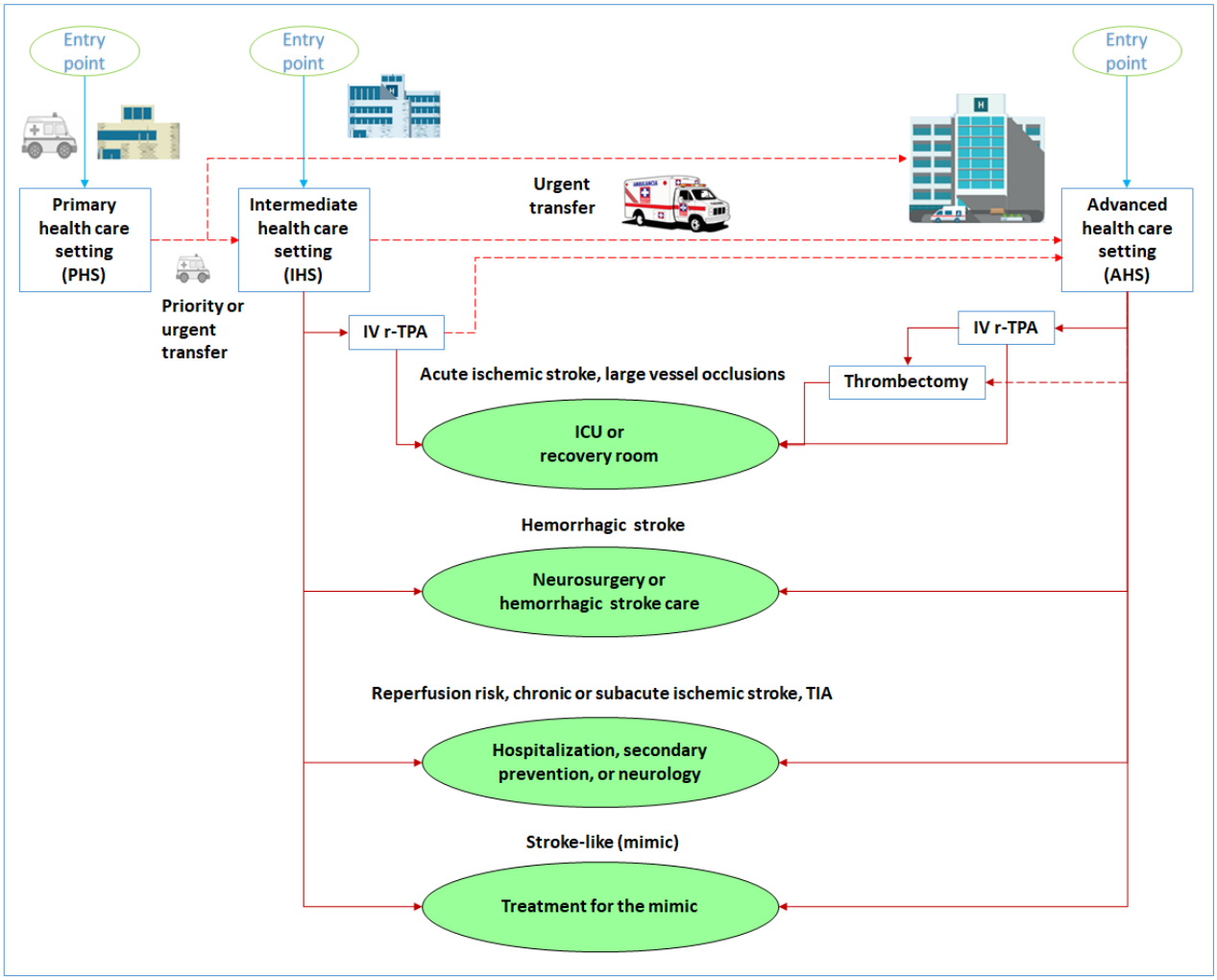
Results

The following health care settings were identified: primary health care setting, intermediate health care setting, and advanced health care setting. The interaction between these three settings is shown in Figure 1, which also shows the optimal health care settings according to a specific patient diagnosis (eg, intensive care unit [ICU], recovery room, facility with neurology or neurosurgery capabilities, regular hospitalization, or ambulatory care). IV rtPA administration can be provided in the intermediate health care setting or advanced health care setting, while thrombectomy is performed only in the advanced health care setting. In both cases, a judicious risk assessment is needed before the administration of any treatment modality [6].

In all settings, the first step is the acquisition of demographic data followed by a clinical background update, blood glucose registry, physical exam, assessment of the level of consciousness using the Glasgow Coma Scale [7], and assessment of the clinical severity of the ischemic stroke using the National Institutes of Health Stroke Scale (NIHSS) [9]. The next step is to determine the onset time to calculate the therapeutic window time (ie, the time between neurological symptom onset and patient arrival to the emergency room). In some cases, the event may be classified as wake-up or unwitnessed stroke. The therapeutic window time may be “within window” (<6 hours) or “out of window” with two possible ranges (6-24 hours or >24 hours) that determine the differences in patient management. The combination of the NIHSS and Glasgow scores as well as the time from onset to care delivery are critical breakpoints for the decision-making process performed at each of the three health care settings.

XSL•FO
RenderX

Figure 1. Interaction between the three health care settings and final possible diagnosis and referrals. ICU: intensive care unit; IV rtPA: intravenous recombinant tissue plasminogen activator; TIA: transient ischemic attack.

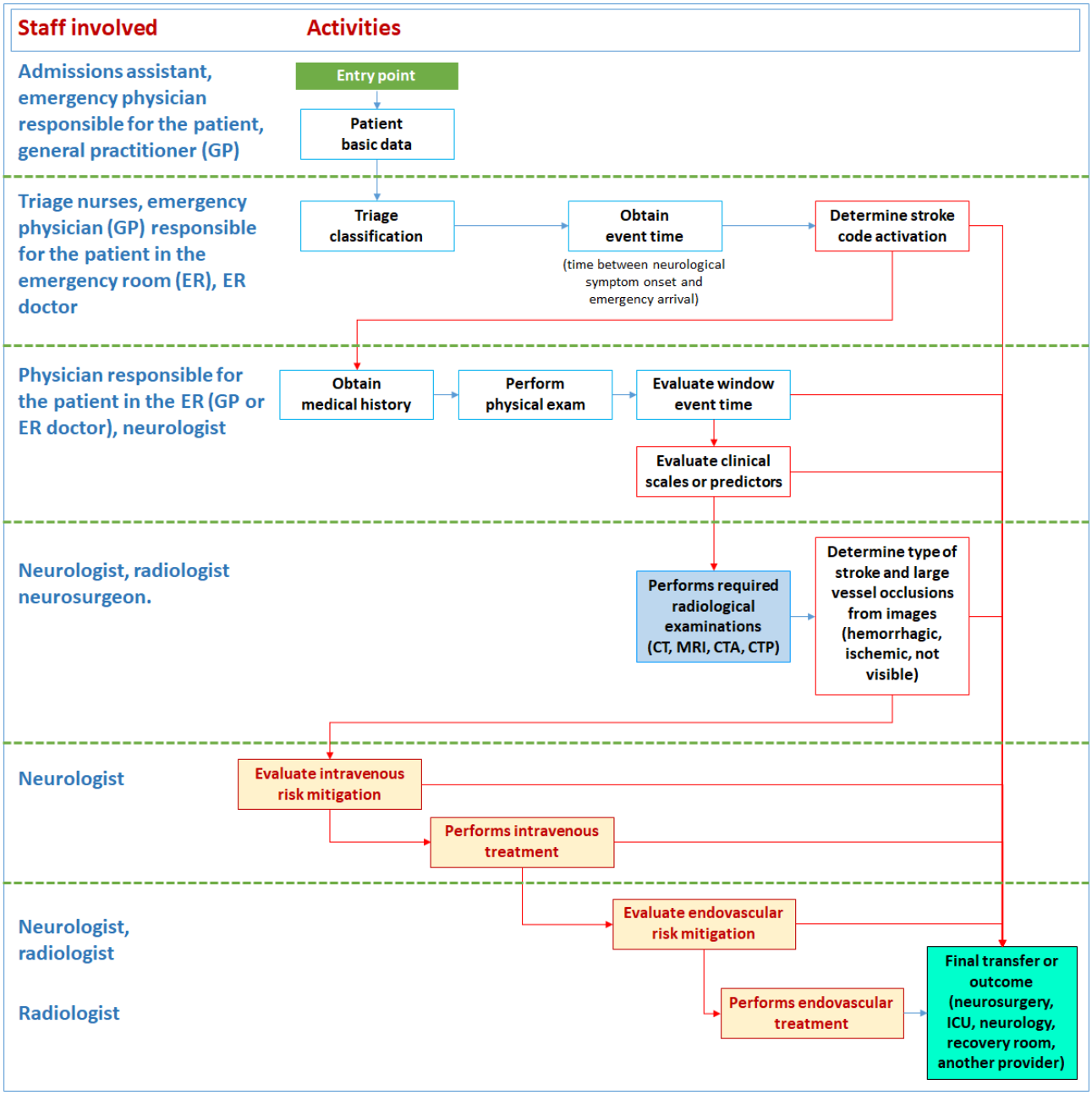


Health Care Settings

The general diagnostic and treatments steps, within each health care setting, are presented in a simplified workflow shown in Figure 2.

XSL•FO
RenderX

Figure 2. General simplified workflow for the three health care settings. CT: computed tomography; CTA: computed tomography angiography; CTP: computed tomography perfusion; ICU: intensive care unit; MRI: magnetic resonance imaging.

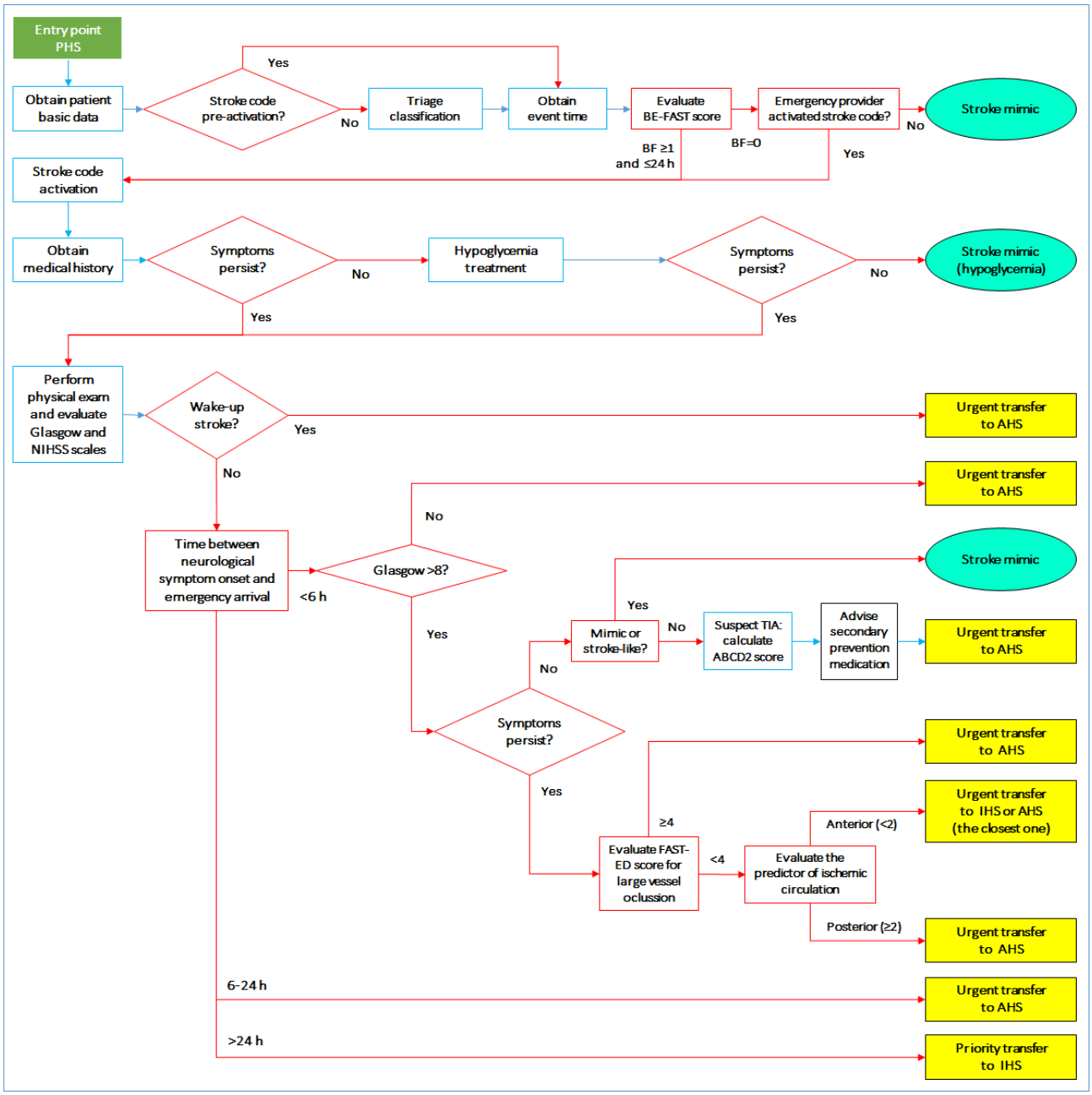


Primary Health Care Setting

In this setting, diagnostic tools are limited to the physical exam performed by a primary care physician as well as basic blood tests (eg, blood glucose). This setting may include ambulances, which may be a possible entry point to the health care system. The purpose of this setting is to provide an initial clinical assessment and determine the patient's transfers to a health care center with reperfusion capabilities, according to possible TIA, large vessel occlusion, or ischemic stroke in the anterior or posterior circulation. In this scenario, there are neither imaging facilities (CT or MRI) nor specialized health care personnel. Therefore, different clinical scales are used to assess the patient's risk at multiple levels and to predict final patient outcomes. For example, patients with possible compromised posterior circulation will be transferred to an advanced health care setting, while those with anterior circulation may be transferred to any intermediate or advanced health care site; this estimation is achieved using ischemic stroke circulation predictors [13,26]. The ABCD2 score [14] is a powerful tool that predicts the subsequent risk of stroke after a TIA, and the Field Assessment Stroke Triage for Emergency Destination score is used to determine the probability of a large-vessel occlusion [16]. The detailed workflow for this setting is shown in Figure 3.

XSL•FO
RenderX

Figure 3. Workflow for the primary health care setting (PHS). AHS: advanced health care setting; Glasgow: Glasgow Coma Scale; FAST-ED: Field Assessment Stroke Triage for Emergency Destination; IHS: intermediate health care setting; NIHSS: National Institutes of Health Stroke Scale; TIA: transient ischemic attack.



Intermediate Health Care Setting

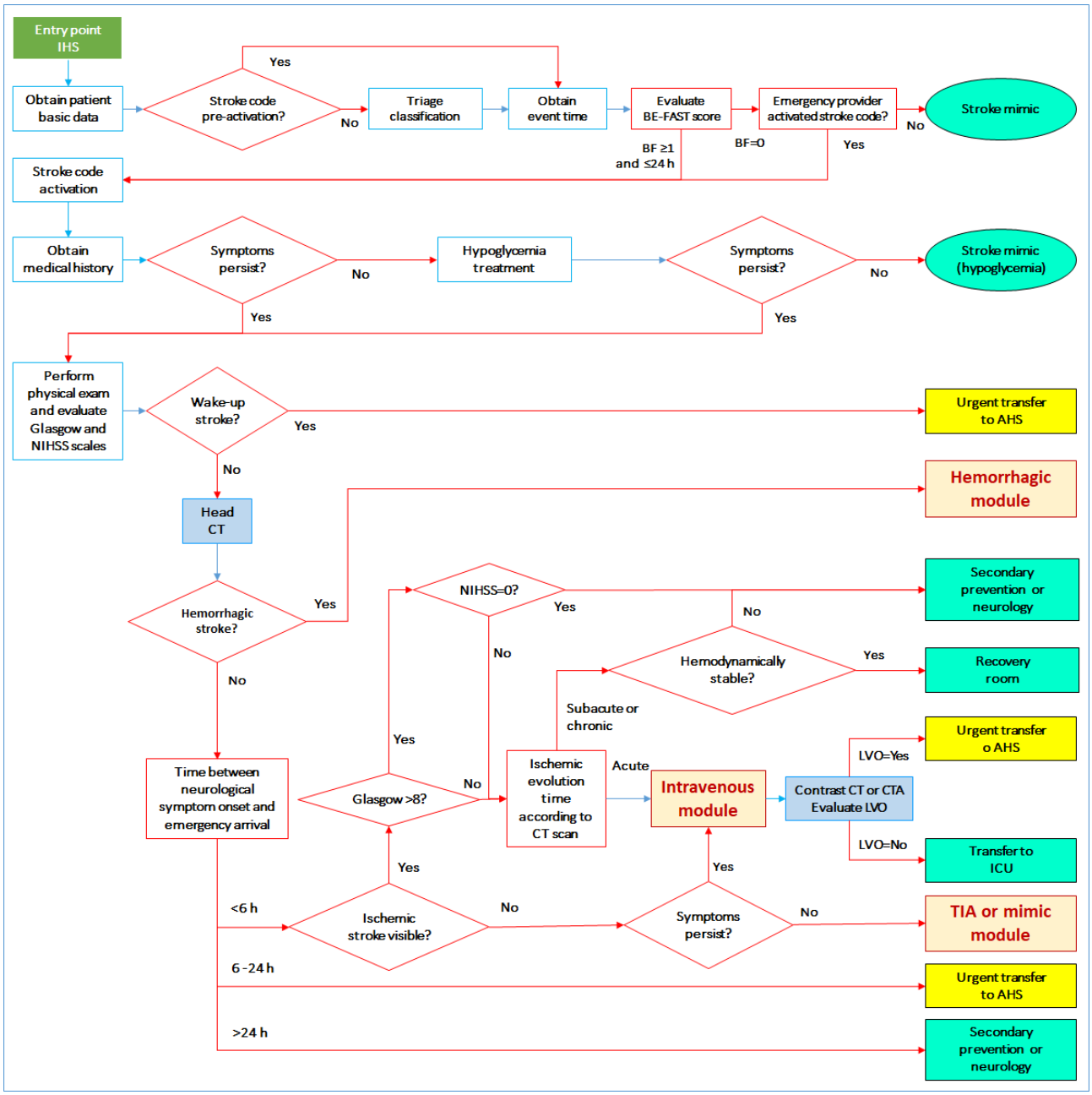
At this level of health care facility, head CT must be available to detect hemorrhagic or ischemic stroke; intravenous thrombolysis capabilities are also required at this level.

To diagnose a potential large vessel occlusion and determine if further transfer to the advanced health care setting is necessary, CTA must be available (or contrast head CT if CTA is not available). At the intermediate health care setting, both neurologists and radiologists may be available, but not fulltime.

This setting works as a mothership for urgent and priority transfers from the primary health care setting. Possible outcomes include priority transfer to an advanced health care setting for thrombectomy purposes, emergent assessment by the neurology or neurosurgery teams, IV r-TPA administration, or ambulatory care. In this setting, if a patient is eligible for IV r-TPA administration, this is done in situ, to enable early treatment, even if the patient will be transferred to the advanced health care setting. The detailed workflow for this setting is shown in Figure 4.

XSL•FO
RenderX

Figure 4. Workflow for the intermediate health care setting (IHS). AHS: advanced health care setting; CT: computed tomography; CTA: computed tomography angiography; Glasgow: Glasgow Coma Scale; ICU: intensive care unit; LVO: large vessel occlusion; NIHSS: National Institutes of Health Stroke Scale; TIA: transient ischemic attack.

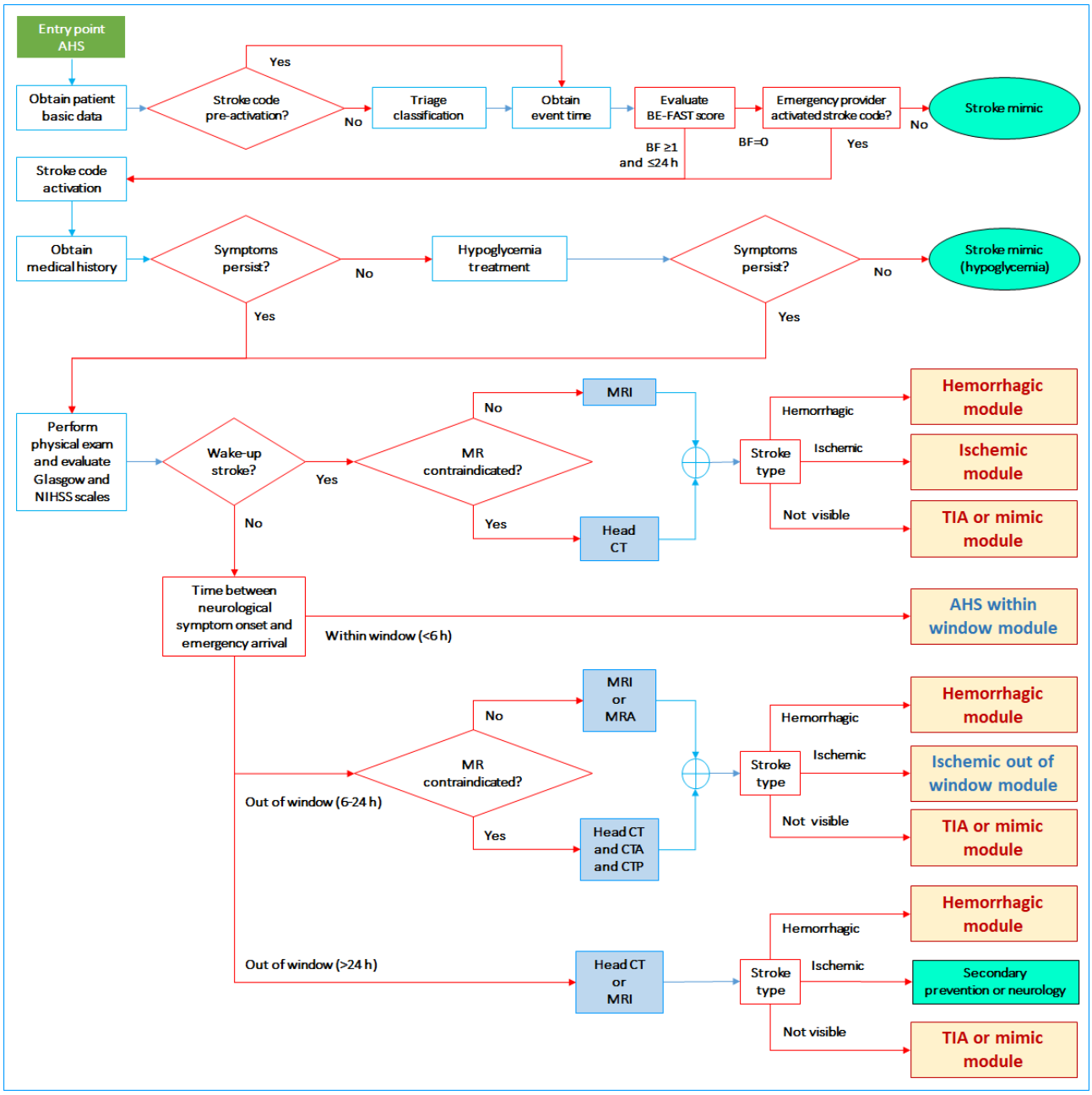


Advanced Health Care Setting

In this health care setting, specialized human and technological resources, such as stroke neurologists, neuroradiologists, CT, CTA, MRI, MRA, and the capacity for thrombolysis and mechanical thrombectomy, are available fulltime. Therefore, interfacility transfers are not necessary. In this setting, CT perfusion images are required for patients with wake-up stroke for which MRI is contraindicated. This setting receives transfers from primary health care settings and intermediate health care settings. The possible outcomes are shown in Figure 5. The workflow in the advanced health care setting for patients within the window for thrombolysis (ie, <6 hours) is shown in Figure 6. To simplify the workflow figures and render each figure on a single page, several common procedures in the intermediate and advanced health care settings were arranged in modules presented in Figures 7-10 (ie, hemorrhagic module, TIA or mimic module, intravenous thrombolysis module, endovascular treatment module, and ischemic stroke out of window module).

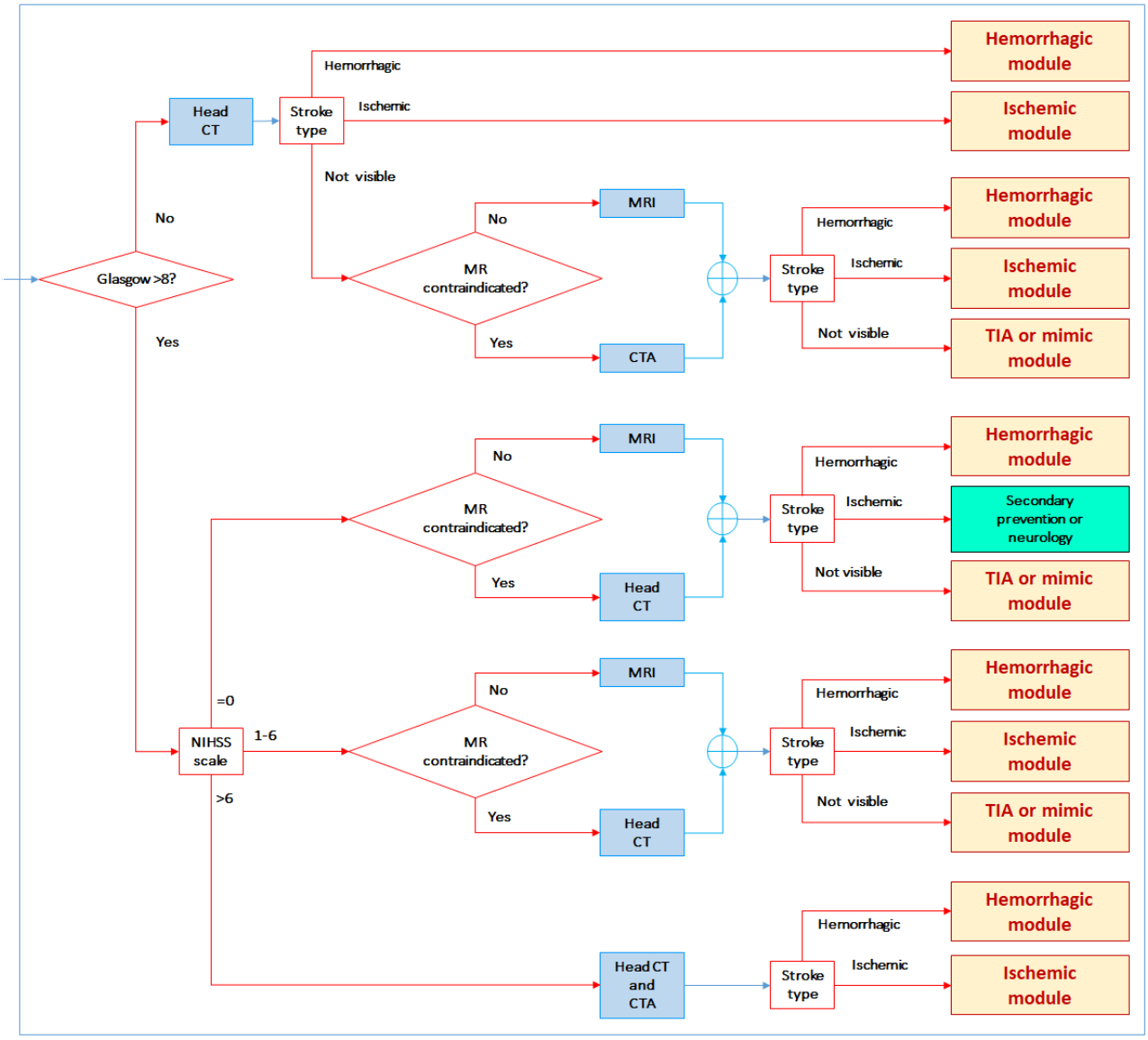
XSL•FO
RenderX

Figure 5. Workflow for the advanced health care setting (AHS). CT: computed tomography; CTA: computed tomography angiography; CTP: computed tomography perfusion; Glasgow: Glasgow Coma Scale; MR: magnetic resonance; MRA: magnetic resonance angiography; MRI: magnetic resonance imaging; NIHSS: National Institutes of Health Stroke Scale; TIA: transient ischemic attack.



XSL•FO
RenderX

Figure 6. Within window module for the advanced health care setting. CT: computed tomography; CTA: computed tomography angiography; Glasgow: Glasgow Coma Scale; MR: magnetic resonance; MRI: magnetic resonance imaging; NIHSS: National Institutes of Health Stroke Scale; TIA: transient ischemic attack.



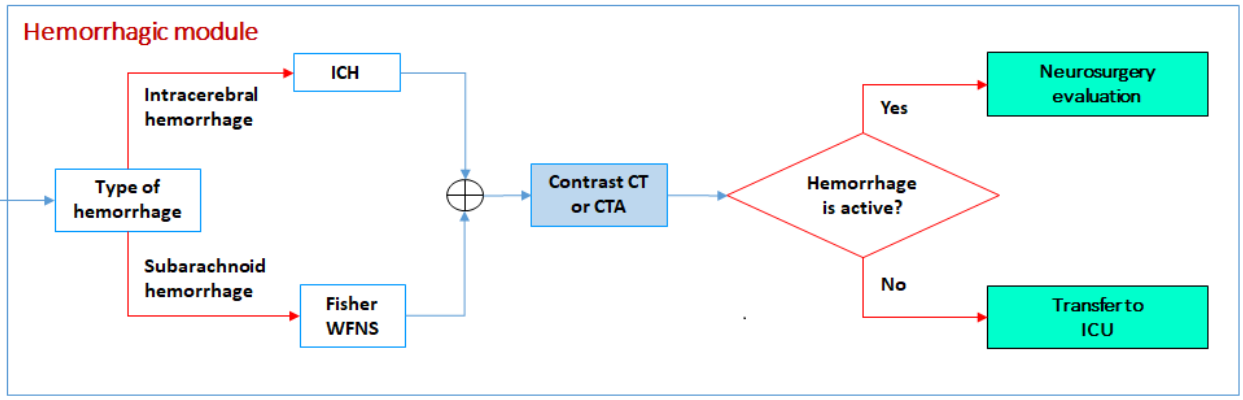
Common Modules

Hemorrhagic Module

If the imaging examination shows a hemorrhagic stroke, the Fisher scale [20] and modified World Federation of Neurosurgical Societies scale [21] are used to evaluate the severity of the subarachnoid hemorrhage. To predict mortality in patients with intracerebral hemorrhage, the intracerebral hemorrhage score [17-19] is used. CT or CTA may reveal active intracranial bleeding; in this case, the patient is referred to the neurosurgery team for urgent care. Otherwise, the patient can be stabilized and treated in the ICU (Figure 7).

XSL•FO
RenderX

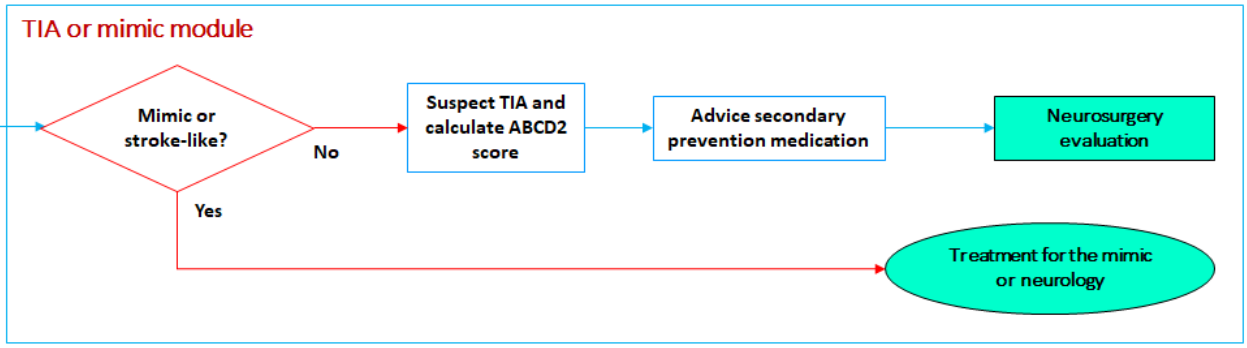
Figure 7. Hemorrhagic module. CT: computed tomography; CTA: computed tomography angiography; ICH: intracerebral hemorrhage score; ICU: intensive care unit; WFNS: Modified World Federation of Neurosurgical Societies.



TIA or Mimic Module

If the symptoms are gone or are not consistent with a vascular territory and imaging examination reveals neither a hemorrhagic stroke nor an ischemic stroke, there are two possibilities: the patient is presenting with a stroke mimic or having a TIA. In the later, the ABCD2 score [14] is calculated to evaluate the actual stroke risk. Secondary prevention using statins and antiaggregant therapy is initiated, and the patient is discharged for neurologic outpatient care (Figure 8).

Figure 8. Transient ischemic attack (TIA) or mimic module.

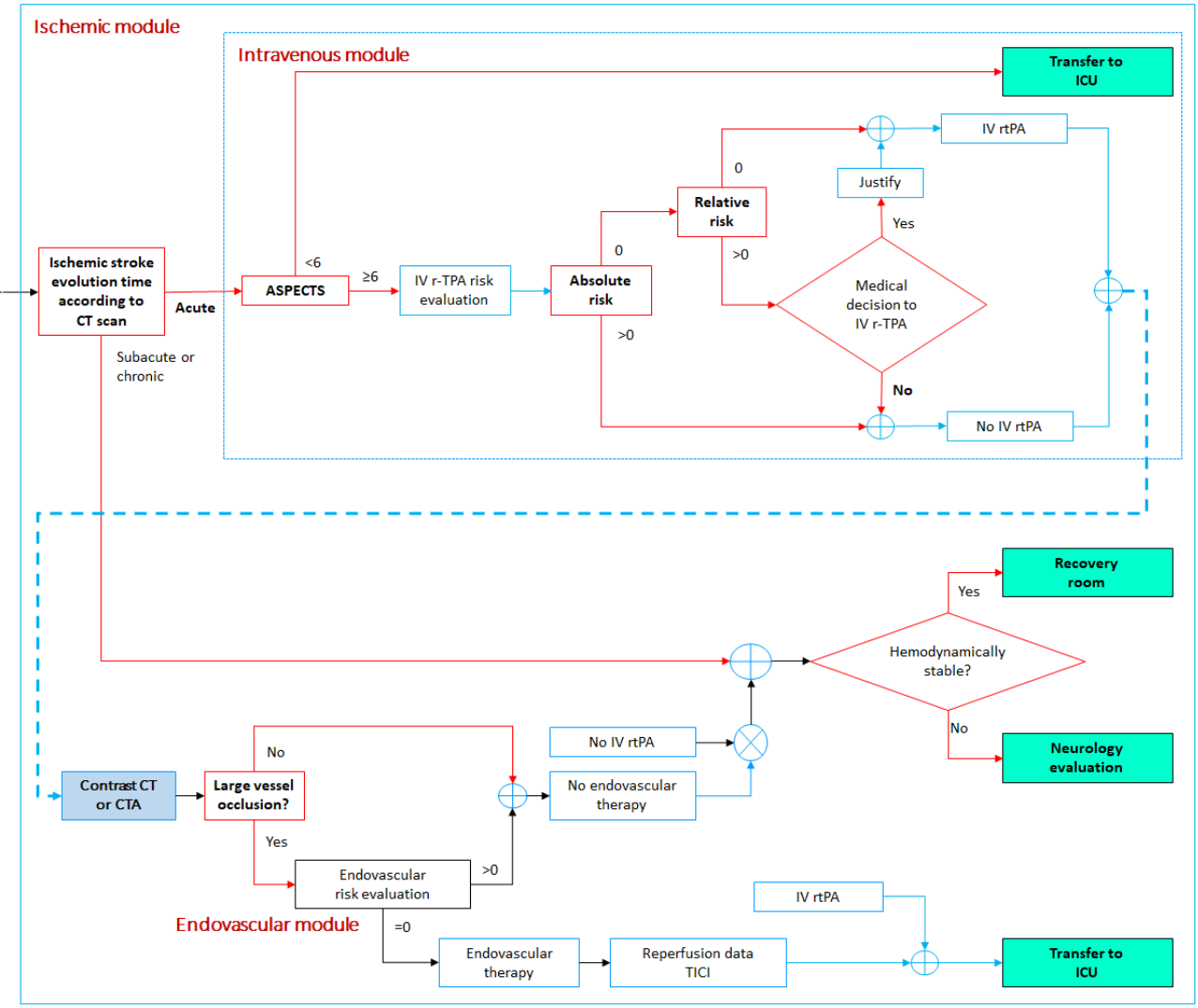


Ischemic Module

This module consists of two submodules: intravenous thrombolysis module for IV rtPA administration and endovascular treatment module, to evaluate thrombectomy treatment (Figure 9).

XSL•FO
RenderX

Figure 9. Ischemic module. ASPECTS: Alberta Stroke Program Early Computed Tomography Scan; CT: computed tomography; CTA: computed tomography angiography; ICU: intensive care unit; IV rtPA: intravenous recombinant tissue plasminogen activator; TICI: Thrombolysis in Cerebral Infarction scale.



Intravenous Thrombolysis Module

If the initial imaging examination reveals an ischemic stroke, an imaging vascular evaluation of the anterior and posterior circulation is performed according to the onset time of the ischemic insult (acute, subacute, or chronic). Patients with subacute or chronic lesions are referred to the recovery room if hemodynamically stable; otherwise, they are hospitalized for neurology assessment. For patients with acute lesions in the middle cerebral artery territory, the Alberta Stroke Program Early CT Scan (ASPECTS) is calculated; patients with ASPECTS <6 are referred to the ICU, whereas patients with ASPECTS ≥6 are evaluated in terms of the absolute and relative contraindications for IV rtPA administration (Table 2) [24]. Next, if there is no risk or only a relative risk after judicious medical assessment, intravenous thrombolysis is performed. At the same time, the patient is evaluated for the presence of large vessel occlusions and possible thrombectomy, as indicated in the endovascular treatment module (Figure 9).

XSL•FO
RenderX

Table 2. Risk mitigation matrices for reperfusion therapies.

Risk mitigation treatment	Contraindications
Intravenous thrombolysis – absolute criteria	
Clinical history	Previous ischemic stroke within 3 months; previous intracranial hemorrhage (excluded hemorrhagic transformation of a stroke); intra-axial neoplasm at this time; craniocerebral trauma or spinal within the inpatient acute period or within the 3 previous months; intracranial or spinal surgery in the last 3 months; infectious endocarditis or actual aortic dissection; extra-axial neoplasia, arteriovenous malformation, or aneurysm not excluded; arterial puncture in noncompressible location not in the last 7 days
Incoming clinical parameters	Suggestive symptoms of subarachnoid hemorrhage; blood pressure ≥185/110 mm Hg, despite management according to the protocol; blood glucose levels <50 mg/dL; active internal bleeding or active hemorrhagic diathesis
Hematologic	Platelet count <100,000/mm^3; INR[a] >1.7 and PT[b] >15 seconds or PTT[c] >40 seconds; heparin (last 48 hours with an abnormal PTT), heparin of low molecular weight in therapeutic doses (last 48 hours); inhibitors of thrombin/Xa factor in the last 48 hours
Diagnostic imaging	Evidence of acute intracranial hemorrhage, infarction size of the middle cerebral artery territory hemisphere ≥1/3 (ie, ASPECTS[d] <6)
Special cases	Start of pregnancy until 14 days postpartum; window 3-4.5 hours, ≥80 years old, diabetes, previous stroke, use of oral anticoagulant
Intravenous thrombolysis – relative criteria	
Clinical history	Surgery or major trauma in the last 14 days; gastrointestinal tract or urinary tract hemorrhage in the last 21 days; acute myocardial infarction, especially with segment ST elevation or pericarditis in the last 3 months
Incoming clinic parameters	NIHSS[e] <4 or NIHSS ≥25
Thrombectomy – absolute criteria	
Physical exam	NIHSS >29
Hematologic	INR >3, PTT >2, thrombocytopenia <30,000/mm^3
Diagnostic imaging	ASPECTS <6, infarction size (on DWI[f]) >70 mL, active hemorrhage

[a]INR: international normalized ratio.

[b]PT: prothrombin time.

[c]PTT: partial thromboplastin time.

[d]ASPECTS: Alberta Stroke Program Early Computed Tomography Scan.

[e]NIHSS: National Institutes of Health Stroke Scale.

[f]DWI: diffusion-weighted magnetic resonance imaging.

Endovascular Treatment Module

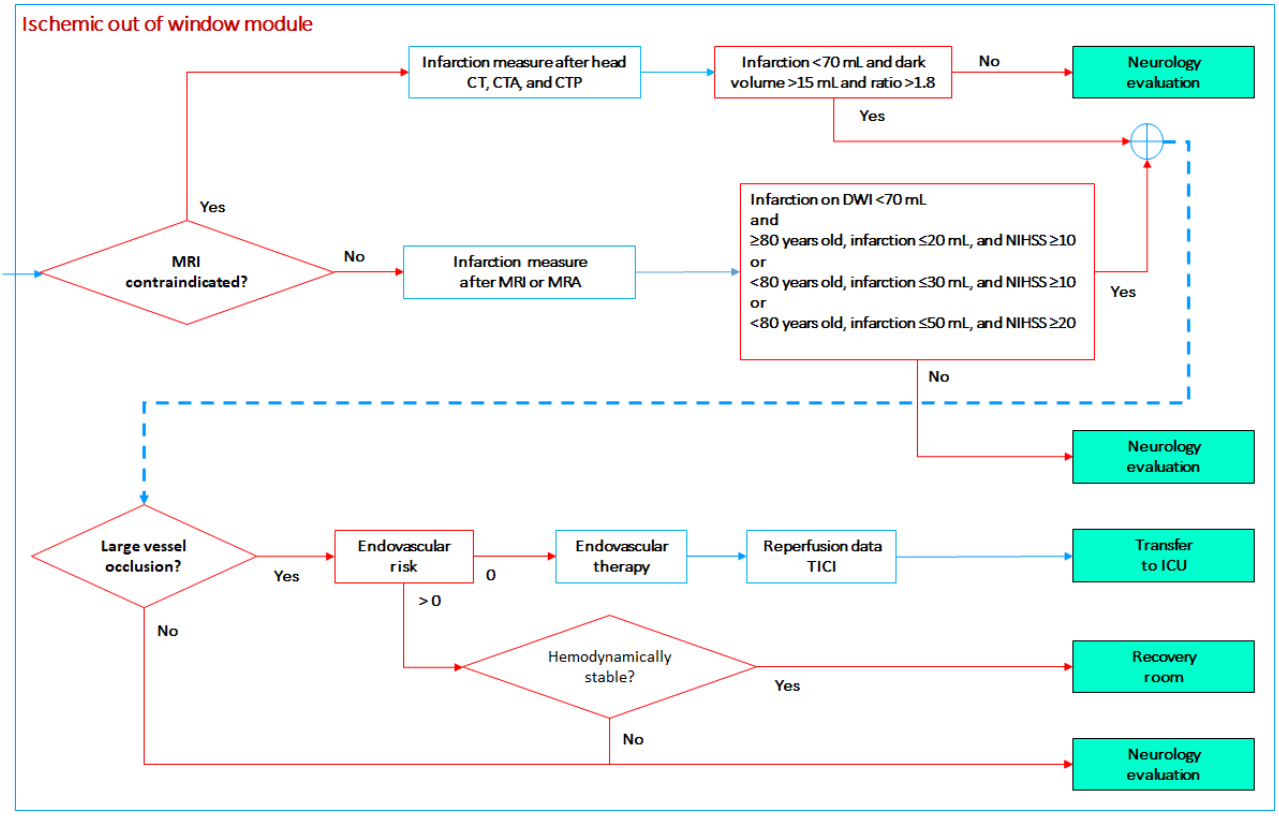
Large vessel occlusions are evaluated using contrast CT or CTA. If there are no occlusions, the patient is referred to the recovery room if hemodynamically stable; otherwise, they are hospitalized for continuous neurologic assessment. If a large vessel occlusion is confirmed, a comprehensive risk evaluation should be performed before thrombectomy [24]. Then, if no risks are identified, thrombectomy should be performed as soon as possible. After thrombectomy is performed, the degree of reperfusion is measured by means of the Thrombolysis in Cerebral Infarction score [25], and the patient is referred to the ICU (Figure 9).

Ischemic Stroke Out of Window Module

Patients arriving to an advanced health care setting after a wake-up stroke, unwitnessed stroke, or “out of window” stroke with symptom onset 6-24 hours before first medical contact may benefit from reperfusion therapies only if certain conditions are met (Figure 10). These conditions rely on the infarct volume as quantified by diffusion-weighted MRI or CT perfusion (if MRI is contraindicated), patient age, and stroke severity (NIHSS score).

XSL•FO
RenderX

Figure 10. Ischemic stroke out of window module. CT: computed tomography; CTA: computed tomography angiography; CTP: computed tomography perfusion; DWI: diffusion-weighted magnetic resonance imaging; ICU: intensive care unit; MRA: magnetic resonance angiography; MRI: magnetic resonance imaging; NIHSS: National Institutes of Health Stroke Scale; TICI: Thrombolysis in Cerebral Infarction scale.



Design of Algorithms

The three algorithms work as a handy framework for the most critical steps in the care of patients with ischemic or hemorrhagic stroke. These algorithms were based on decision trees that represent the clinical requirements and specifications of the system and consist of checklists and questionnaires [27] evaluating common physiological variables, the patient's clinical background, different predictors and scales, and specific laboratory tests according to each algorithm stage. The responses to these questionnaires determine the next step to be performed.

The predictors, scales, and scores implemented in this system allow the assessment of the vascular insult severity and neurological status of the patient. Other factors that determine patient management, either in situ or in a distant health care setting, are shown in Table 1. The clinical background, physical exam, stroke severity, and radiological findings were stored in basic modules within the clinical algorithms. These modules, as implemented in our telestroke system, are shown in Table 3. The software outputs corresponding to specific diagnoses, clinical scenarios, and transfer decisions are shown in Table 4. The checklists for risk assessment before the administration of reperfusion therapies [24] are detailed in Table 2.

XSL•FO
RenderX

Table 3. Input information modules implemented in the telestroke system.

Input data	Examples of possible data
Identification data	Name, age, sex, identification number, date of birth
Triage module	
Short physical exam	Heart rate, blood pressure, respiratory rate, weight
Triage level	1 or 2
Incoming patient medium	By their own means, referral from another service, ambulance transfer
Cincinnati scale	0-3
Stroke code activation?	Yes/no
Clinical history module	
Neurological exam	Dysarthria, hemiparesis, hemiplegia, agnosia, aphasia, dysphagia, paresthesia, mutism, headache, delirium, loss of touch, facial weakness sparing the forehead
Past relevant illnesses	Previous stroke, diabetes mellitus, dyslipidemia, coronary disease, chronic kidney disease, atrial fibrillation, high blood pressure, sleep apnea or hypopnea syndrome, smoking, thrombophilia
Relevant findings	Patient found unconscious; patient woke with loss of sensation in the one side of her body
Use of anticoagulant?	Yes/no; if yes, which one: apixaban, dabigatran, rivaroxaban, enoxaparin, warfarin, edoxaban
Laboratory	Blood glucose levels, clotting times (PT[a] and aPTT[b]), platelet count, INR[c], blood count
Physical exam module	
Cardiopulmonary resuscitation?	Yes/no
Intubation?	Yes/no
Glasgow Coma Scale	3-15
NIHSS[d]	0-37
Event time	
Time of onset of symptoms	Observed date and time or wake-up or unwitnessed stroke
Time of arrival at emergency door	YYYY/MM/DD hh:mm
Evolution time (hours)	Calculated from the previous times or typed directly
Therapeutic window type	Within therapeutic window (<6 hours), out of therapeutic window (>6 hours) or (6-24 hours; >24 hours), or wake-up or unwitnessed stroke
ABCD2 score; stroke risk after a TIA[e]	0-7; percentage (%) at 2, 7, and 90 days
Imaging evaluation (CT[f], CT perfusion, CTA[g], MRI[h])	Compromised cerebral territory, ischemic event dating (acute, subacute, chronic), ASPECTS[i], large vessel occlusions, or infarct volume, dark volume
Risk evaluation[j]	Absolute and relative evaluation for thrombolysis, risk evaluation for thrombectomy
Interventional procedure times	Door to needle time, door to inguinal puncture time
Thrombectomy	Date and time of inguinal puncture or stent implant or reperfusion; TICI[k] scale, thrombolysis reperfusion

[a]PT: prothrombin time.

[b]aPTT: activated partial thromboplastin time.

[c]INR: international normalized ratio.

[d]NIHSS: National Institutes of Health Stroke Scale.

[e]TIA: transient ischemic attack.

[f]CT: computed tomography.

[g]CTA: computed tomography angiography.

[h]MRI=magnetic resonance imaging.

[i]ASPECTS=Alberta Stroke Program Early CT Scan.

[j]See details for risk mitigation in Table 4.

[k]TICI: Thrombolysis in Cerebral Infarction.

XSL•FO
RenderX

Table 4. Output information in the telestroke system.

Module	Value suggested by the system
Diagnoses	Hemorrhagic stroke; acute, subacute, or chronic ischemic stroke; large vessel occlusions; TIA[a]; stroke mimic
Outcomes	Neurosurgery; hemorrhagic care; hospitalization or neurology (if reperfusion risk, chronic or subacute ischemic stroke, TIA); recovery room (if hemodynamic stable); ICU[b], after thrombolysis and/or thrombectomy; ambulatory care (with secondary prevention using statins or anticoagulants)
Transfer	Transfer from PHS[c] to IHS[d] or AHS[e], transfer from IHS to AHS

[a]TIA: transient ischemic attack.

[b]ICU: intensive care unit.

[c]PHS: primary health care setting.

[d]IHS: intermediate health care setting.

[e]AHS: advanced health care setting.

Software Development and Validation

The algorithms included in the three health care settings were incorporated into web-based software. Individual user profiles were created for the administrative staff and health care providers, who were assigned specific privileges.

A user-friendly interface reduces human error and assures the completeness and integrity of the information. The questionnaires implemented were straightforward and only required single-click selections instead of free-text typing for easy and rapid data input. The software was developed using the Hypertext Preprocessor and JavaScript languages and could be executed in any web browser on a laptop, tablet, or smartphone.

For data storage, a MySQL 5.1.40 database (Oracle Corporation, Redwood City, CA) was used, wherein sensitive data were encrypted (ie, patient identification). Data were stored in a structured relational database, allowing future evaluation of the system performance as well as a strong foundation for public health policies. The database included the administrative information of each facility in the telestroke network and the possible referral facilities (ie, those that have a given facility that was contracted to receive patients when a transfer is required). This information allowed the rapid selection of the most suitable stroke center according to the patient's needs after a judicious assessment of the clinical requirements and transfer times. Given that "time is brain," potential administrative pitfalls between primary health care settings and advanced health care settings also had to be considered for a quick and effective transfer. In our country, patients may be transported to various emergency departments until they are accepted in one of them, producing a critical delay in the required care known as "the death ride."

Since a patient can arrive at any given hospital and may be transferred across several health care settings without receiving adequate stroke care, a "case" starting point was defined as the time when the first medical contact was documented in the last visited hospital until the final patient outcome was reported before the patient's discharge. Therefore, when a patient is transferred to a second health care setting, all the information for the case is available in the receiving facility given that all data are stored in a server database and shared with all the facilities. This design decision allows common access to the patient's health condition at any moment from any health care setting while also avoiding the entry of redundant information. Hence, past medical history, current clinical condition, blood test results, imaging evaluations, and procedures are available in real-time for all health care facilities across the whole spectrum of stroke patient care. In addition, this allows transfer reporting to the referral facilities ahead of the patient's arrival, avoiding prolonged waiting times at emergency departments.

Software validation was performed in different phases. The first phase consisted of a simulation of the test scripts on all possible workflow pathways for each of the health care settings, which was performed to validate adequate software representation of each the algorithms. To validate the software implementation, more than 1211 test scripts were performed covering all possible pathways in each of the predefined health care settings. The second phase consisted of a retrospective registry of cases from our stroke database (nearly 600 patients in the last 5 years). The third phase was the validation of the software by neurology residents, who utilized the software while also performing a usual clinical assessment with printed forms. The final phase is to be performed between different health care facilities to test the performance of our telestroke network with real-life cases and transfers based on the information broadcast. The total number of possible pathways documented after this initial experience was as follows: 10 in the primary health care setting, 39 in the intermediate health care setting, and 1162 in the advanced health care setting.

The final system was named Telestroke-RU (copyright 13-70-240, 03/12/2018 from the National Copyright Office, Colombia) and is available for authorized users [28]. This system is not a product intended for commercial or profit uses and may be used for educational purposes.

Discussion

Principal Findings

The comprehensive telestroke platform developed in this work is the first stage to optimizing health care delivery for patients with stroke symptoms regardless of the entry point into our local emergency network in both urban and rural regions.

XSL•FO
RenderX

This system supports health care personnel by providing adequate stroke care and facilitating the prompt transfer of patients to a more appropriate health care setting according to the specific cerebrovascular disease at presentation. This system facilitates stroke care delivery for patients with acute ischemic stroke within the therapeutic window who are candidates for reperfusion therapies. Therefore, the system will contribute to mitigating the well-known mortality and morbidity associated with stroke.

Further evaluations will be performed to assess the true impact of this tool in terms of reductions in critical time windows, such as the time between symptom onset and reperfusion, door to needle time, primary health care setting to advanced health care setting transfer times, discharge clinical outcomes, accuracy of the final diagnosis, and the clinical outcomes of patients at 30 and 90 days using the modified Rankin scale [29].

Comparison With Prior Work

To the best of our knowledge, in our country, there are no software tools for the assessment and management of patients with stroke symptoms. Worldwide, smartphone apps and web-based tools are available [30-34]. These solutions were designed for acute ischemic stroke care, for triage protocols, and as an aid for transfers when reperfusion therapies are needed (eg, Field Assessment Stroke Triage for Emergency Destination score) [30] as well as the delivery of efficient inpatient or ambulatory stroke care [32,34]. Other tools have been developed with the purpose of the evaluation of specific clinical scales or radiological scores [33]; these tools use the same algorithm independent of the level of resources available at the entry point to the health care system. In contrast, our system integrates 12 clinical and radiological scales, scores, and predictors according to the specific health care setting or the referral facility to provide a specific diagnosis (hemorrhagic stroke; acute, subacute, or chronic ischemic stroke; TIA; stroke mimic). Our system allows either immediate treatment or further transfer to the appropriate health care setting. Since our solution is not integrated with the hospital electronic health record (EHR) system, it can be used in all health care services, independent of the EHR system used at each facility. It is worth mentioning that, in our country, EHR systems are not available at several rural primary health care settings.

Limitations

The continuous improvement of evidence-based stroke care guidelines motivates the continuous review of health care setting algorithms and, therefore, software updates. Further work includes the use of GPS and traffic applications to calculate the actual duration of real-time patient transfers and adequate selection of the referral facility with the shorter transfer time. In the short-term, this system will be migrated to a smartphone app to allow for a greater number of system users in a friendlier interface.

Conclusions

The implementation of this system in a telestroke network contributes to the fulfillment of and adherence to recently published stroke care guidelines, providing evidence-based practice, improving patient outcomes, and supporting the achievement of several requirements to achieve and maintain primary stroke center certification.

This telestroke system allows the assessment of different therapeutic alternatives according to the specific patient's clinical condition, thus improving efficiency and providing high-quality delivery of care. Finally, the epidemiological information stored in the database will inform public health care policies to design and implement better national policies for remote regions with significant underreporting of acute cerebrovascular diseases.

Acknowledgments

We thank our institutions and the National Department of Science, Technology and Innovation of Colombia for funding this study (Grant 1204-744-55680).

Conflicts of Interest

None declared.

References

1. World Health Organization. Stroke--1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. Stroke 1989 Oct;20(10):1407-1431. [doi: 10.1161/01.STR.20.10.1407] [Medline: 2799873]
2. Taylor TN, Davis PH, Torner JC, Holmes J, Meyer JW, Jacobson MF. Lifetime Cost of Stroke in the United States. Stroke 1996 Sep;27(9):1459-1466. [doi: 10.1161/01.str.27.9.1459]
3. Bonita R. Epidemiology of stroke. The Lancet 1992 Feb;339(8789):342-344. [doi: 10.1016/0140-6736(92)91658-u]
4. Saposnik G, Del Brutto OH. Stroke in South America. Stroke 2003 Sep;34(9):2103-2107. [doi: 10.1161/01.str.0000088063.74250.db]
5. Kulcsar M, Gilchrist S, George MG. Improving stroke outcomes in rural areas through telestroke programs: an examination of barriers, facilitators, and state policies. Telemed J E Health 2014 Jan;20(1):3-10. [doi: 10.1089/tmj.2013.0048] [Medline: 24286197]

XSL•FO
RenderX

6. Powers W, Rabinstein A, Ackerson T, Adevoe O, Bambakidis N, Becker K. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Journal of Vascular Surgery 2018 Jun;67(6):1934. [doi: 10.1016/j.jvs.2018.04.007]
7. Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, Murray G. The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol 2014 Aug;13(8):844-854. [doi: 10.1016/S1474-4422(14)70120-6] [Medline: 25030516]
8. Kothari R, Jauch E, Broderick J, Brott T, Sauerbeck L, Khoury J, et al. Acute stroke: delays to presentation and emergency department evaluation. Ann Emerg Med 1999 Jan;33(1):3-8. [doi: 10.1016/s0196-0644(99)70431-2] [Medline: 9867880]
9. Kwah LK, Diong J. National Institutes of Health Stroke Scale (NIHSS). J Physiother 2014 Mar;60(1):61 [FREE Full text] [doi: 10.1016/j.jphys.2013.12.012] [Medline: 24856948]
10. Kothari R, Hall K, Brott T, Broderick J. Early stroke recognition: developing an out-of-hospital NIH Stroke Scale. Acad Emerg Med 1997 Oct;4(10):986-990 [FREE Full text] [doi: 10.1111/j.1553-2712.1997.tb03665.x] [Medline: 9332632]
11. Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA Stroke Study Group. Stroke 1999 Nov;30(11):2355-2359. [doi: 10.1161/01.str.30.11.2355] [Medline: 10548670]
12. Broderick JP, Lu M, Kothari R, Levine SR, Lyden PD, Haley EC, et al. Finding the most powerful measures of the effectiveness of tissue plasminogen activator in the NINDS tPA stroke trial. Stroke 2000 Oct;31(10):2335-2341. [doi: 10.1161/01.str.31.10.2335] [Medline: 11022060]
13. Tarnutzer AA, Berkowitz AL, Robinson KA, Hsieh Y, Newman-Toker DE. Does my dizzy patient have a stroke? A systematic review of bedside diagnosis in acute vestibular syndrome. CMAJ 2011 Jun 14;183(9):E571-E592 [FREE Full text] [doi: 10.1503/cmaj.100174] [Medline: 21576300]
14. Ghandehari K, Ahmadi F, Ebrahimzadeh S, Shariatinezhad K, Ghandehari K. Assessment of ABCD² scale in patients with transient ischaemic attack or stroke. Neurol Neurochir Pol 2012;46(5):421-427. [doi: 10.5114/ninp.2012.31351] [Medline: 23161185]
15. Josephson SA, Sidney S, Pham TN, Bernstein AL, Johnston SC. Higher ABCD2 score predicts patients most likely to have true transient ischemic attack. Stroke 2008 Nov;39(11):3096-3098. [doi: 10.1161/STROKEAHA.108.514562] [Medline: 18688003]
16. Lima FO, Silva GS, Furie KL, Frankel MR, Lev MH, Camargo , et al. Field Assessment Stroke Triage for Emergency Destination: A Simple and Accurate Prehospital Scale to Detect Large Vessel Occlusion Strokes. Stroke 2016 Aug;47(8):1997-2002 [FREE Full text] [doi: 10.1161/STROKEAHA.116.013301] [Medline: 27364531]
17. Brott T, Broderick J, Kothari R, Barsan W, Tomsick T, Sauerbeck L, et al. Early hemorrhage growth in patients with intracerebral hemorrhage. Stroke 1997 Jan;28(1):1-5. [doi: 10.1161/01.str.28.1.1] [Medline: 8996478]
18. Hemphill JC, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. Stroke 2001 Apr;32(4):891-897. [doi: 10.1161/01.str.32.4.891] [Medline: 11283388]
19. Zuccarello M, Brott T, Derex L, Kothari R, Sauerbeck L, Tew J, et al. Early surgical treatment for supratentorial intracerebral hemorrhage: a randomized feasibility study. Stroke 1999 Sep;30(9):1833-1839. [doi: 10.1161/01.str.30.9.1833] [Medline: 10471432]
20. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. Neurosurgery 1980 Jan;6(1):1-9. [doi: 10.1227/00006123-198001000-00001] [Medline: 7354892]
21. Sano H, Satoh A, Murayama Y, Kato Y, Origasa H, Inamasu J, members of the 38 registered institutionsWFNS Cerebrovascular Disease & Treatment Committee. Modified World Federation of Neurosurgical Societies subarachnoid hemorrhage grading system. World Neurosurg 2015 May;83(5):801-807. [doi: 10.1016/j.wneu.2014.12.032] [Medline: 25535068]
22. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati Prehospital Stroke Scale: reproducibility and validity. Ann Emerg Med 1999 Apr;33(4):373-378. [doi: 10.1016/s0196-0644(99)70299-4] [Medline: 10092713]
23. Pexman JH, Barber PA, Hill MD, Sevick RJ, Demchuk AM, Hudon ME, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. AJNR Am J Neuroradiol 2001 Sep;22(8):1534-1542 [FREE Full text] [Medline: 11559501]
24. Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJB, Demaerschalk BM, American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Peripheral Vascular Disease, Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013 Mar;44(3):870-947 [FREE Full text] [doi: 10.1161/STR.0b013e318284056a] [Medline: 23370205]
25. Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, von Kummer R, Saver JL, Cerebral Angiographic Revascularization Grading (CARG) Collaborators, STIR Revascularization working group, STIR Thrombolysis in Cerebral Infarction (TICI) Task Force. Recommendations on angiographic revascularization grading standards for acute ischemic stroke: a consensus statement. Stroke 2013 Sep;44(9):2650-2663 [FREE Full text] [doi: 10.1161/STROKEAHA.113.001972] [Medline: 23920012]
26. Runchey S, McGee S. Does this patient have a hemorrhagic stroke?: clinical findings distinguishing hemorrhagic stroke from ischemic stroke. JAMA 2010 Jun 09;303(22):2280-2286. [doi: 10.1001/jama.2010.754] [Medline: 20530782]

XSL•FO
RenderX

27. Every NR, Hochman J, Becker R, Kopecky S, Cannon CP. Critical Pathways. Circulation 2000 Feb;101(4):461-465. [doi: 10.1161/01.cir.101.4.461]
28. Telestroke-RU. URL: https://telestroke.uniandes.academy [accessed 2020-06-27]
29. Wilson JTL, Hareendran A, Grant M, Baird T, Schulz UGR, Muir KW, et al. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. Stroke 2002 Sep;33(9):2243-2246 [FREE Full text] [Medline: 12215594]
30. Nogueira RG, Silva GS, Lima FO, Yeh Y, Fleming C, Branco D, et al. The FAST-ED App: A Smartphone Platform for the Field Triage of Patients With Stroke. Stroke 2017 May;48(5):1278-1284. [doi: 10.1161/STROKEAHA.116.016026] [Medline: 28411260]
31. Andrew BY, Stack CM, Yang JP, Dodds JA. mStroke: "Mobile Stroke"-Improving Acute Stroke Care with Smartphone Technology. J Stroke Cerebrovasc Dis 2017 Jul;26(7):1449-1456. [doi: 10.1016/j.jstrokecerebrovasdis.2017.03.016] [Medline: 28434773]
32. Dickson RL, Sumathipala D, Reeves J. Stop Stroke© Acute Care Coordination Medical Application: A Brief Report on Postimplementation Performance at a Primary Stroke Center. J Stroke Cerebrovasc Dis 2016 May;25(5):1275-1279 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2015.12.001] [Medline: 26971040]
33. Herweh C, Ringleb PA, Rauch G, Gerry S, Behrens L, Möhlenbruch M, et al. Performance of e-ASPECTS software in comparison to that of stroke physicians on assessing CT scans of acute ischemic stroke patients. Int J Stroke 2016 Jun;11(4):438-445. [doi: 10.1177/1747493016632244] [Medline: 26880058]
34. Katzan IL, Fan Y, Speck M, Morton J, Fromwiller L, Urchek J, et al. Electronic Stroke CarePath. Circ Cardiovasc Qual Outcomes 2015 Oct 29;8(6 suppl 3):S179-S189. [doi: 10.1161/circoutcomes.115.001808]

Abbreviations

AHS: advanced health care setting.
aPTT: activated partial thromboplastin time.
ASPECTS: Alberta Stroke Program Early Computed Tomography Scan.
CT: computed tomography.
CTA: computed tomography angiography.
CTP: computed tomography perfusion.
DWI: diffusion-weighted magnetic resonance imaging.
EHR: electronic health record.
FAST-ED: Field Assessment Stroke Triage for Emergency Destination.
ICH: intracerebral hemorrhage.
ICU: intensive care unit.
IHS: intermediate health care setting.
INR: international normalized ratio.
IV rtPA: intravenous recombinant tissue plasminogen activator.
MRA: magnetic resonance angiography.
MRI: magnetic resonance imaging.
NIHSS: National Institutes of Health Stroke Scale.
PHS: primary health care setting.
PT: prothrombin time.
PTT: partial thromboplastin time.
TIA: transient ischemic attack.
TICI: Thrombolysis in Cerebral Infarction.
WFNS: World Federation of Neurosurgical Societies.

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by L Moscote, L Rusu; comments to author 13.03.20; revised version received 16.03.20; accepted 20.03.20; published 27.07.20.

Please cite as:
Bayona H, Ropero B, Salazar AJ, Pérez JC, Granja MF, Martínez CF, Useche JN
Comprehensive Telestroke Network to Optimize Health Care Delivery for Cerebrovascular Diseases: Algorithm Development
J Med Internet Res 2020;22(7):e18058
URL: http://www.jmir.org/2020/7/e18058/
doi:10.2196/18058
PMID:32716302

XSL•FO
RenderX

©Hernán Bayona, Brenda Ropero, Antonio José Salazar, Juan Camilo Pérez, Manuel Felipe Granja, Carlos Fernando Martínez, Juan Nicolás Useche. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis

Yangyang Han[1], BS; Reidar K Lie[2], MD, PhD; Rui Guo[1], PhD

[1]School of Public Health, Capital Medical University, Beijing, China

[2]Department of Philosophy, University of Bergen, Bergen, Norway

Corresponding Author:
Rui Guo, PhD
School of Public Health
Capital Medical University
No 10 Xitoutiao, Youanmenwai
Fengtai District
Beijing, 100069
China
Phone: 86 01083911573
Email: guorui@ccmu.edu.cn

Abstract

Background: The internet hospital is an innovative organizational form and service mode under the tide of internet plus in the Chinese medical industry. It is the product of the interaction between consumer health needs and supply-side reform. However, there has still been no systematic summary of its establishment and definition, nor has there been an analysis of its service content.

Objective: The primary purpose of this study was to understand the definition, establishment, and development status of internet hospitals.

Methods: Data on internet hospitals were obtained via the Baidu search engine for results up until January 1, 2019. Based on the results of the search, we obtained more detailed information from the official websites and apps of 130 online hospitals and formed a database for descriptive analysis.

Results: By January 2019, the number of registered internet hospitals had expanded to approximately 130 in 25 provinces, accounting for 73.5% of all provinces or province-level municipalities in China. Internet hospitals, as a new telehealth model, are distinct but overlap with online health, telemedicine, and mobile medical. They offer four kinds of services—convenience services, online medical services, telemedicine, and related industries. In general, there is an underlying common treatment flowchart of care in ordinary and internet hospitals. There are three different sponsors—government-led integration, hospital-led, and enterprise-led internet hospitals—for which stakeholders have different supporting content and responsibilities.

Conclusions: Internet hospitals are booming in China, and it is the joint effort of the government and the market to alleviate the coexistence of shortages of medical resources and wasted medical supplies. The origin of internet hospitals in the eastern and western regions, the purpose of the establishment initiator, and the content of online and offline services are different. Only further standardized management and reasonable industry freedom can realize the original intention of the internet hospital of meeting various health needs.

(J Med Internet Res 2020;22(7):e17995) doi:10.2196/17995

KEYWORDS

Internet hospital; telehealth; telemedicine; ehealth; digital health; digital medicine; health services research; China

Introduction

During the past five years, China has been committed to promoting the use of the internet to provide medical services, although it is still in the exploratory stage [1]. Internet connections, in general, have brought about a significant change to the Chinese economy, and there is no doubt that it can improve the quality of service and provide convenience for people. As a supplement to offline health services, online health services are of great significance in alleviating the relative shortage of medical resources in China [2]. With the development of technology, online health service levels have

XSL•FO
RenderX

continually improved, which has led traditional medical resources to gradually become networked.

China has 904 million internet users, with an internet penetration rate of 64.5%, according to the China Internet Network Information Center in April 2020 [3]. The vast number of internet users constitute China's booming consumer market. Mobile payment is becoming more and more popular in daily life, and user habits of paying for online medical services based on knowledge are also becoming established [4]. From April 2018 to April 2019, China's internet medical users grew from 28 million to 45 million, a year-on-year increase of 59.9% [5]. Ping An Good Doctor, one of the representative internet hospitals, had 300 million registered users as of October 2019, equivalent to 1 in every 3 Chinese netizens having used it [6], and the number of visitors exceeded 1.11 billion from January to April 2020 during the COVID-19 pandemic [7]. The age range of internet hospital users is wide, and the number of middle-aged and elderly users is increasing [8].

Internet hospitals is a term with a somewhat vague meaning. In general, it is an internet medical platform combining online and offline access for medical institutions to provide a variety of telehealth services directly to patients. It extends medical resources from the hospital to the internet by using information technology, and it develops online medical services and health services [9]. It can be as simple as using the internet to make and manage appointments or renew prescriptions; however, it can also be a means for Chinese citizens to communicate face-to-face with doctors via internet pages or public therapeutic platforms in apps such as WeChat (mobile social media apps) [10]. Internet hospitals are online platforms but need to be based in existing institutions with consistent online and offline service supervision. They include both medical and nonclinical services but can only use internet technology to provide safe and appropriate medical services for the follow-up of common and chronic diseases online [11]. Doctors are allowed to prescribe treatment for related diseases online after accessing the patient's medical records [11].

Telehealth is a medical service model similar to internet hospitals in China and is used in settings throughout the world [12,13]. It commonly describes a wide range of diagnostic and management modalities, education, and other related activities within health care [14]. It is expanding beyond telemedicine to cover nonclinical events such as appointment scheduling, continuing medical education, and physician training [14].

In the United States, telehealth or telemedicine has been defined in all states, and some of these services by video can be reimbursed [15,16]. Many states consider online services inadequate for building doctor-patient relationships and prescribing treatments, so some allow online diagnosis and prescription of drugs while others restrict them [15]. A growing number of states are passing legislation to guide health care professional boards to adopt standards of practice for providers using telehealth [15]. One US telehealth platform, Amwell, an internet-based health care giant, has more than 150 million insured users, serving over 55 health plans, more than 240 medical groups, and more than 2,000 hospitals [17]. In terms of strategy and governance, studies have shown that only 2 countries in the European Union are in the management stage of policy supervision, and 16 countries are considered to be in the initial and temporary stage, meaning that most countries in the EU have no formal definition of telemedicine services or only small short-term and independent pilot developments [18]. Meanwhile, in Japan, access to telehealth clinical functions was restricted to health consultations, and only during the COVID-19 pandemic did the government allow patients to receive medical care and receive prescription via the internet rather than by going to a physical hospital [19,20].

However, in China, the current situation developed after an initial period of growth in internet hospital services when Chinese health authorities ordered a temporary halt in the development in order to develop a more appropriate set of regulations [21]. The first internet hospital in China was established in Guangdong Province in 2014 [22]. Then, China began to promote this new model of online health care to develop internet health services actively, to provide convenience services through the mobile internet, to encourage internet enterprises to cooperate with medical institutions, and to establish a medical network information platform. Thus, internet hospitals began to appear in some areas, gradually. In 2016, the First Affiliated Hospital of Zhejiang University established Zhejiang First internet hospital and was headed by the tertiary hospital [23]. However, by 2017, the government re-examined and approved the abolition of internet hospitals, cloud hospitals, and network hospitals [21]. By 2018, the state again encouraged medical institutions to use information technology such as the internet to expand the space and content of medical services and to build a comprehensive online and offline medical service model covering prognosis, diagnosis, and postdiagnosis [11]. Medical institutions were again allowed to develop internet hospitals and to provide safe and appropriate medical services using internet technology [11]. In 2018, the government formulated specific regulations on internet hospital management, including diagnosis and treatment, telemedicine, and other online aspects. This marked the beginning of the standardized development of internet hospitals [11].

Three documents were issued on July 17, 2018 by the National Health Commission of the People’s Republic of China and National Administration of Traditional Chinese Medicine: Internet Diagnosis and Treatment Management Measures (trial implementation), Internet Hospital Management Measures (trial implementation), and Telemedicine Service Management Standards (trial implementation) [11]. It was the first time that China put forward detailed regulations on internet hospitals, signifying that internet hospitals had entered the stage of standardized development [11]. Chinese internet hospitals have experienced ups and downs, after twists and turns of the embryonic stage, and are now entering the rapid development stage of standardization.

In order to capture the current state of development of internet hospitals in China, we carried out a study covering internet hospitals (up to the beginning of 2019) to analyze the construction mode and content of internet hospitals and provide corresponding suggestions for construction.

XSL•FO
RenderX

Methods

This study obtained data on internet hospitals through the Baidu search engine and Apple and Android app stores, manually. Baidu's dominant role in internet searching in China is comparable to the role of Google in Western countries, on both desktop and mobile devices, with approximately 70% to 80% market share. People can search for queries by entering keywords.

First, we searched for the official internet hospital name from April 6, 2019 to April 12, 2019. Before searching in Baidu, we sorted and analyzed policies of internet hospitals at the national level on the Website of the State Council. Internet Hospital Management Regulations stressed that internet hospitals should use the name of the hospital with the addition of the term *internet hospital* [11]. Some hospitals use the name of *cloud hospital* or *web hospital* as an auxiliary reference. So, we searched according to the internet hospital naming rules, following the Baidu Robots Protocol. Our inclusion criteria were internet hospitals established before January 1, 2019, and we included direct and indirect search results (direct: web items; indirect: government announcements and page content under web item). This resulted in a list of 130 internet hospital names in our Excel (Microsoft Inc) database.

Second, we queried the relevant information about the internet hospital name in Baidu, inquired from the internet hospital official website, and downloaded the internet hospital app from the app store to obtain the following variables to build a database used in the statistical analysis: establishment time (precise to the month), establishment location, service content provided by technical cooperation parties, and media (pages, public medical platforms, or apps) in contact with people. The study focused on the overall development of the internet hospital industry, so we did not collect the number of patients and doctors in each hospital.

Finally, the data obtained from the 130 internet hospitals were entered into the Excel database and analyzed.

Results

The Structure of Internet Hospitals

Our study of the available information about internet hospitals revealed that the term is used for a number of different types of activities. Telemedicine was launched first, and then mobile medicine and online medicine became popular. Generally speaking, online medicine, mobile medicine, and telemedicine are all covered by the concept of internet hospital. In Table 1, we have summarized the relationships between the various uses of the term internet hospital.

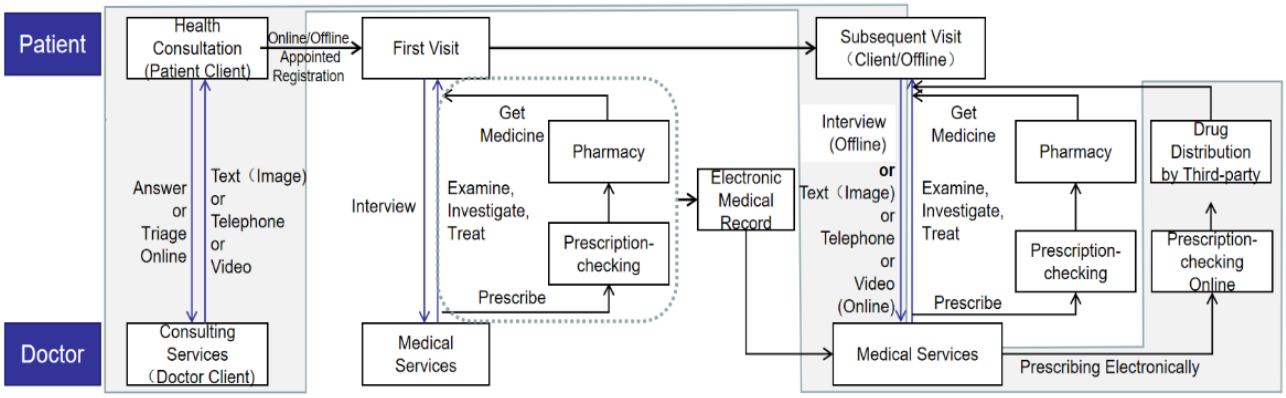
Based on the data we obtained, we constructed a flowchart between offline hospitals and online hospitals by summarizing the treatment process of existing internet hospitals based on doctor-patient interaction (Figure 1). Patients can consult online before going to the offline hospital to collect relevant professional health information and understand their current situation. If they need to go to a hospital, there can be a more extensive and reliable consultation that can assist the hospital in triaging. Next, patients can choose to register online or go to a hospital for a preliminary examination. After the first visit, patients can choose between online or offline ways to check the disease prognosis, treatment effects, and further treatment. Moreover, sometimes subsequent visits that require examination and testing need to be done offline. Therefore, the two areas are divided into white and gray, indicating that they may be online or offline (Figure 1). Then, pharmacists check the electronic prescriptions, and drugs can be delivered by a third-party logistics enterprise cooperating with the hospital or can be picked up at a nearby pharmacy. Patients can also accept door-to-door diagnosis and treatment out of the hospital.

Table 1. The relationship and difference between others with internet hospitals.

Noun	Definition	Relationship and difference with internet hospitals
Online health	The use of internet technology to provide patients and the public with disease diagnosis, treatment programs, prescriptions, and other services [21].	It is one of the services provided by internet hospitals.
Mobile medicine	Emerging mobile communications and network technologies for health care systems [24]. It is via mobile and wireless devices, and the sharing of that information between patients and providers such as personal digital assistants, mobile phones, and satellite communications.	Internet hospitals have a variety of digital media, such as portals, public medical platforms, and apps.
Telemedicine	Use of telecommunications technologies to provide medical information and services [25].	Internet hospitals are a platform to provide technical support for remote diagnosis and treatment.

XSL•FO
RenderX

Figure 1. Hospital treatment flowchart.



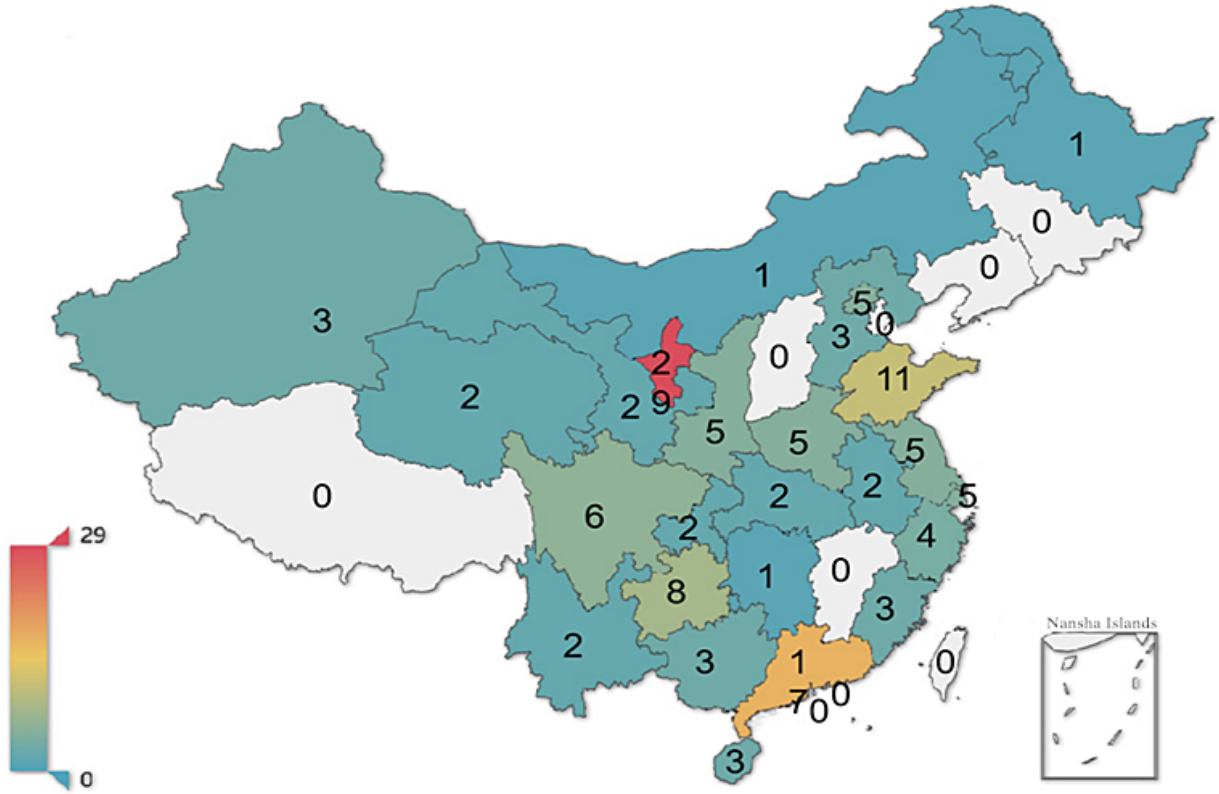
Status Analysis of Internet Hospitals

As of January 2019, 25 of the 34 provinces or municipalities (73.5%) had registered a total of 130 internet hospitals in China, ranging from 0 to 29 per administrative region (Figure 2). There has been a growth trend from the establishment of internet hospitals up to the beginning of 2019. In the years from 2014 to 2018, 1 (0.8%), 5 (3.8%), 42 (32.3%), 65 (50.0%), and 17 (13.1%) hospitals were launched, respectively. The top 11 provinces and cities accounted for 100/130 (76.9%) internet hospitals. We can conclude that the number of internet hospitals grew very slowly between the end of 2017, when the establishment of internet hospitals was discouraged, and the end of 2018, when policies for comprehensive regulation were introduced.

Based on the general classification of internet hospital services and based on our data and the national policy documents, services provided by internet hospitals include four classifications, each referring to the service content for patients (Table 2). Convenience services help patients to operate online to simplify offline procedures and learn health knowledge independently. Online health means that patients can communicate with doctors to consult for health problems or obtain the diagnosis and prescriptions through the internet. Telemedicine is medical cooperation between institutions initiated by the hospital to other hospitals when patients seek medical treatment offline. Moreover, related industries refer to the content that needs to be provided by a third party, such as health records and drug-related enterprises.

We randomly selected the screenshot of the homepage of 1 Chinese internet hospitals app (Patient Client) (Figure 3); the content is based on 3 of the 4 classifications above except telemedicine between institutions.

Figure 2. Heat map of number of internet hospitals in Chinese provinces.

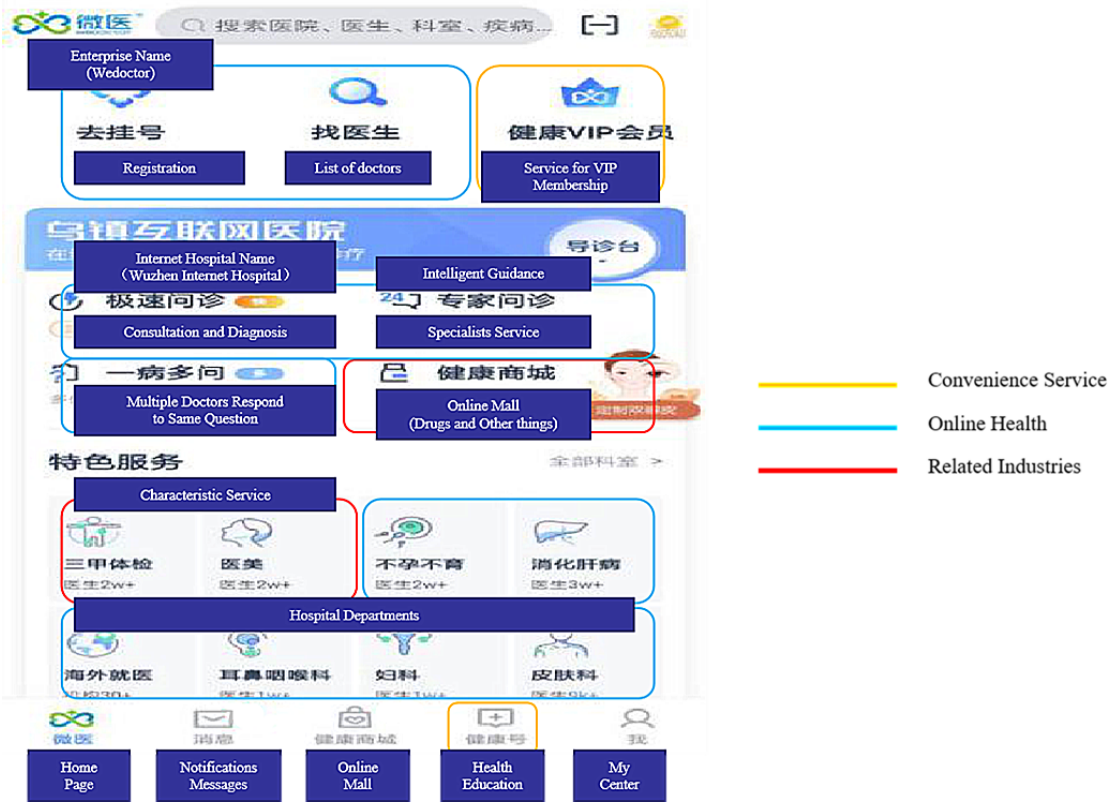


XSL•FO
RenderX

Table 2. Internet hospital service classification.

Classification	Content
Convenience services	Intelligent guidance, registration, guidance, mobile payment, inspection report query, cost query [26], medical feedback, health education
Online health	Health consultation and follow-up for common and chronic diseases
Telemedicine (institution)	Providing on-line medical services initiated by institutions
Related industries	Drug order, drug distribution, health records

Figure 3. Example of internet hospital app interface.



Every internet hospital contains specific content in the introduction, such as a booking service (online booking of a medical expert appointment within the medical facilities), but also health education (the hospital provides a wide range of medical knowledge about drugs, chronic disease knowledge lectures, information about national health policy, and so forth). Through internet portals, the patient can query to book appointments and can check reports. Doctors, via the outpatient internet service, can provide electronic prescription service to patients, including online prescriptions and online payment records. The network outpatient service for patients offers electronic prescription services, including online prescription, online payment dispensing, and others.

Based on our data, internet hospitals can be divided into three categories, and each has relatively successful typical cases. In the first category, the government is the initiator whose primary purpose is to achieve a unified standard of regional population health management, such as the Sichuan Province internet hospital, or the Zhejiang Province internet hospital. The second category is initiated by hospitals whose primary purpose is to expand the scope and intensity of hospital services so that people can continue to see their doctors. For example, the First Affiliated Hospital of Zhejiang University, a well-known tertiary hospital in China, opened their internet hospital to provide a platform for their doctors to provide services online. The third category is initiated by enterprises whose purpose is to connect patients with practicing doctors who, through these efforts, can increase the number of patients, and thereby, gain benefits. Also, they can improve the accessibility of health services for patients, and the increase in patients can improve doctors' income and skills. This category includes some enterprises such as Haodf (Yinchuan Intelligent internet hospital), Ping An Good Doctor (Ping An Good Doctor internet hospital), Wedoctor (Wuzhen internet hospital). These were the first to utilize internet plus in the medical industry to carry out the sales of light drug consultations, health management, and other services. Even though these three categories can be separate, the effective delivery of medical services requires the cooperation of doctors, patients, hospitals, governments, and enterprises. Based on our database of internet hospitals, 38/130 (29.2%) were initiated by the government and hospital, and 92/130 (70.8%) were initiated by enterprises; we can see that each of them undertakes different roles from the perspective of stakeholders. The sources of doctors and patients are also different (Table 3).

XSL•FO
RenderX

Table 3. Relevant parties of internet hospital model.

Initiator	Integration of government services for guidance	Longitudinal hospital-led service	Horizontal enterprise-led service
Government	Establishment of internet medical platform at administrative level	The government is responsible for examination, approval and supervision	The government is responsible for examination, approval and supervision
Hospital	Establishment of online hospital districts in local medical institutions	Transfer offline services to online	Conditions and requirements on which services must provide and carry out entity-based
Enterprise	Contractor provides technical and medical resources support	Contractor provides technical support	Invest in construction for profit
Doctor	The contractor platform registered physicians and physicians from regional medical institutions provide services.	Doctors belonging to the hospital	Multipoint practicing doctor
Patient	Patients in administrative areas	Hospital service area mainly for patients, supplemented by network patients	Internet patients

Discussion

Principal Findings

The most striking finding in our study is the level of integration of ordinary hospital services with the internet hospital. While enterprises have initiated most internet hospitals (92/130, 70.8%), the dominant model is an integration of the internet platform with existing hospital and government services. The construction of the internet hospitals in China is a two-way force, to a certain extent, as a joint effort by the government and the market to alleviate the coexistence of a shortage of medical resources and wasted medical supplies.

On the one hand, health resources are abundant in the east, while they are relatively few in the midwest regions [27]. In rural areas with scarce medical resources, the government and hospitals hope to optimize resources through telehealth, so that patients can contact other remote doctors to solve medical problems, and promote high-quality medical resources in the east to meet the needs of patients in the midwest; therefore, there are many internet hospitals in the central and western regions of China. The internet enables people to overcome geographic barriers to access health care services.

On the other hand, the e-commerce giants (such as Taobao or Jingdong) have been involved with the pharmaceutical industry, providing online drug shopping and health consulting for consumers, but they cannot directly sell prescription drugs [28]. The enterprises follow China's internet medical regulation policies to create offline hospitals or cooperate with the existing medical institutions in building internet hospitals to ensure that they have implemented the whole process by the closed loop (doctor-patient-hospital medical services). With the rise of the industry, many regions with developed internet technologies took the initiative to enter the internet medical field and set up internet hospitals in developed cities to serve more patients.

Internet Hospital Service Depends on the Integration of Online and Offline Services

The development of online medical services and health services in the new integrated model has several clear benefits. Patients can use the portal and medical information systems to improve communication with health care providers. They can take advantage of internet technology, reduce registration time, and reduce treatment times [29]. The richness of internet medical treatment can meet the diversified health needs of patients, to a certain extent, and also help to realize the continuity of medical services for patients [30,31].

Despite these advantages, it is also clear that internet hospitals cannot replace many of the core functions of physical hospitals. The internet hospital Management Measures stipulates two situations for the use of internet medical treatment in practice. In the first case, when a patient visits a physical medical institution and the institution initiates contact with other doctors to consult through an internet hospital, the physician can provide diagnostic advice and prescriptions. In the other case, patients do not consult with physical medical institutions. Doctors can then only offer follow-up to patients for common and chronic diseases through the internet hospital. The internet hospital includes many internet-based medical services, but it does not have many necessary medical resources and physical infrastructure online such as sickbeds, diagnostic equipment, and operating conditions for medical services. Therefore, the reality is that most internet hospital online services still only offer a supplement to the traditional nonclinical parts such as booking, registration, guidance, payment, and other convenience services [32].

Moreover, internet hospitals cannot provide the first diagnosis and complex disease follow-up after a diagnosis, so they cannot be involved in some of the core business of medical care, limiting the utilization rate [11,33]. For example, dermatology mainly relies on visual data to make a diagnosis. Usually, doctors with rich clinical experience can make a clear judgment after seeing the skin photos, while some departments rely more on examination and monitoring for diagnosis [34]. Meanwhile, in health consultation, doctors can conduct triage based on patient descriptions or video images, and doctors need to master additional characteristics of the patient's integrated circumstance and ensure that they provide valid information for the medical records on the internet to provide online follow-up. Some medical services (first visit and severe illness) must be provided by offline hospitals, while others (nonclinical, health consultation, and follow-up diagnosis) can conveniently be realized online. Internet hospitals have stimulated the advantages of online and offline services to meet the medical needs of

XSL•FO
RenderX

patients better and release high-quality medical resources (health workforce, equipment, etc) for offline hospitals [35].

Different Initiators of Internet Hospitals May Have Different Development Directions

The government, hospitals, and enterprises are the initiators of the three ways of building internet hospitals, and they are also relevant stakeholders. The government is the policymaker, the main body of industry supervision, and plays an irreplaceable role [36]. The primary purpose of the government as the initiator is to integrate medical resource information and population health information in the region to form a database of population information, electronic health records, electronic medical records, and other data in the entire region [37,38]. As the foundation for the construction of internet hospitals, hospitals are providers of offline medical services. The primary purpose of an internet hospital with a physical hospital as the initiator is to extend the hospital's service process to the internet, extend the medical service capabilities of the physical hospital to the patient's home, and provide remote follow-up services for patients [39]. The enterprise is the technical support built by the internet hospital and maintains the regular operation of the platform. When enterprises are the main initiators of the internet hospital, their focus is on increasing operating income and achieving profitability. From the perspective of doctors and patients in internet hospitals, the source of doctors and patients for internet hospitals initiated by enterprises is significantly different from the source of doctors and patients when initiated by the government and hospitals. The doctors provided by internet hospitals built by the government and hospitals are qualified doctors within the administrative area, and the scope of their service is still local patients with a small number of cross-province patients. The service scope of the internet hospital built by enterprises is registered doctors and patients nationwide. Nevertheless, at least for the time being, enterprise-initiated internet hospitals still have to rely on existing physical infrastructure. If that should change and such hospitals can operate more independently, it would present a significant challenge for government regulation and supervision.

Limitations

This study has several limitations. First, although this study attempts to describe the current situation of all internet hospitals in China, there may be missing items because data collection relied upon web-based search engines. Second, this research sample only included internet hospitals up to the beginning of 2019, which is the initial stage of the internet hospital industry (from its initiation and growth from very few to a substantial number), but it did not include the continuous replication of the internet hospital model in 2019. We will include these data in future studies to further analyze the relationship between the number of internet hospitals and the amount of services they provide. Third, this study did not investigate the actual online-offline conversion rate of internet hospitals to evaluate the service capacity of the hospitals. It is necessary to conduct research from the perspective of doctors, patients, and managers to further study stakeholders' views on internet hospitals and to analyze the influencing factors of internet hospital usage in the current situation.

Conclusions

Internet hospitals redefined the concept of medical service provision through telecommunications technology and began to revolutionize the health sector [40]. It is praiseworthy that China innovates the internet hospital model in the field of internet health care so that more people can receive medical resources to which they did not previously have access. The most significant value of the internet lies in the realization of the transfer of medical centers, from traditional hospitals as the main body to a more patient-oriented model where people can receive medical services and health care without leaving home. Currently, although the government has put forward management measures, they are still unable to achieve an internet hospital that provides the core functions of medical service. The government still needs to develop more suitable content to form a systematic internet medical model and to explore the regulatory scope to provide more development freedom for the internet hospital industry. It is essential to realize the central idea of the internet hospital—hospitals have more space to treat diseases, and patients save time [41].

Acknowledgments

This study was funded by the National Natural Science Foundation of China (No. 71704118).

Authors' Contributions

YH and RG designed the main concepts of this work. YH performed data collection and wrote this paper. RG and RKL edited and promoted the manuscript.

Conflicts of Interest

None declared.

References

1. Xie XX, Zhou WM, Lin LY, Fan S, Lin F, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
2. Tu J, Wang CX, Wu SL. The internet hospital: an emerging innovation in China. Lancet Global Health 2015 Aug;3(8):e445-e446. [doi: 10.1016/S2214-109X(15)00042-X]

XSL•FO
RenderX

3. 45th statistical report on internet de-velopment in China [in Chinese]. China Internet Network Information Center. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202004/t20200428_70974.htm [accessed 2020-06-01]
4. Comprehensive analysis of China's internet healthcare in 2018 [in Chinese]. Yunzhang Finance. URL: http://www.123.com.cn/kline/67027.html [accessed 2020-06-01]
5. Tu Y, Zhou C. Market analysis of China's internet medical industry in 2020 [in Chinese]. Sohu. URL: https://www.sohu.com/a/377757603_114835 [accessed 2020-06-01]
6. The registered users of Ping An internet hospital exceeded 300 million [in Chinese]. People.cn. URL: http://sh.people.com.cn/n2/2019/0923/c134768-33379897.html [accessed 2020-06-01]
7. Performance of Ping An Good Doctor: 1 billion visits during the epidemic [in Chinese]. Sina Finance. URL: https://baijiahao.baidu.com/s?id=1658298035741040628&wfr=spider&for=pc [accessed 2020-06-07]
8. Li JM, Wang XL, Chen H, Li P, Chen JW, et al. Exploration of building large public hospitals into internet hospitals [in Chinese]. Journal of Medical Informatic 2018 Dec;39(2):22-25.
9. Chang ZD, Chen M. Analysis of medical service mode and the trend of internet hospital [in Chinese]. J Health Inform Manag 2016 Dec;13(6):e557-e560.
10. Xu JH, Kang Q, Song Z, Clarke CP. Applications of mobile social media: WeChat among academic libraries in China. The Journal of Academic Librarianship 2015 Jan;41(1):21-30. [doi: 10.1016/j.acalib.2014.10.012]
11. Measures for the administration of internet diagnosis and treatment (for trial implementation). National Health Commission of the People's Republic of China. 2018 Jul 17. URL: http://www.gov.cn/gongbao/content/2019/content_5358684.htm [accessed 2020-06-01]
12. Scott R, Mars M. Telehealth in the developing world: current status and future prospects. Smart Homecare Technology & Telehealth 2015 Feb;3:25-37. [doi: 10.2147/shtt.s75184]
13. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
14. What is telehealth? The National Telehealth Policy Resource Center. URL: https://www.cchpca.org/about/about-telehealth [accessed 2020-06-01]
15. State telehealth laws & medicaid program policies. Center for Connected Health Policy. 2020. URL: https://www.cchpca.org/sites/default/files/2020-05/CCHP_%2050_STATE_REPORT_SPRING_2020_FINAL.pdf [accessed 2020-06-01]
16. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
17. Amwell. URL: https://business.amwell.com/ [accessed 2020-06-01]
18. Carrasqueiro S, Ramalho A, Esteves A, Pereira C, Martins D, et al. Report on EU state of play on telemedicine services and uptake recommendations. Joint Action to Support the eHealth Network. 2017. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20171128_co09_en.pdf [accessed 2020-06-01]
19. Shimoda T, Yoshimura S, Yoshida Y, Okazaki M, Gotou T, et al. Development and current status of an advanced telehealth consultation system in Japan. J Telemed Telecare 2015 Feb;21(3):176-178. [doi: 10.1177/1357633x15571998]
20. Japan will allow online treatment of first-visit patients. Xinhua News Agency. URL: https://www.sohu.com/a/387105493_267106 [accessed 2020-06-01]
21. On soliciting measures for the administration of internet diagnosis and treatment (for trial implementation). National Health and Family Planning Commission of China. 2017 May 11. URL: https://www.sohu.com/a/139754985_648425 [accessed 2020-06-01]
22. Guangdong Network Hospital officially launched [in Chinese]. Youdeyi. URL: http://www.youdeyi.com/newinfo.html?M_type=1&id=1 [accessed 2020-06-01]
23. The country's first public tertiary-hospital-led internet hospital officially launched [in Chinese]. CN-Healthcare. URL: https://www.cn-healthcare.com/article/20160216/content-481397.html?from=timeline&isappinstalled=0 [accessed 2020-06-01]
24. Istepanian RSH, Laxminarayn S, Pattichis CS. M-health: Emerging Mobile Health Systems. Boston: Springer Science + Business Media; 2006.
25. Perednia DA, Allen A. Telemedicine technology and clinical applications. JAMA 1995 Feb 08;273(6):483-488. [doi: 10.1001/jama.1995.03520300057037] [Medline: 7837367]
26. Opinions of the General Office of the State Council on Promoting the Development Internet + Health and Medical Care. General Office of the State Council. 2018 Apr 28. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-06-01]
27. He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, et al. Using an internet-based hospital to address maldistribution of health care resources in rural areas of Guangdong province, China: retrospective and descriptive study. JMIR Med Inform 2018 Dec 21;6(4):e51 [FREE Full text] [doi: 10.2196/medinform.9495] [Medline: 30578195]
28. Ruan X. Pharmaceutical e-commerce: break the old pattern and find a new way out [in Chinese]. New Economy Weekly 2016 Jan(1):84-87.

XSL•FO
RenderX

29. Min C. Internet + medical health: creating a new mode of intelligent medical service [in Chinese]. Chinese Cadres Tribune 2018(10):30-33. [doi: 10.14117/j.cnki.cn11/d.2018.10.007]
30. Deng SY. Research on internet medical service model from the perspective of value co-creation [in Chinese]. Science and Technology Innovation Herald 2018 Jul;15(07):225-226. [doi: 10.16660/j.cnki.1674-098X.2018.07.225]
31. Ji H, Xie YF, Li L. Discussion on the application status and prospect of the internet medical service mode [in Chinese]. Financial and Economic Circles (Academic Edition) 2018(16):140. [doi: 10.16266/j.cnki.cn11-4098/f.2018.16.103]
32. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: multicenter user profiling study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
33. Ye Q, Deng Z, Chen Y, Liao J, Li G, Lu Y. How resource scarcity and accessibility affect patients? usage of mobile health in China: resource competition perspective. JMIR Mhealth Uhealth 2019 Aug 09;7(8):e13491. [doi: 10.2196/13491] [Medline: 31400104]
34. Krupinski EA, Patterson T, Norman CD, Roth Y, ElNasser Z, et al. Successful models for telehealth. Otolaryngol Clin North Am 2011 Dec;44(6):1275-1288 [FREE Full text] [doi: 10.1016/j.otc.2011.08.004] [Medline: 22032481]
35. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: a new health care mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
36. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014 Feb;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
37. Tian J. Construction of network hospital of Guangdong province in the background of ?Internet + Medical? [in Chinese]. China Digital Medicine 2016 Apr;11(01):28-30.
38. Sichuan to build a population health information platform. The People's Government of Sichuan Province. URL: http://www.sc.gov.cn/10462/12771/2016/10/28/10401778.shtml [accessed 2020-06-01]
39. Yang Y. Analysis of medical service mode and the trend of internet hospital [in Chinese]. Medical and Health 2017 Sep(01):18.
40. Xu LW. Internet hospital: challenges and opportunities in China. In: Springer International Publishing.: Springer International Publishing; 2016 Nov Presented at: International Conference on Health Information Science; 2016; Cham p. 85-90. [doi: 10.1007/978-3-319-48335-1_9]
41. Kou JQ, Yan ZZ, Zhang WX, Jiang M. Design and implementation of the internet hospital based on the substantial hospital [in Chinese]. China Medical Devices 2018 Nov;33(11):135-139. [doi: 10.3969/j.issn.1674-1633.2018.11.039]

Edited by G Eysenbach; submitted 27.01.20; peer-reviewed by H Hu, L Sbaffi; comments to author 20.04.20; revised version received 02.06.20; accepted 14.06.20; published 29.07.20.

Please cite as:
Han Y, Lie RK, Guo R
The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis
J Med Internet Res 2020;22(7):e17995
URL: http://www.jmir.org/2020/7/e17995/
doi:10.2196/17995
PMID:32723721

©Yangyang Han, Reidar K Lie, Rui Guo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Technology-Enabled Self-Management of Chronic Obstructive Pulmonary Disease With or Without Asynchronous Remote Monitoring: Randomized Controlled Trial

Vess Stamenova[1], PhD; Kyle Liang[1], MSc; Rebecca Yang[1], MPH; Katrina Engel[2], RRT; Florence van Lieshout[1], BSc; Elizabeth Lalingo[3], RRT; Angelica Cheung[2], RRT; Adam Erwood[4], MMI; Maria Radina[5], MD; Allen Greenwald[5], MD; Payal Agarwal[1,6], MD; Aman Sidhu[7,8], MSc, MD; R Sacha Bhatia[1,8,9], MBA, MD; James Shaw[1,9], PhD, PT; Roshan Shafai[10], MD; Onil Bhattacharyya[1,6,9], PhD, MD

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Respiratory Therapy Department, Markham Stouffville Hospital, Markham, ON, Canada

[3]Medicine, Care Transitions, Access & Flow, Respiratory Therapy, Markham Stouffville Hospital, Markham, ON, Canada

[4]Support Services & Transformation, Markham Stouffville Hospital, Markham, ON, Canada

[5]Center for Respiratory Health, Markham Stouffville Hospital, Markham, ON, Canada

[6]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[7]University Health Network, Toronto, ON, Canada

[8]Department of Medicine, University of Toronto, Toronto, ON, Canada

[9]Women's College Research Institute, Toronto, ON, Canada

[10]Hospital to Home and Community Medicine Clinic, Markham Stouffville Hospital, Markham, ON, Canada

Corresponding Author:
Vess Stamenova, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 4163236000 ext 5112
Email: vess.stamenova@wchospital.ca

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality and leads to frequent hospital admissions and emergency department (ED) visits. COPD exacerbations are an important patient outcome, and reducing their frequency would result in significant cost savings. Remote monitoring and self-monitoring could both help patients manage their symptoms and reduce the frequency of exacerbations, but they have different resource implications and have not been directly compared.

Objective: This study aims to compare the effectiveness of implementing a technology-enabled self-monitoring program versus a technology-enabled remote monitoring program in patients with COPD compared with a standard care group.

Methods: We conducted a 3-arm randomized controlled trial evaluating the effectiveness of a remote monitoring and a self-monitoring program relative to standard care. Patients with COPD were recruited from outpatient clinics and a pulmonary rehabilitation program. Patients in both interventions used a Bluetooth-enabled device kit to monitor oxygen saturation, blood pressure, temperature, weight, and symptoms, but only patients in the remote monitoring group were monitored by a respiratory therapist. All patients were assessed at baseline and at 3 and 6 months after program initiation. Outcomes included self-management skills, as measured by the Partners in Health (PIH) Scale; patient symptoms measured with the St George's Respiratory Questionnaire (SGRQ); and the Bristol COPD Knowledge Questionnaire (BCKQ). Patients were also asked to self-report on health system use, and data on health use were collected from the hospital.

Results: A total of 122 patients participated in the study: 40 in the standard care, 41 in the self-monitoring, and 41 in the remote monitoring groups. Although all 3 groups improved in PIH scores, BCKQ scores, and SGRQ impact scores, there were no

XSL•FO
RenderX

significant differences among any of the groups. No effects were observed on the SGRQ activity or symptom scores or on hospitalizations, ED visits, or clinic visits.

Conclusions: Despite regular use of the technology, patients with COPD assigned to remote monitoring or self-monitoring did not have any improvement in patient outcomes such as self-management skills, knowledge, or symptoms, or in health care use compared with each other or with a standard care group. This may be owing to low health care use at baseline, the lack of structured educational components in the intervention groups, and the lack of integration of the action plan with the technology.

Trial Registration: ClinicalTrials.gov NCT03741855; https://clinicaltrials.gov/ct2/show/ NCT03741855

(J Med Internet Res 2020;22(7):e18598) doi:10.2196/18598

KEYWORDS

COPD; eHealth; telemedicine; remote consultation; self-care

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is the third leading cause of mortality worldwide [1], with 65 million people having moderate to severe COPD worldwide [2]. In Ontario, Canada, COPD accounts for 24% of hospital admissions and 24% of emergency department (ED) visits and is responsible for the highest percentage (18.8%) of 30-day ED readmissions [3]. Reducing the frequency of COPD exacerbations is an important patient outcome and could result in significant cost savings.

One approach to reducing exacerbations is to provide regular remote monitoring of patients from their homes. Remote monitoring requires patients to take measurements of their vital signs (oxygen, blood pressure, and symptoms) and to record them manually on paper [4] or to transmit them with devices using a phone or an internet line [5-7]. Recent developments such as Bluetooth technology, cloud-based storage, and Wi-Fi–enabled tablets have allowed data from remote devices to be uploaded automatically to a database accessible to patients, caregivers, and health care providers, either periodically or on an *as-needed* basis [8-10]. Remote monitoring programs are always monitored by a health care provider, even though they are sometimes referred to as *self-management* programs, as the recordings are taken by the patients [10]. Some remote monitoring programs also often have an educational component [11,12] such as coaching sessions to support self-management. COPD self-management behaviors include self-recognition and self-treatment of exacerbations (eg, taking medications); coping with breathlessness; and lifestyle changes such as quitting smoking, eating healthy, and exercising [13]. Self-management COPD interventions have generally been shown to be effective in improving quality of life measures [13,14], but a recent meta-analysis failed to show significant improvements in quality of life [15].

There is a large body of literature on COPD reporting on the effects of remote monitoring on patient outcomes and health care utilization, and several recent reviews have summarized these findings [16-18]. For example, Kruse et al [16] reported that the number of articles stating that patient outcomes improved overall with telemonitoring was approximately equal to that showing no improvement. Another review [17] reported that remote monitoring decreased ED admissions and hospitalizations but failed to impact other patient outcomes (mortality, outpatient visits, and length of stay). Hong and Lee [17] suggested that *integrated remote monitoring* programs (those that have educational components) may be more effective, especially when they target patients with more advanced diseases.

Educational components come at an additional cost to these programs [19] and even the simple act of monitoring patients remotely and connecting with them only when alerts are received requires dedicated staff. Few studies have looked at the effectiveness of *self-monitoring* programs. Self-monitoring programs ask patients to take their readings and receive automated feedback based on these readings without being actively monitored by a health provider [4,20]. Results from the few studies available in the literature have shown some promise in improving patient outcomes but they were feasibility trials that required larger samples and control group designs. A self-monitoring program, if noninferior to a remote monitoring program, would provide the opportunity for significant cost savings without compromising on patient outcomes. To our knowledge, no studies to date have directly compared a remote monitoring program with a self-monitoring program for patients with COPD.

Objectives

The objective of our study was to compare the effectiveness of implementing a technology-enabled self-monitoring program versus a technology-enabled self- and remote-monitoring program (or simply remote monitoring) in a population of patients with COPD, compared with a standard care group. We hypothesized that both intervention programs would lead to improvements in self-management skills and respiratory symptoms relative to the standard care program. In addition, the technology-enabled remote monitoring programs may be more effective at increasing COPD knowledge than self-monitoring alone.

Methods

Study Setting

The study was conducted at a 309-bed community-based hospital in Ontario. Recruitment took place in a hospital-based outpatient COPD clinic, from the private practice of respirologists affiliated with the hospital, and from an outpatient COPD rehabilitation program.

XSL•FO
RenderX

Trial Design

We conducted an open-label randomized controlled trial (RCT) comparing 2 technology-enabled interventions, a self-monitoring group and a remote monitoring group, relative to standard care. Patients were randomized in a 1:1:1 ratio to 1 of the 3 groups. The study recruitment started in April 2018 and was completed in September 2019. A full description of the protocol has been published [21]. We report here only on the quantitative portion of the evaluation. The qualitative results are published separately [22].

Participants

Eligibility Criteria

Patients were included if they were aged 18 years or older and had an established clinical diagnosis of COPD by a respirologist, according to clinical guidelines [23]. Exclusion criteria included a diagnosis of other significant lung diseases (eg, interstitial lung disease), patients without Wi-Fi internet access in their homes, inability to read English (required for filling out the questionnaires), participation in other remote monitoring programs, or inability to use the technology because of physical or cognitive impairment.

Recruitment Process

The main site of recruitment was the hospital-based outpatient clinic, where all eligible patients seen within the past year were contacted for participation. Patients could also be referred to the study from outside the clinic, through the private practice of hospital-affiliated respirologists or through an outpatient COPD rehabilitation program. Patients were contacted by phone, directly approached at an appointment, or approached at the hospital's exercise rehabilitation program by a clinical staff member (respirologist or respiratory therapist [RT]). Those who were interested were referred to the clinical project specialist and scheduled for a baseline evaluation, at which time informed consent was obtained, group allocation was revealed, and the kit was provided (if in the self-monitoring or remote monitoring group).

Allocation

We used a web-based random number generator [24] to allocate patients to groups, as described in the trial protocol [21]. Sequential patient group allocation was placed in a sealed envelope and revealed to the patient by the clinical project specialist after consent was obtained.

Intervention

Technology

The *Cloud DX Connected Health Kit* (Cloud DX Inc) [25] (Multimedia Appendix 1) was used in the 2 intervention groups. It was selected as it was made by a local Ontario company (a requirement from the granting agency), was fully developed, was on the market at the time of the study, and was capable of monitoring oxygen saturation. The kit comprised the following Bluetooth devices: a custom tablet computer, a Pulsewave wrist cuff monitor (which measures blood pressure), an oximeter, a weighing scale, and a thermometer. The devices were approved by the US Food and Drug Administration and Health Canada. A digital version of the COPD Assessment Test (CAT) [26] and the modified Medical Research Council (MRC) Scale [27] were also embedded in the technology. The data from all devices were transmitted to a database, and patients and health care providers interacted with it through a web-based portal. Regular bug fixes were occurring throughout the trial, and no major revisions of the content of the platform were done. Throughout the trial, there were 3 releases and 1 service pack installed in the platform and 2 releases, 5 hotfixes, and 1 service pack released for the companion app (see Multimedia Appendix 2 for details).

Intervention Procedures

The intervention lasted for 6 months. Patients in the intervention groups were asked to record their vitals (oximetry and blood pressure were required, whereas temperature and weight were optional) and symptoms (CAT and MRC) with the Cloud DX platform every day. They were also provided with a written version of a personalized COPD action plan that instructed patients on what to do if their readings fell outside predetermined thresholds (Multimedia Appendix 1). Individual patient thresholds were determined by the clinical project specialist (who was an RT), in consultation with the patient's respirologist. Patients in the self-monitoring and remote monitoring groups were additionally contacted by the clinical project specialist 2 weeks after receiving their kit to reassess the appropriateness of the thresholds. In addition, all patients had the option to email or call the clinic with any nonemergency questions they may have. All patients were advised to go to the ED if necessary, at any point in the study. Patients were also informed that data were not monitored 24 hours, 7 days a week and to respond to their clinical needs as they would normally do outside of the study.

When a patient's readings fell outside the predetermined thresholds, a notification was sent to both the clinical project specialist and the patient through email. The clinical project specialist reviewed the readings and responded when clinically indicated only for the remote monitoring group. Follow-up calls were made only when the readings exceeded thresholds twice or more within 2 days and were made only on weekdays. An attempt to complete the follow-up call was performed within 24 hours of receiving the notification. If the patient was unavailable, a message was left to return the call. In addition, the RT called the patients in the remote monitoring group once a week, irrespective of the values of the vitals. The purpose of the call was to check the patients, prompt action plan usage as needed, and provide education to the patients about their COPD as needed. The clinical project specialist received the readings for the self-monitoring group, but they were not actively monitored, and no follow-up calls were made in this group. Patients in the self-monitoring group were informed that their data were not actively monitored by the clinic. Patients in both intervention groups had secondary threshold levels (extreme measures) preset by the site investigator. Cloud DX staff monitored these levels and contacted the patients when necessary. For details, please refer to the protocol [21].

Patients in the standard care group were not provided with a technology or an action plan. This group received otherwise

XSL•FO
RenderX

standard care from the respiratory clinic, including routine in-person follow-up appointments and access to a certified respiratory educator. Patients in the standard care group were told that they would receive the equipment at the end of the trial to incentivize them to stay in the trial and to ensure that all participants had equal access.

Outcomes

All patients completed 3 assessments, at baseline, at 3 months, and at 6 months, on a series of questionnaires. Visit 1 (baseline) was in person, whereas visits 2 (3 months) and 3 (6 months) could be done in person or remotely (online through REDCap [28,29] or over the phone).

Primary Outcome

The primary outcome for the trial was self-management as assessed using the Partners in Health (PIH) Scale [30], a validated scale measuring the current status of self-management, with items on the knowledge of the condition and skills to monitor and respond to symptoms. This scale was selected to measure the primary outcome as we believed that both interventions could lead to self-management improvement.

Secondary Outcomes

The secondary outcomes included measures of COPD severity and COPD knowledge and were measured with the St George's Respiratory Questionnaire (SGRQ) [31] and the Bristol COPD Knowledge Questionnaire (BCKQ) [32]. The SGRQ contains subsections on respiratory symptoms, activities that are limited because of breathlessness, and impacts on daily life. The BCKQ [32] is a measurement of the level of knowledge of the disease in patients with COPD. Patients were also asked to self-report at baseline, 3 months, and 6 months on their COPD-related ED presentations, hospital admissions, length of hospital stays, number of exacerbations (episodes in which antibiotics or steroids were prescribed or hospital/clinic visits because of a respiratory issue), number of COPD-related visits to a family doctor, number of COPD-related nurse contacts, self-reported use of medication, and self-reported smoking cessation. The number of contacts/calls to the outpatient clinic and deaths were tracked and reported by the clinical project specialist. In addition, hospital admission data and ED usage from the local hospital were also obtained.

Vendor-recorded use data were also documented and sent for analysis at the end of the trial. This included the frequency of recordings for oxygen, blood pressure, temperature, weight, MRC and CAT scores, and the number of times thresholds were exceeded.

Statistical Analysis

Patient characteristics were summarized using descriptive statistics, including mean and SD for continuous variables (if normally distributed) or median, median absolute deviation, and absolute numbers for categorical variables.

All quantitative continuous data were analyzed by conducting a between-group repeated measures analysis of variance (ANOVA) comparing the scores at baseline versus 3-month follow-up and baseline versus 6-month follow-up assessments. This deviation from the original protocol (where we had planned to include all 3 time points in each analysis) was done to maximize the data and avoid excluding participants who did not have data on all 3 time points. Kruskal-Wallis tests were used where data were not normally distributed or group variances were heterogeneous.

Ethics and Dissemination

The study was approved by the research ethics boards of the Markham Stouffville Hospital and Women's College Hospital, Ontario, Canada (protocol version 1.8, December 7, 2018). The study was also retrospectively registered with ClinicalTrials.gov (NCT03741855).

Patient and Public Involvement

During the initial planning stages of the study, we used a co-design approach in the development of the intervention. Patients were given access to the technology for 2 weeks and were subsequently interviewed about their experiences. Health care providers were also interviewed about their current models of care and their experience with the technology. The goal of this process was to establish whether the technology met the needs of its users (patients and health care providers) and to determine whether any modifications to the technology and the service it provided were needed. Modifications to both service and technology were done in response to this feedback. Some of this feedback was also used to inform the decisions about primary and secondary outcome selection.

Patient advisers were not involved directly in the development of the research question and outcome measures or recruitment. The burden of the intervention was assessed by the research ethics boards who had public member representatives. Any participants interested in receiving information about the results of the study will be provided with a summary once the results are available.

Results

Study Participants

A total of 122 patients participated in the study: 40 in the standard care, 41 in the self-monitoring, and 41 in the remote monitoring groups. Of these patients, 7 in the standard care, 5 in the self-monitoring, and 6 in the remote monitoring group did not complete the study (8 patients withdrew from the trial for various reasons; 6 patients were noncompliant with their readings; 4 patients died: 1 from a COPD exacerbation, 1 from complications of comorbid conditions, 1 from cardiac arrest, and 1 from unknown causes; and 1 patient dropped out because of difficulty using the technology; Figure 1). There were no significant differences in the rates of study completion among the groups (P=.80). Patients were excluded from individual analyses, if they had missing data.

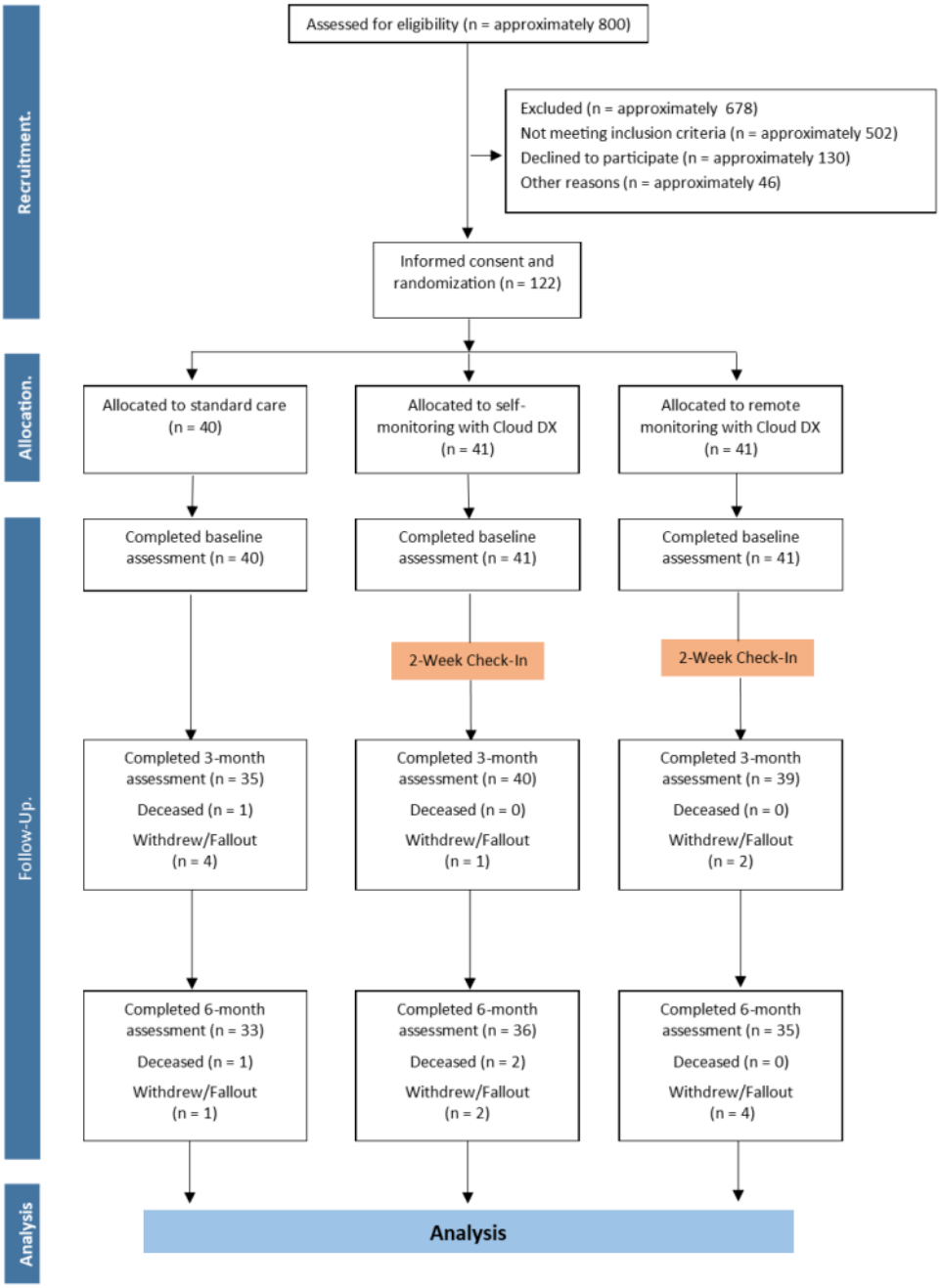
The baseline characteristics of the patients are described in Table 1. Comparisons among groups were made with ANOVA for normally distributed variables, Kruskal-Wallis tests for variables that were not normally distributed, and chi-square tests for categorical data. Patients matched at baseline on all characteristics, except for the CAT baseline scores, where the

XSL•FO
RenderX

self-monitoring group had significantly lower scores than the standard care group (P=.02).

All patients were able to speak and read English, except for 3 patients in the remote monitoring group who were included as they had support from caregivers in completing the questionnaires. There was an equal distribution of education level across the 3 groups (P=.64; Multimedia Appendix 1). Patients from all 3 groups were present in all income brackets, except for the highest income bracket, where there were no self-monitoring patients (P=.01; Multimedia Appendix 1). Patients in all 3 groups were also matched on a series of medical conditions at baseline (Multimedia Appendix 1), except for osteoporosis, for which the rates were lower in the remote monitoring group (P=.02), and pulmonary hypertension, which was reported only in 3 cases, all in the standard care group (P=.04).

Figure 1. Patient flow through the study. Patient flow through each arm of the study. A total of approximately 800 patients were screened for eligibility in order to obtain the final sample of 122 participants.



XSL•FO
RenderX

Table 1. Baseline characteristics of patients in each group.

Characteristics	Standard care group	Self-monitoring group	Remote monitoring group	*P* value
Age (years)				.86
Patients, n	40	41	41	
Mean (SD)	72.78 (9.16)	71.76 (7.28)	71.98 (9.52)	
Gender, n (%)				.93
Female	19 (48)	18 (44)	18 (44)	
Male	21 (52)	23 (56)	23 (56)	
Patients with caregiver (%)	75	83	76	.63
Years since diagnosis of COPD[a]				.32
Patients, n	36	40	38	
Median	4	4.5	7	
MAD[b]	3.71	5.19	4.45	
Currently smoking (%)	23	12	24	.32
Never smoked (%)	3	11	23	.07
Years since quitting smoking				.63
Patients, n	29	31	24	
Median	15	15	13	
MAD	17.79	10.38	11.86	
FEV1 % Pre[c]				.22
Patients, n	35	37	36	
Median	0.45	0.53	0.50	
MAD	0.22	0.13	0.25	
FEV1 (L)[d]				.09
Patients, n	35	37	36	
Median	1.09	1.26	1.17	
MAD	0.47	0.67	0.43	
FEV1/FVC[e]				.49
Patients, n	35	37	36	
Median	0.56	0.58	0.54	
MAD	0.18	0.16	0.19	
COPD Assessment Test				.02
Patients, n	31	41	41	
Mean (SD)	20.42 (7.68)	15.54 (7.65)	19.15 (8.18)	
Systolic blood pressure				.54
Patients, n	25	41	40	
Mean (SD)	129.36 (15.86)	129.90 (20.50)	125.65 (17.34)	
Diastolic blood pressure				.63
Patients, n	25	41	40	
Mean (SD)	76.76 (7.55)	75.37 (10.63)	74.48 (8.83)	
BMI				.08
Patients, n	32	35	31	
Median	24.65	28.10	23.40	

XSL•FO
RenderX

Characteristics	Standard care group	Self-monitoring group	Remote monitoring group	*P* value
MAD	6.67	4.15	6.52	
Never been in exercise/rehabilitation (%)	58	68	80	.08
Never used technology (%)	68	78	85	.16
No medications on hold (%)	63	59	63	.91
COPD exacerbations in the past 12 months				.17
Patients, n	40	40	41	
Median	1	1	2	
MAD	1.48	1.48	2.97	
Emergency department visits in the past 12 months				.61
Patients, n	40	41	41	
Median	0	0	0	
MAD	0	0	0	
Hospitalizations in the past 12 months				.72
Patients, n	39	41	41	
Median	0	0	0	
MAD	0	0	0	
Primary care visits in the past 12 months				.79
Patients, n	38	40	41	
Median	1.5	1	1	
MAD	2.22	1.48	1.48	

[a]COPD: chronic obstructive pulmonary disease.

[b]MAD: median absolute deviation.

[c]FEV1 % pre: % of predicted forced expiratory volume in 1 second.

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

[e]FVC: forced vital capacity.

Readings and Notifications

There were no significant differences in the number of readings completed by each intervention group on any of the measures. There were also no significant differences in the number of notifications received on any of the measures. Patients took their readings almost daily, with a median number of 160 readings in the self-monitoring group and 162 readings in the remote monitoring group over a 182-day period (Multimedia Appendix 1).

Incoming Calls

There were differences among the groups in the number of incoming calls completed during the intervention (P<.001). Post hoc comparisons showed that the standard care group made significantly fewer calls (mean 0.13, SD 0.40) than the self-monitoring (mean 4.17, SD 4.17; P<.001) and the remote monitoring groups (mean 3.27, SD 4.29; P<.001). There was no difference between the number of calls made by the self-monitoring and remote monitoring groups (P=.11).

Primary and Secondary Outcomes

A repeated measures ANOVA showed a significant improvement in PIH scores from baseline to 3 months (P=.001) and from baseline to 6 months (P=.008) but no group effects or interactions, suggesting that there was no differential effect among the groups (Figure 2).

XSL•FO
RenderX

Figure 2. PIH at baseline and 3 months and baseline and 6 months for each group. Significant improvement in PIH scores from baseline to 3 months (P=.001) and from baseline to 6 months (P=.008) were observed, but no group effects or interactions, suggesting no differential effect among the groups. PIH: Partners in Health; RM: remote monitoring SC: standard care; SM: self-monitoring.



A repeated measures ANOVA showed a significant improvement in BCKQ scores from baseline to 3 months (P<.001) and from baseline to 6 months (P<.001; Multimedia Appendix 1). Steeper gains were observed in the remote monitoring group compared with the self-monitoring and standard care groups, both from baseline to 3 months and from baseline to 6 months, but the interaction effect did not reach statistical significance (P=.13 and P=.07, respectively). The gains in accuracy were less than 10%, and all groups had initial scores of just above 30% accuracy, which is lower than the average of 54% accuracy reported by the original BCKQ study [32]. No group main effects were observed.

A repeated measures ANOVA showed no changes in SGRQ activity scores (Multimedia Appendix 1) from baseline to 3 months (P=.49) or from baseline to 6 months (P=.76) and no group effects or interactions.

A repeated measures ANOVA showed a significant improvement in SGRQ impact scores (Multimedia Appendix 1) from baseline to 3 months (P=.047), but no significant group effect or interaction, suggesting that there was no differential effect among the groups. When comparing baseline to 6 months, a repeated measures ANOVA showed a significant effect of time (P=.006) and a significant interaction effect (P=.005). Separate pairwise comparison analyses were performed to examine the interaction effect. The standard care group improved from baseline to 6 months, whereas the remote monitoring group scores deteriorated (higher score) as demonstrated by a significant interaction effect (P=.02). A significant interaction effect (P=.003) was also observed when the self-monitoring and remote monitoring groups were analyzed separately, showing that the self-monitoring group improved, whereas the remote monitoring group worsened. Both standard care and self-monitoring groups improved significantly with time in their SGRQ impact scores (P=.002), and there was no interaction or group effect.

A repeated measures ANOVA showed no changes in SGRQ symptom scores from baseline to 3 months (P=.56) or from baseline to 6 months (P=.62) and no group effects or interactions (Multimedia Appendix 1).

Finally, for the remote monitoring and self-monitoring groups, a repeated measures ANOVA comparing the second CAT readings to their previous CAT reading, with time as a within-subject variable and group as a between-subject variable, showed no significant main effects or interactions. The same was observed with the MRC scores. Therefore, there were no changes in the CAT and MRC scores from the beginning of the intervention until the end.

Correlation Between Changes in Partners in Health Scores and the Number of Readings

We ran a series of Pearson correlations between the number of readings (CAT, MRC, and oxygen saturation) and the change in score from baseline to 6 months for the participants in the self-monitoring and remote monitoring groups. No correlations were observed in any of these analyses.

Health Care Use

In comparing baseline to 3 months, there were no significant effects of time, group, or interaction on any of the measures, except for a decrease in primary care COPD-related visits (P=.04). This reduction was most evident in the standard care and remote monitoring groups. In comparing baseline to 6 months, there were no effects of time, group, or interaction on any of the measures. The incidence of any of the above events was quite low (Multimedia Appendix 1).

In addition to the self-reported measures, the hospital charts of patients were reviewed to estimate the number of ED visits and hospitalizations that had occurred at the hospital during the 6 months preceding their enrollment and during their participation in the trial. The charts were also reviewed to assess the total number of clinic visits that the patients had completed during these periods (clinic visits could be ascertained only for patients who were seeing a physician at the COPD clinic). Nonparametric comparisons were run to estimate the effect of time, group, and time×group interaction. No significant main effects or interactions were observed.

Separate data analyses were performed to compare only COPD-related ED visits and hospitalizations. These analyses showed a significant decrease in COPD-related ED visits during the 6 months before trial enrollment to the 6 months during trial

XSL•FO
RenderX

enrollment (P=.007); however, there was no group effect or interaction, suggesting that there was a general decline in visits across groups. For COPD-related hospital admissions, there was a decrease but not a statistically significant effect across the 3 groups (P=.07; Table 2).

Table 2. Health care use based on hospital data.

Group	Patients, n	Six months before enrollment				Six months post enrollment			
		Mean (SD)	Median	MAD[a]	Maximum value	Mean (SD)	Median	MAD	Maximum value
Emergency department admissions									
SC[b]	40	0.7 (0.99)	0	0	3	0.43 (0.9)	0	0	4
SM[c]	41	0.22 (0.52)	0	0	2	0.32 (0.79)	0	0	4
RM[d]	41	0.46 (0.9)	0	0	4	0.37 (0.77)	0	0	4
Hospital admissions									
SC	40	0.25 (0.54)	0	0	2	0.3 (0.85)	0	0	5
SM	41	0.15 (0.48)	0	0	2	0.12 (0.4)	0	0	2
RM	41	0.32 (0.82)	0	0	4	0.15 (0.42)	0	0	2
Emergency department admissions for COPD[e]									
SC	40	0.38 (0.74)	0	0	3	0.13 (0.52)	0	0	3
SM	41	0.15 (0.48)	0	0	2	0.07 (0.35)	0	0	2
RM	41	0.27 (0.71)	0	0	4	0.1 (0.3)	0	0	1
Hospital admissions for COPD									
SC	40	0.15 (0.43)	0	0	2	0.18 (0.81)	0	0	5
SM	41	0.07 (0.35)	0	0	2	0.02 (0.16)	0	0	1
RM	41	0.2 (0.56)	0	0	3	0.05 (0.22)	0	0	1
Clinic visits									
SC	32	1.25 (1.08)	1	0	6	1.41 (0.98)	1	0	4
SM	35	1.23 (0.97)	1	1.5	4	1.57 (0.78)	1	1.5	4
RM	30	1.37 (0.72)	1	0	3	1.5 (0.97)	1	0.7	4

[a]MAD: median absolute deviation.

[b]SC: standard care.

[c]SM: self-monitoring.

[d]RM: remote monitoring.

[e]COPD: chronic obstructive pulmonary disease.

Discussion

Principal Findings

This study compared the effectiveness of a technology-enabled self-monitoring program to a remote monitoring program and standard care in a population of patients with COPD. Despite high adherence to the intervention and a low dropout rate, the study found no difference in self-efficacy or disease knowledge and disease severity measures among the groups. All 3 groups, including the standard care group, improved self-efficacy and disease knowledge measures. These changes were significant over time and were evident at both 3- and 6-month evaluations. The standard care and self-monitoring groups, but not the remote monitoring group, also reported a lower impact of COPD on their lives when comparing baseline to 6-month evaluations. There were no changes in symptoms or activity scores in any of the groups. There were also no differences (increases or decreases) in patient health care utilization, including ED visits, hospital admissions, primary care visits, or nursing visits, during participants' participation in the trial relative to the 6 months preceding the trial, although these were secondary outcomes that the study was not powered for.

The lack of effect in this study is not unique, as current studies and reviews on the effects of remote monitoring on patient outcomes and health utilization have shown mixed results [16-18]. Some studies have reported positive results on some quality of life measures and symptoms [10,33-35], but many have reported no effects [8,12,36-39]. With respect to health care utilization, the effects are also mixed, with some reporting reductions in hospitalizations [5,7,33,40-42], length of stay [7,33,40,43], and ED visits [33,40], but many failing to find significant effects [6,10,34,36-38]. A recent systematic review concluded that the evidence on the effectiveness of remote

XSL•FO
RenderX

monitoring is mixed [16], although some meta-analyses have reported significant reductions in hospitalizations and ED visits [17,44]. Hong and Lee [17] took their meta-analysis one step further and examined the effects of patient severity and intervention type (interventions with or without an educational component). They concluded that interventions with an educational component (such as those seen in self-management programs) and those targeting patients with more severe diseases had the greatest effects, especially on health care utilization (hospitalizations and ED visits). Our intervention lacked both these components.

First, although spirometry readings on an average suggested moderate to severe disease, there was a large variability among patients, with many patients having milder symptoms. Most patients had no hospital or ED admissions in the 12 months before joining the intervention, which also suggested that even those with moderate to severe disease had enough clinical support to avoid hospitalizations. However, we know that there was room for improvement as all 3 groups improved on several patient outcomes, including self-efficacy and disease knowledge measures. This may mean that the standard care delivered at the local clinic may already have been quite effective in improving patient outcomes, which would have made it harder for us to detect any additional effect of remote monitoring or self-monitoring over those provided in standard care during the same period. If future studies exclude milder patients from similar interventions and find effects, it may provide evidence that remote monitoring and self-monitoring monitoring programs are better suited for moderate to severe patients. This may also make such programs more affordable for health care systems as only a subset of patients will have to be monitored. If the primary goal is to examine the impact on health care utilization (which was not our focus in this study), we recommend focusing on patients with at least one ED or hospital admission, using administrative data sets and examining longer intervention periods.

Second, our intervention did not have a formal educational component. Although the clinical project specialist was making regular calls to patients in the remote monitoring group and providing them with guidance and education when needed, there was no structured educational component in the form of coaching sessions. Some education was also already delivered as part of standard care in the clinic, which may also explain why patients with moderate to severe disease had relatively few hospitalizations. Future studies should include structured coaching sessions, covering topics on self-recognition and self-treatment of exacerbations (eg, taking medications), coping with breathlessness, and lifestyle changes [13] to make the intervention groups more distinct from standard care. More personalized educational components can be delivered by a health care professional in a remote monitoring program, whereas more standardized education modules can be delivered directly through the device in response to readings.

In this intervention, we provided patients with action plans that were not integrated with the devices. Evidence suggests that action plans can be effective in reducing the effects of exacerbations when they are followed, but few patients follow written action plans [45]. The integration of action plans within the platform, with patients receiving feedback directly from the device, may improve their use, although a recent study found no additional benefits when action plans were embedded within a self-monitoring mobile app over those provided by a written action plan [38]. Further developments in trend analyses and predictive analytics of remote monitoring data [46] may allow for early detection of exacerbations as relative changes in vital signs may be more important than detecting vitals reaching absolute thresholds.

Strengths and Limitations

With regard to strengths, this is the first study to directly compare remote monitoring relative to self-monitoring and standard care, and the comparison was conducted through an RCT, which offers strong internal validity for the findings. There was a high adherence rate from both intervention groups with hundreds of oximeter recordings, subjective symptom scores, and blood pressure measurements taken from patients. Furthermore, follow-up was very good, with only 7% of the participants withdrawing, 7% being noncompliant, and only one patient finding the technology to be too difficult to use. Despite this, our design suffered from drawbacks such as a relatively short intervention period (6 months) and inclusion criteria that allowed any patients with a diagnosis of COPD, irrespective of disease severity, to participate. Some of these decisions were made because of time constraints surrounding the trial funding and associated recruitment challenges. Many studies in the literature had disease severity inclusion criteria that often required patients to be admitted at least once and often twice in the previous year [8,42,47]. With respect to intervention duration, a full-year intervention seems to be the most common, but we noted interventions ranging from 3 to 24 months. Although our intervention was only 6 months, it is worth noting that some 6-month interventions have shown positive effects [7,34,35].

Conclusions

Our 6-month intervention comparing technology-enabled remote monitoring and self-monitoring programs showed no intervention specific improvements in self-efficacy, disease knowledge, or quality of life. No effects were observed in health care utilization, including hospital admissions and ED visits. Future studies should focus on patients with higher health care system use and moderate to severe disease. We also recommend including structured educational components (potentially both in remote monitoring and self-monitoring programs) and predictive analytics of vitals data that detect relative rather than absolute changes in vitals.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary results.
[DOCX File , 4543 KB - jmir_v22i7e18598_app1.docx]

Multimedia Appendix 2
Bug fixes and releases.
[DOCX File , 19 KB - jmir_v22i7e18598_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1701 KB - jmir_v22i7e18598_app3.pdf]

References

1. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent ehealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
2. Chronic Respiratory Diseases: Burden of COPD. World Health Organization. URL: https://www.who.int/respiratory/copd/burden/en/ [accessed 2020-02-10]
3. All-Cause Readmission to Acute Care and Return to the Emergency Department. CIHI: Canadian Institute for Health Information. 2012. URL: https://secure.cihi.ca/free_products/Readmission_to_acutecare_en.pdf [accessed 2018-10-18]
4. MacNab M, Lee SH, McCloughan L, Hanley J, McKinstry B, Pinnock H. Oximetry-supported self-management for chronic obstructive pulmonary disease: mixed method feasibility pilot project. BMC Health Serv Res 2015 Oct 26;15:485 [FREE Full text] [doi: 10.1186/s12913-015-1135-2] [Medline: 26503028]
5. Vitacca M, Paneroni M, Grossetti F, Ambrosino N. Is there any additional effect of tele-assistance on long-term care programmes in hypercapnic copd patients? A retrospective study. COPD 2016 Oct;13(5):576-582. [doi: 10.3109/15412555.2016.1147542] [Medline: 27018995]
6. Sorknaes AD, Madsen H, Hallas J, Jest P, Hansen-Nord M. Nurse tele-consultations with discharged COPD patients reduce early readmissions--an interventional study. Clin Respir J 2011 Jan;5(1):26-34. [doi: 10.1111/j.1752-699X.2010.00187.x] [Medline: 21159138]
7. de san Miguel K, Smith J, Lewin G. Telehealth remote monitoring for community-dwelling older adults with chronic obstructive pulmonary disease. Telemed J E Health 2013 Sep;19(9):652-657. [doi: 10.1089/tmj.2012.0244] [Medline: 23808885]
8. Lilholt PH, Udsen FW, Ehlers L, Hejlesen OK. Telehealthcare for patients suffering from chronic obstructive pulmonary disease: effects on health-related quality of life: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 9;7(5):e014587 [FREE Full text] [doi: 10.1136/bmjopen-2016-014587] [Medline: 28490555]
9. Tupper O, Gregersen T, Ringbaek T, Brøndum E, Frausing E, Green A, et al. Effect of tele-health care on quality of life in patients with severe COPD: a randomized clinical trial. Int J Chron Obstruct Pulmon Dis 2018;13:2657-2662 [FREE Full text] [doi: 10.2147/COPD.S164121] [Medline: 30214183]
10. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 3;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
11. Farias R, Sedeno M, Beaucage D, Drouin I, Ouellet I, Joubert A, et al. Innovating the treatment of COPD exacerbations: a phone interactive telesystem to increase COPD action plan adherence. BMJ Open Respir Res 2019;6(1):e000379 [FREE Full text] [doi: 10.1136/bmjresp-2018-000379] [Medline: 31178998]
12. Mudiyanselage SB, Stevens J, Watts J, Toscano J, Kotowicz M, Steinfort C, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul;25(6):343-352. [doi: 10.1177/1357633X18775850] [Medline: 29793387]
13. Yadav UN, Hosseinzadeh H, Baral KP. Self-management and patient activation in COPD patients: an evidence summary of randomized controlled trials. Clin Epidemiol Glob Health 2018 Sep;6(3):148-154 [FREE Full text] [doi: 10.1016/j.cegh.2017.10.004]
14. Wang T, Tan J, Xiao LD, Deng R. Effectiveness of disease-specific self-management education on health outcomes in patients with chronic obstructive pulmonary disease: an updated systematic review and meta-analysis. Patient Educ Couns 2017 Aug;100(8):1432-1446. [doi: 10.1016/j.pec.2017.02.026] [Medline: 28318846]
15. Jolly K, Sidhu M, Bates E, Majothi S, Sitch A, Bayliss S, et al. Systematic review of the effectiveness of community-based self-management interventions among primary care COPD patients. NPJ Prim Care Respir Med 2018 Nov 23;28(1):44. [doi: 10.1038/s41533-018-0111-9] [Medline: 30470741]

XSL•FO
RenderX

16. Kruse C, Pesek B, Anderson M, Brennan K, Comfort H. Telemonitoring to manage chronic obstructive pulmonary disease: systematic literature review. JMIR Med Inform 2019 Mar 20;7(1):e11496 [FREE Full text] [doi: 10.2196/11496] [Medline: 30892276]
17. Hong Y, Lee SH. Effectiveness of tele-monitoring by patient severity and intervention type in chronic obstructive pulmonary disease patients: a systematic review and meta-analysis. Int J Nurs Stud 2019 Apr;92:1-15. [doi: 10.1016/j.ijnurstu.2018.12.006] [Medline: 30690162]
18. Yang F, Wang Y, Yang C, Hu H, Xiong Z. Mobile health applications in self-management of patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis of their efficacy. BMC Pulm Med 2018 Sep 4;18(1):147 [FREE Full text] [doi: 10.1186/s12890-018-0671-z] [Medline: 30180835]
19. Bourbeau J, Collet J, Schwartzman K, Ducruet T, Nault D, Bradley C. Economic benefits of self-management education in COPD. Chest 2006 Dec;130(6):1704-1711. [doi: 10.1378/chest.130.6.1704] [Medline: 17166985]
20. Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2014;9:935-944 [FREE Full text] [doi: 10.2147/COPD.S60179] [Medline: 25246781]
21. Stamenova V, Yang R, Engel K, Liang K, van Lieshout F, Lalingo E, et al. Technology-enabled self-monitoring of chronic obstructive pulmonary disease with or without asynchronous remote monitoring: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Aug 19;8(8):e13920 [FREE Full text] [doi: 10.2196/13920] [Medline: 31429418]
22. van Lieshout F, Yang R, Stamenova V, Agarwal P, Cornejo D, Sidhu A, et al. Evaluating the implementation of a remote monitoring program for chronic obstructive pulmonary disease using qualitative methods: a service design perspective (Preprint). J Med Interent Res 2020:- (forthcoming). [doi: 10.2196/preprints.18148]
23. Dewar M, Curry RW. Chronic obstructive pulmonary disease: diagnostic considerations. Am Fam Physician 2006 Feb 15;73(4):669-676 [FREE Full text] [Medline: 16506711]
24. Urbaniak G, Plous S. Research Randomizer. 2019. URL: https://www.randomizer.org/ [accessed 2019-09-26]
25. Connected Health Monitoring Kit. Cloud DX. URL: https://www.clouddx.com/#/connectedhealthkit [accessed 2019-04-04]
26. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009 Sep;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
27. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988 Mar;93(3):580-586. [doi: 10.1378/chest.93.3.580] [Medline: 3342669]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
30. Petkov J, Harvey P, Battersby M. The internal consistency and construct validity of the partners in health scale: validation of a patient rated chronic condition self-management measure. Qual Life Res 2010 Sep;19(7):1079-1085. [doi: 10.1007/s11136-010-9661-1] [Medline: 20437206]
31. Jones P, Quirk F, Baveystock C. The St George's respiratory questionnaire. Resp Med 1991 Sep;85:25-31. [doi: 10.1016/s0954-6111(06)80166-6]
32. White R, Walker P, Roberts S, Kalisky S, White P. Bristol COPD knowledge questionnaire (BCKQ): testing what we teach patients about COPD. Chron Respir Dis 2006;3(3):123-131. [doi: 10.1191/1479972306cd117oa] [Medline: 16916006]
33. Esteban C, Moraza J, Iriberri M, Aguirre U, Goiria B, Quintana JM, et al. Outcomes of a telemonitoring-based program (telEPOC) in frequently hospitalized COPD patients. Int J Chron Obstruct Pulmon Dis 2016;11:2919-2930 [FREE Full text] [doi: 10.2147/COPD.S115350] [Medline: 27920519]
34. McDowell JE, McClean S, FitzGibbon F, Tate S. A randomised clinical trial of the effectiveness of home-based health care with telemonitoring in patients with COPD. J Telemed Telecare 2015 Mar;21(2):80-87. [doi: 10.1177/1357633X14566575] [Medline: 25586812]
35. Davis C, Bender M, Smith T, Broad J. Feasibility and acute care utilization outcomes of a post-acute transitional telemonitoring program for underserved chronic disease patients. Telemed J E Health 2015 Sep;21(9):705-713. [doi: 10.1089/tmj.2014.0181] [Medline: 25955129]
36. Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al. A multicentre, randomized controlled trial of telehealth for the management of COPD. Respir Med 2018 Nov;144:74-81. [doi: 10.1016/j.rmed.2018.10.008] [Medline: 30366588]
37. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (CHROMED). A randomized clinical trial. Am J Respir Crit Care Med 2018 Sep 1;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]
38. Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 9;7(10):e14408 [FREE Full text] [doi: 10.2196/14408] [Medline: 31599729]

XSL•FO
RenderX

39. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. Br Med J 2013 Oct 17;347:f6070 [FREE Full text] [doi: 10.1136/bmj.f6070] [Medline: 24136634]
40. Achelrod D, Schreyögg J, Stargardt T. Health-economic evaluation of home telemonitoring for COPD in Germany: evidence from a large population-based cohort. Eur J Health Econ 2017 Sep;18(7):869-882 [FREE Full text] [doi: 10.1007/s10198-016-0834-x] [Medline: 27699567]
41. Pedone C, Chiurco D, Scarlata S, Incalzi RA. Efficacy of multiparametric telemonitoring on respiratory outcomes in elderly people with COPD: a randomized controlled trial. BMC Health Serv Res 2013 Mar 6;13:82 [FREE Full text] [doi: 10.1186/1472-6963-13-82] [Medline: 23497109]
42. Lyth J, Lind L, Persson HL, Wiréhn AB. Can a telemonitoring system lead to decreased hospitalization in elderly patients? J Telemed Telecare 2019 Jul 10:1357633X19858178 epub ahead of print. [doi: 10.1177/1357633X19858178] [Medline: 31291794]
43. Shany T, Hession M, Pryce D, Roberts M, Basilakis J, Redmond S, et al. A small-scale randomised controlled trial of home telemonitoring in patients with severe chronic obstructive pulmonary disease. J Telemed Telecare 2017 Aug;23(7):650-656. [doi: 10.1177/1357633X16659410] [Medline: 27464957]
44. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease: Cochrane review and meta-analysis. Br J Gen Pract 2012 Nov;62(604):e739-e749 [FREE Full text] [doi: 10.3399/bjgp12X658269] [Medline: 23211177]
45. Bischoff EW, Hamd DH, Sedeno M, Benedetti A, Schermer TR, Bernard S, et al. Effects of written action plan adherence on COPD exacerbation recovery. Thorax 2011 Jan;66(1):26-31. [doi: 10.1136/thx.2009.127621] [Medline: 21037270]
46. Riis HC, Jensen MH, Cichosz SL, Hejlesen OK. Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. J Med Eng Technol 2016;40(1):1-7. [doi: 10.3109/03091902.2015.1105317] [Medline: 26745746]
47. Udsen FW, Lilholt PH, Hejlesen O, Ehlers L. Cost-effectiveness of telehealthcare to patients with chronic obstructive pulmonary disease: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 17;7(5):e014616 [FREE Full text] [doi: 10.1136/bmjopen-2016-014616] [Medline: 28515193]

Abbreviations

ANOVA: analysis of variance
BCKQ: Bristol COPD Knowledge Questionnaire
CAT: COPD Assessment Test
COPD: chronic obstructive pulmonary disease
ED: emergency department
MRC: Medical Research Council
PIH: Partners in Health
RCT: randomized controlled trial
RT: respiratory therapist
SGRQ: St George's Respiratory Questionnaire

Edited by E Borycki; submitted 12.03.20; peer-reviewed by L Friesen, L Wang, GE Iyawa; comments to author 05.05.20; revised version received 21.05.20; accepted 31.05.20; published 30.07.20.

Please cite as:
Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E, Cheung A, Erwood A, Radina M, Greenwald A, Agarwal P, Sidhu A, Bhatia RS, Shaw J, Shafai R, Bhattacharyya O
Technology-Enabled Self-Management of Chronic Obstructive Pulmonary Disease With or Without Asynchronous Remote Monitoring: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e18598
URL: http://www.jmir.org/2020/7/e18598/
doi:10.2196/18598
PMID:32729843

©Vess Stamenova, Kyle Liang, Rebecca Yang, Katrina Engel, Florence van Lieshout, Elizabeth Lalingo, Angelica Cheung, Adam Erwood, Maria Radina, Allen Greenwald, Payal Agarwal, Aman Sidhu, R Sacha Bhatia, James Shaw, Roshan Shafai, Onil Bhattacharyya. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Characteristics of an Internet-Based Cohort of Patient-Reported Diagnosis of Granulomatosis With Polyangiitis and Microscopic Polyangiitis: Observational Study

Jason Michael Springer[1*], MD, MS; Tanaz A Kermani[2*], MD, MS; Antoine Sreih[3], MD; Dianne G Shaw[4], MA; Kalen Young[4], MA; Cristina M Burroughs[5], BS; Peter A Merkel[3,6], MD, MPH

[1]University of Kansas Medical Center, Kansas City, KS, United States

[2]Division of Rheumatology, Department of Medicine, University of California, Santa Monica, CA, United States

[3]Division of Rheumatology, Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Vasculitis Foundation, Kansas City, MO, United States

[5]Health Informatics Institute, University of South Florida, Tampa, FL, United States

[6]Division of Clinical Epidemiology, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Jason Michael Springer, MD, MS
University of Kansas Medical Center
3901 Rainbow Blvd MS 2026
Kansas City, KS, 66160
United States
Phone: 1 913 588 6009
Email: jspringer2@kumc.edu

Abstract

Background: Utilizing the traditional centers of excellence approach to conduct clinical trials involving rare diseases remains challenging. Patient-based registries have been shown to be both feasible and valid in several other diseases.

Objective: This report outlines the clinical characteristics of a large internet registry cohort of participants with a self-reported diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis.

Methods: Patients with a self-reported diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis in an internet-based prospective longitudinal cohort (from the Vasculitis Patient-Powered Research Network) were included. Data on symptoms, diagnostic testing, and treatment were collected using standardized questionnaires.

Results: The study compared patients with granulomatosis with polyangiitis (n=762) and patients with microscopic polyangiitis (n=164). Of the cohort, 97.7% (904/925) reported the diagnosis had been confirmed by a physician. Compared to microscopic polyangiitis, patients with granulomatosis with polyangiitis reported significantly more ear, nose, and throat manifestations (granulomatosis with polyangiitis: 641/723, 88.7%; microscopic polyangiitis: 89/164, 54.3%; z=10.42, P<.001), fevers (granulomatosis with polyangiitis: 325/588, 55.3%; microscopic polyangiitis: 64/139, 46.0%; z=1.96, P=.05), joint involvement (granulomatosis with polyangiitis: 549/688, 79.8%; microscopic polyangiitis: 106/154, 68.8%; z=2.96, P=.003), and pulmonary involvement (granulomatosis with polyangiitis: 523/734, 71.3%; microscopic polyangiitis: 90/154, 58.4%; z=3.13, P=.002). Compared to microscopic polyangiitis, patients with granulomatosis with polyangiitis reported significantly less renal involvement (granulomatosis with polyangiitis: 457/743, 61.5%; microscopic polyangiitis: 135/163, 82.8%; z=–5.18, P<.001) and renal transplantation (granulomatosis with polyangiitis: 10/721, 1.4%; microscopic polyangiitis: 7/164, 4.3%; z=–2.43, P=.02). Antineutrophil cytoplasmic antibody positivity was reported in 94.2% (652/692) of patients with granulomatosis with polyangiitis and 96.1% (147/153) of patients with microscopic polyangiitis. A biopsy showing vasculitis was reported in 77.0% (562/730) of patients with granulomatosis with polyangiitis and 81.9% (131/160) of patients with microscopic polyangiitis.

Conclusions: In this large, internet-based cohort of patients with a self-reported diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis, disease manifestations were consistent with expectations for each type of vasculitis. Given the rarity

XSL•FO
RenderX

of these and other vasculitides, conducting some types of research through internet-based registries may provide an efficient alternative to inperson, center-of-excellence clinical trials.

(J Med Internet Res 2020;22(7):e17231) doi:10.2196/17231

KEYWORDS

granulomatosis with polyangiitis; microscopic polyangiitis; vasculitis; patient-reported outcomes; patient registry; electronic health records; questionnaire; online cohort

Introduction

Granulomatosis with polyangiitis and microscopic polyangiitis are forms of antineutrophil cytoplasmic antibody–associated vasculitis that primarily target small arteries. These are rare diseases, with annual prevalence of granulomatosis with polyangiitis estimated from 24 to 160 per 1,000,000 and annual prevalence of microscopic polyangiitis estimated from 39 to 94 per 1,000,000 [1]. There are several challenges in conducting clinical trials involving rare diseases, including the need to involve multiple centers, high costs, and other logistical challenges. Novel methods for obtaining both meaningful and reliable data are needed. Multiple studies [2-6] of other diseases have proven the validity of patient-reported diagnoses and outcomes. The aims of this study were to describe the self-reported clinical features of patients with granulomatosis with polyangiitis and microscopic polyangiitis who participated in the Vasculitis Patient-Powered Research Network and to establish to what extent this internet-based cohort is representative of the general population of patients with these forms of antineutrophil cytoplasmic antibody–associated vasculitis.

Methods

Established in 2014, the Vasculitis Patient-Powered Research Network is an international, internet-based prospective longitudinal registry of patient- or caregiver-reported information. The Vasculitis Patient-Powered Research Network was established as a partnership between the Vasculitis Clinical Research Network (a vasculitis research network) and the Vasculitis Foundation (the largest patient advocacy group for vasculitis). The network represents a collaboration among a variety of vasculitis stakeholders including patients, patient advocacy organizations, academic clinical investigators, expert clinicians, biomedical informaticians, methodologists, and funding organizations. Patient-partners are an integral part of team and are involved in strategically planning, developing, reviewing, and approving research studies. Patient-partners receive training in patient participation in research.

The Vasculitis Patient-Powered Research Network is the largest patient-based registry for primary systemic vasculitis with over 3000 patients enrolled to date; patients (or caregivers) consent to participate in studies and provide self-reported information longitudinally using the internet-based platform [7]. For this study, advertisement for recruitment was done via social media, the Vasculitis Foundation website, and flyers at national and regional vasculitis conferences.

Only patients with a self-reported diagnosis of granulomatosis with polyangiitis and microscopic polyangiitis were included in this study. Patients provided consent online and were enrolled between November 2014 and May 2019. The data were obtained via convenience (or opportunity) sampling methods. Using the internet-based patient portal, participants filled out standardized questionnaires (open survey, Multimedia Appendix 1) in English which requested information such as demographics (age, gender, ethnicity), signs and symptoms of vasculitis at the time of diagnosis, results of diagnostic studies performed (ie, laboratory testing, biopsies, imaging, etc), prescribed treatments, and outcomes. For each item, respondents could select *yes*, *no*, or *I don't know* and were able to review or change previously answered questions. Responses that were left blank or where the answer was not known were excluded from the analysis. Patient participation was encouraged by sending email reminders to participants; however, no incentives were offered for completion of the questionnaires. Multiple entries by the same individual were prevented through the use of password-protected user log-ins.

This study was approved by the institutional review board of the University of South Florida. To address data completeness and compliance, the Vasculitis Patient-Powered Research Network operates a comprehensive data compliance strategy using a variety of tools and approaches. The data compliance reports are regularly monitored by the network and data managers of the Vasculitis Patient-Powered Research Network to identify emerging trends. If a participant has not completed all forms, a series of automated email reminders to the participant are triggered. After an initial, generic reminder email message has been sent, a form-specific email reminder is sent. This message specifies which forms are incomplete and also describes the scientific significance and need for the information requested by each form.

Data were analyzed to compare clinical manifestations and diagnostic testing of patients with granulomatosis with polyangiitis to those of patients with microscopic polyangiitis. Two-tailed z scores were performed for comparisons of proportions. Two-tailed independent t tests were used for comparisons of means and medians. P values≤.05 were considered significant.

Results

Participant Characteristics

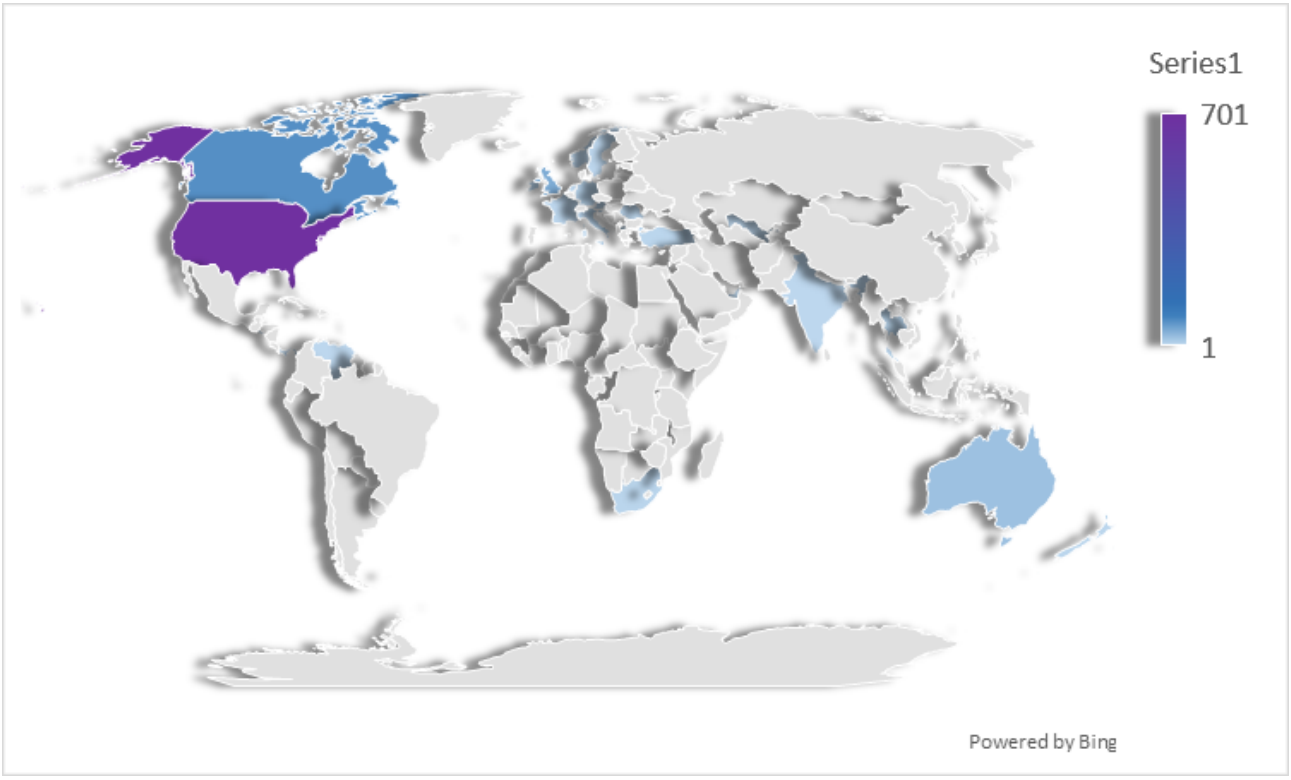
Granulomatosis With Polyangiitis

A total of 762 participants reported a diagnosis of granulomatosis with polyangiitis; 518 (68.0%) were female and

XSL•FO
RenderX

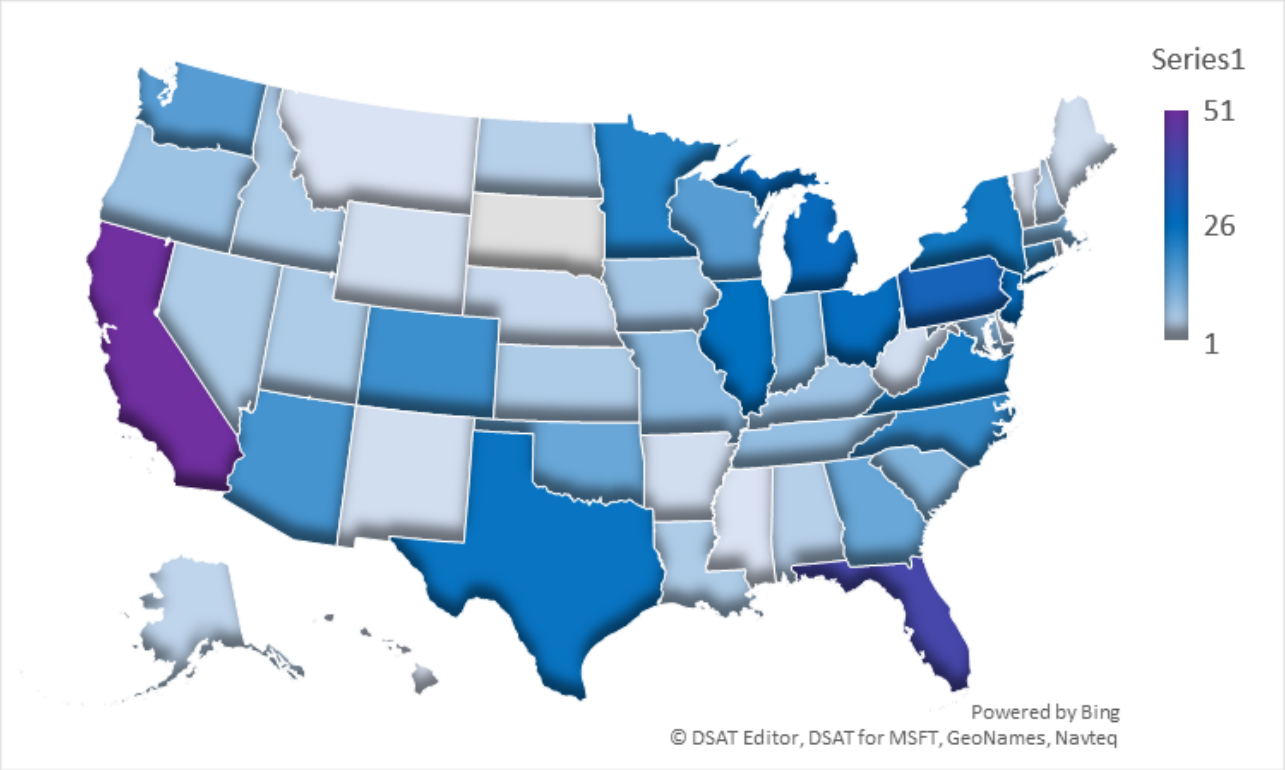
244 (32.0%) were male. The median age of patients at the onset of symptoms was 45 (IQR 31-57) for 619 respondents, and the median age at diagnosis was 48 (IQR 35-57) for 683 respondents. Out of 761 respondents, 248 (32.6%) reported their disease as active, 465 (61.1%) reported their disease as being in remission, and 48 (6.3%) reported they were unsure. Respondents (648/727, 89.1%) reported seeing one or more of the following specialists: rheumatologist (528/727, 72.6%), nephrologist (202/727, 27.8%), pulmonologist (139/727, 19.1%), otolaryngologist (135/727, 18.6%), neurologist (29/727, 4.0%), immunologist (9/727, 1.2%), dermatologist (3/727, 0.4%). Respondents (720/762, 94.5%) reported their country of origin as the United States (572/720, 79.4%, of which state unreported: 72/572 and state reported: 500/572; northeastern states: 102/500, 20.4%; midwestern states: 132/500, 26.4%; southern states: 164/500, 32.8%; western states: 102/500, 20.4%), Canada (62/720, 8.6%), United Kingdom (36/720, 5.0%), Australia (18/720, 2.5%), or other (32/720, 4.4%) (Figure 1 and Figure 2).

Figure 1. Heat maps showing geographic distribution of participants with granulomatosis with polyangiitis and microscopic polyangiitis internationally.



XSL•FO
RenderX

Figure 2. Heat maps showing geographic distribution of participants with granulomatosis with polyangiitis and microscopic polyangiitis within the United States of America.



Microscopic Polyangiitis

A total of 164 participants reported a diagnosis of microscopic polyangiitis; 133 (81.1%) were female and 31 (18.9%) were male. The median age of patients at the onset of symptoms was 52 (IQR 36-61) for 129 respondents, and the median age at diagnosis was 53 (IQR 41-62) for 158 respondents. Out of 160 respondents, 55 (34.4%) reported their disease as active, 96 (60.0%) reported their disease as being in remission, and 13 (8.1%) reported they were unsure. Respondents (141/160, 88.1%) reported seeing one or more of the following specialists: rheumatologist (95/160, 59.4%), nephrologist (85/160, 53.1%), pulmonologist (28/160, 17.5%), neurologist (6/160, 3.8%), immunologist (3/160, 1.9%), dermatologist (2/160, 1.2%) and otolaryngologist (1/160, 0.6%). Respondents (150/164, 91.5%) reported their country of origin as the United States (121/150, 80.6%, of which state unreported: 12/121 and state reported: 109/121; northeastern states: 15/109, 13.8%; midwestern states: 30/109, 27.6%; southern states: 31/109, 28.4%; western states: 33/109, 30.3%), Canada (10/150, 6.6%), United Kingdom (8/150, 5.3%), Australia (5/150, 3.3%), or other (6/150, 4.0%)

Self-Reported Manifestations

In patients with granulomatosis with polyangiitis, the most common self-reported manifestations were nasal or sinus in 82.9% (600/723), joint pain in 79.8% (549/688), pulmonary in 71.3% (523/734), peripheral nerve in 62.8% (411/654), and renal in 61.5% (457/743) of respondents. Pulmonary–renal syndrome (diffuse alveolar hemorrhage in combination with renal disease) was reported by 26.9% (184/684) of participants. Venous thromboembolisms were reported by 13.7% (96/701) of respondents.

In patients with microscopic polyangiitis, the most common self-reported manifestations were renal in 82.8% (135/163), joint pain in 68.8% (106/154), peripheral nerve in 64.6% (95/147), nasal or sinus in 61.0% (89/146), rash in 59.6% (93/156), and pulmonary in 58% (90/154) of respondents. Pulmonary–renal syndrome was reported by 29% (47/161). Venous thromboembolism was reported in 13.6% (21/154) of respondents. Patients also reported isolated renal manifestations without other organ manifestations (14/164, 8.5%), although constitutional symptoms such as myalgia (1/164, 0.6%), fever (2/164, 1.2%), and weight loss (6/164, 3.7%) were reported.

Diagnostic Testing

Confirmation of diagnosis by a physician was reported by 97.5% (742/761) of respondents with granulomatosis with polyangiitis. Patients with granulomatosis with polyangiitis underwent antineutrophil cytoplasmic antibody testing (652/692, 94.2%), laboratory testing (566/762, 74.3%), biopsy (475/762, 62.3%), and imaging (292/762, 38.3%) for diagnosis; symptom-based diagnosis was also reported (533/762, 69.9%). Out of 730 respondents, 562 (77.0%) reported having a biopsy showing vasculitis at some point, of which the biopsy sites included kidney (283/562, 50.4%), lung (192/562, 34.2%), nasal or sinus (146/562, 26.0%), skin (70/562, 12.5%), nerve (10/562, 1.8%), artery (7/562, 1.2%), and other sites (36/562, 6.4%).

Confirmation of diagnosis by a physician was reported by 98.8% (162/164) of respondents with microscopic polyangiitis. Patients with microscopic polyangiitis underwent antineutrophil cytoplasmic antibody testing (147/153, 96.1%), laboratory testing (129/164, 78.7%), biopsy (126/164, 76.8%), and imaging (69/164, 42.1%) for diagnosis; symptom-based diagnosis was also reported (110/164, 67.1%). Of 160 respondents, 131

XSL•FO
RenderX

(81.9%) reported having a biopsy showing vasculitis at some point, of which the biopsy sites included kidney (103/160, 78.6%), lung (20/160, 15.3%), skin (15/160, 11.5%), nerve (6/160, 4.6%), nasal or sinus (4/160, 3.1%), artery (3/160, 2.3%), and other sites (2/160, 1.5%). Additional diagnostic testing information can be found in Multimedia Appendix 2.

Medications

For respondents with granulomatosis, the most commonly used medications included glucocorticoids (707/762, 92.8%), cyclophosphamide (total: 439/762, 57.6%; oral: 313/762, 41.1%; intravenous: 226/762, 29.7%), sulfamethoxazole and trimethoprim (418/762, 54.9%), rituximab (396/762, 52.0%), methotrexate (354/762, 46.5%), azathioprine (311/762, 40.8%), mycophenolate (140/762, 18.4%), and intravenous immunoglobulin (38/762, 5.0%); 8.4% (64/762) reported receiving plasma exchange at some point. For respondents with microscopic granulomatosis, the most commonly used medications included glucocorticoids (151/161, 93.8%), rituximab (89/161, 55.3%), cyclophosphamide (total: 79/161, 49.1%; oral: 44/161, 27.3%; intravenous: 49/161, 30.4%), azathioprine in (79/161, 49.1%), sulfamethoxazole and trimethoprim (53/161, 32.9%), mycophenolate (39/161, 24.2%), methotrexate (34/161, 21.1%), and intravenous immunoglobulin (9/161, 5.6%); 9.9% (16/161) reported receiving plasma exchange at some point. Additional data on the medications used by patients with granulomatosis with polyangiitis and by patients with microscopic polyangiitis to treat their vasculitis can be found in Multimedia Appendix 3.

Granulomatosis With Polyangiitis Versus Microscopic Polyangiitis

Compared to those with a self-reported diagnosis of microscopic polyangiitis, those with a self-reported diagnosis of granulomatosis with polyangiitis were younger at both onset of symptoms (45 years of age versus 52 years of age, t_{746}=–2.46, P=.01) and diagnosis (48 years of age versus 53 years of age, t_{839}=–2.60, P=.01). Patients with granulomatosis with polyangiitis reported more sinonasal disease (83.0% versus 61.0%, z=5.99, P<.001), hearing loss (48.3% versus 15.3%, z=7.28, P<.001), tracheal involvement (30.0% versus 9.8%, z=4.80, P<.001), pulmonary involvement (71.3% versus 58.4%, z=3.13, P=.002), eye involvement (54.2% versus 33.8%, z=3.14, P=.002), joint involvement (79.8% versus 68.8%, z=2.96, P=.003) and fevers (55.3% versus 46.0%, z=1.96, P=.050) than patients with microscopic polyangiitis; whereas, those with microscopic polyangiitis reported more renal involvement (82.8% versus 61.5%, z=–5.18, P<.001) and were more likely to have undergone a renal transplant (4.3% versus 1.4%, z=–2.43, P=.02) than patients with granulomatosis with polyangiitis (Table 1). There was no difference in the proportion of venous thromboembolism (13.7% versus 13.6%, z=0.02, P=.99), no difference in the percentage of patients reporting a positive antineutrophil cytoplasmic antibody test (94.2% versus 96.1%, z=–0.92, P=.36), and no difference in the percentage of patients reporting a diagnosis based on biopsy (77.0% versus 81.9%, z=–1.35, P=.18). There were more diagnoses from lung (34.2% versus 15.3%, z=4.23, P<.001) and nasal or sinus biopsies (26.0% versus 3.1%, z=5.74, P<.001) in participants with granulomatosis with polyangiitis than those in participants with microscopic polyangiitis. Conversely, there were more diagnoses from kidney biopsy in participants with microscopic polyangiitis (78.6% versus 50.4%, z=–5.87, P<.001) than those in participants with granulomatosis with polyangiitis. A biopsy was more likely to have been performed in participants with microscopic polyangiitis than in those with granulomatosis with polyangiitis (76.8% versus 62.3%, z=–3.53, P<.001). Oral cyclophosphamide (z=3.25, P=.001), methotrexate (z=5.92, P<.001), and sulfamethoxazole and trimethoprim (z=5.06, P<.001) were more commonly used by patients with granulomatosis with polyangiitis than by those with microscopic polyangiitis (Multimedia Appendix 3).

XSL•FO
RenderX

Table 1. Self-reported clinical manifestations.

Clinical Manifestation	Granulomatosis with polyangiitis		Microscopic polyangiitis			
	N[a]	n (%)	N[a]	n (%)	*Z* score	*P* value
Rash	698	401 (57.4)	156	93 (59.6)	–0.50	.62
Weight loss	690	400 (58.0)	151	86 (57.0)	0.23	.82
Fever	588	325 (55.3)	139	64 (46.0)	1.96	.05
Joint	688	549 (79.8)	154	106 (68.8)	2.96	.003
Ear, nose, and throat						
Any	723	641 (88.7)	164	89 (54.3)	10.42	<.001
Nasal/sinus symptoms	723	600 (83.0)	146	89 (61.0)	5.99	<.001
Hearing loss	666	322 (48.3)	144	22 (15.3)	7.28	<.001
Tracheal	610	183 (30.0)	133	13 (9.8)	4.80	<.001
Pulmonary						
Any	734	523 (71.3)	154	90 (58.4)	3.13	.002
Alveolar hemorrhage	700	282 (40.3)	161	58 (36.0)	1.00	.32
Renal						
Any	743	457 (61.5)	163	135 (82.8)	–5.18	<.001
Dialysis	721	73 (10.1)	163	18 (11.0)	–0.35	.73
Renal transplant	721	10 (1.4)	164	7 (4.3)	–2.43	.02
Pulmonary–renal	684	184 (26.9)	161	47 (29.2)	–0.59	.56
Peripheral nerve	654	411 (62.8)	147	95 (64.6)	–0.40	.69
Gastrointestinal	683	14 (2.0)	147	4 (2.7)	–0.51	.61
Venous thromboembolism	701	96 (13.7)	154	21 (13.6)	0.02	.99
Pericardial	609	49 (8.0)	132	6 (4.6)	1.39	.16
Eye	609	330 (54.2)	151	51 (33.8)	3.14	.002

[a]The number of patients who responded yes or no (unknown and blank responses were excluded).

Discussion

Principal Findings

This study was conducted using exclusively patient-reported information. It is important to understand how patient-derived data may differ from those obtained through traditional physician reports used in observational cohorts and clinical trials. This is the first large, prospective cohort of patients with a self-reported diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis that included detailed information from standardized forms to evaluate clinical manifestations, the results of diagnostic testing, and types of treatment. The participants were from across the United States, Canada, and from multiple other countries. An analysis [8] of this cohort found that more than 90% of patients met the 1990 American College of Rheumatology classification criteria for granulomatosis with polyangiitis and the Chapel Hill Consensus Conference definition of microscopic polyangiitis.

The type and distribution of clinical manifestations among patients in this cohort were similar to those reported for observational cohorts and in clinical trials [9-11]. Ear, nose, and throat; pulmonary; and renal manifestations were common in granulomatosis with polyangiitis, as expected, while renal and lung involvement were common in microscopic polyangiitis. The proportions of positive antineutrophil cytoplasmic antibody tests (94.2% and 96.1%) were similar to what would be expected and to what has been in the literature [12]. It has been recognized that there is a high risk of venous thromboembolism in antineutrophil cytoplasmic antibody–associated vasculitis, especially during active disease. There was a 14% prevalence of venous thromboembolism in the combined cohort which was similar to that reported in the literature [13,14].

The most common means of diagnosis reported by participants (symptom-based, laboratory testing, biopsy, and imaging) closely reflected what was emphasized by the 1990 American College of Rheumatology classification criteria for granulomatosis with polyangiitis and the Chapel Hill Consensus Conference definition of microscopic polyangiitis. Only 1% of participants were not sure of the means of their diagnosis, reflecting good insight and recall of participants regarding the basis for their original diagnosis.

There were also important differences between this cohort and center-based cohorts. First, there was a higher than expected prevalence of skin (57.4% and 59.6%) and peripheral nerve

XSL•FO
RenderX

involvement (62.8% and 64.6%). This could reflect the nonspecificity of the queries where patients were asked to attribute manifestations to vasculitis. Patients may have had difficulty adjudicating whether their symptoms were caused by vasculitis or as a result of complications of therapy.

Ear, nose, and throat manifestations are a hallmark manifestation for granulomatosis with polyangiitis distinguishing it from microscopic polyangiitis. In a study [15] comparing physician-reported clinical characteristics of granulomatosis with polyangiitis and microscopic polyangiitis in observational cohorts versus randomized controlled trials, 19% patients with microscopic polyangiitis had ear, nose, and throat manifestations compared to 76% patients with granulomatosis with polyangiitis. In this study cohort, there was a statistically higher (z=10.42, P<.001) prevalence of these manifestations in granulomatosis with polyangiitis as expected, but there was also a higher than expected prevalence in microscopic polyangiitis with nasal-sinus manifestations being reported in 54.3% of patients. This may represent the nonspecificity of queries which included asking patients to attribute manifestations to vasculitis. Alternatively, it may in part also reflect misclassification in the medical community in which patients who test positive for myeloperoxidase antineutrophil cytoplasmic antibodies are classified as having microscopic polyangiitis; and only patients who test positive for proteinase 3 antineutrophil cytoplasmic antibodies are classified as having granulomatosis with polyangiitis.

As expected, patients with microscopic polyangiitis were significantly more likely (z=–5.18, P<.001) to have had renal involvement than patients with granulomatosis with polyangiitis. In the microscopic polyangiitis group, 8.5% patients had isolated renal manifestations (renal-limited vasculitis). Overall, the proportion of renal involvement was similar to that reported in the literature [12]; however, in earlier reports, up to one-third of those with antineutrophil cytoplasmic antibody–associated vasculitis with renal involvement went on to develop end-stage renal disease [16] which was higher than what was found in this study (granulomatosis with polyangiitis: 73/457, 16.0%; microscopic polyangiitis: 18/135, 13.3%). This could be related to the improvement of renal outcomes in granulomatosis with polyangiitis and microscopic polyangiitis patients over time reflecting the changes in diagnosis and management [16,17]. This could also reflect the fact that there was a possible bias as a result of less morbidity in participants from the online portal. Furthermore, this study enrolled patients with a clinical diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis; patients who had isolated renal manifestations and part of the spectrum of antineutrophil cytoplasmic antibody–associated vasculitis could have been underrepresented, especially if they had not been clinically diagnosed with granulomatosis with polyangiitis or microscopic polyangiitis by their physicians. Patients with renal-limited disease who have end-stage renal disease may also no longer be followed or monitored for vasculitis, and therefore, may not have enrolled in the study.

Strengths and Limitations

This study had several strengths. First, the study cohort is large for such rare diseases, and is geographically diverse. The data elements of interest were selected and designed by highly experienced investigators in this field and collected using standardized forms. Patient input was obtained at each stage of the process, especially on the design of the user interface for the Vasculitis Patient-Powered Research Network website.

This study had several limitations. First, direct physician confirmation of the diagnosis was not obtained as part of this study; however, based on the patient information that was provided, more than 90% of patients with granulomatosis with polyangiitis and microscopic polyangiitis met either the 1990 American College of Rheumatology Classification Criteria for granulomatosis with polyangiitis or the Chapel Hill Consensus Conference definition for microscopic polyangiitis [8]. Almost all of the participants reported physician confirmation of the diagnosis. In addition, the manifestations described by patients were similar to those reported in the literature. Second, questions that were left blank or for which the participant answered, “I don’t know” were not included in the analysis. This may represent a response bias that is common among studies using self-reported data; however, for most questions this represented a small percentage of responses. Third, bias sampling may have led to an overrepresentation of certain subgroups. For instance, there were more women than men (2:1 for granulomatosis with polyangiitis and 4:3 for microscopic polyangiitis), despite the sex ratio in the general population being close to 1 for both diseases [1]. Internet-based participation in many surveys and studies is more common among women [18-20]. Age may also have been a factor in participation in this internet-based cohort. This appears to be less of an issue in this study which includes patients from across the age spectrum, including individuals older than 70 years. Fourth, the mode of survey response (internet) may have prevented some patients who lack access or higher education and since many of the participants were recruited through the Vasculitis Foundation, and it may also represent a bias toward participants who are savvier about their disease.

Conclusions

Granulomatosis with polyangiitis and microscopic polyangiitis are both rare (also known as orphan diseases), which makes clinical research for these disorders difficult and requires the use of multiple centers which can be financially and logistically challenging. Patient registries offer an alternative to the centers of excellence approach to conducting clinical research. The data reported herein for an internet-based cohort demonstrated the feasibility of such registries across broad geographic regions and the high level of comparability between an internet-based and traditional academic center-based participant populations. These data provide investigators and patients with confidence that internet-based patient-reported data are reliable and can be used to conduct novel, cost-efficient medical research. Such internet-based registries offer an advantage in capturing participant data from those who would not otherwise be able to participate in studies.

XSL•FO
RenderX

Acknowledgments

The Vasculitis Patient-Powered Research Network has received support from the Patient-Centered Research Institute (PCORI award PPRN-1306-04758) and through the Vasculitis Clinical Research Consortium. The Vasculitis Clinical Research Consortium is part of the Rare Diseases Clinical Research Network, an initiative of the Office of Rare Diseases Research, National Center for Advancing Translational Science. The Vasculitis Clinical Research Consortium is funded through a collaboration between National Center for Advancing Translational Science and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (U54 AR057319).

Conflicts of Interest

PAM reports receiving funds for various activities in the past two years. They have done consulting for the following institutions: AbbVie, AstraZeneca, Biogen, Boeringher-Ingelheim, Bristol-Myers Squibb, Celgene, ChemoCentryx, CSL Behring, Forbius, Genentech/Roche, Genzyme/Sanofi, GlaxoSmithKline, InflaRx, Insmed, Jannsen, Kiniksa, Magenta, Pfizer, Sparrow, Talaris. They recieved research support from the following institutions: AstraZeneca, Boeringher-Ingelheim, Bristol-Myers Squibb, Celgene, ChemoCentryx, Forbius, Genentech/Roche, Genzyme/Sanofi, GlaxoSmithKline, InflaRx. They recieve royalties from the following institutions: UpToDate.

Multimedia Appendix 1
Vasculitis Patient-Powered Research Network online questionnaire forms for granulomatosis with polyangiitis and microscopic polyangiitis.
[DOCX File , 65 KB - jmir_v22i7e17231_app1.docx]

Multimedia Appendix 2
Patient-reported testing methods to establish a diagnosis of granulomatosis with polyangiitis or microscopic polyangiitis.
[DOCX File , 15 KB - jmir_v22i7e17231_app2.docx]

Multimedia Appendix 3
Patient-reported medications used to treat granulomatosis with polyangiitis and microscopic polyangiitis.
[DOCX File , 14 KB - jmir_v22i7e17231_app3.docx]

References

1. Ntatsaki E, Watts RA, Scott DGI. Epidemiology of ANCA-associated vasculitis. Rheum Dis Clin North Am 2010 Aug;36(3):447-461. [doi: 10.1016/j.rdc.2010.04.002] [Medline: 20688243]
2. Peeters GMEE, Tett SE, Dobson AJ, Mishra GD. Validity of self-reported osteoporosis in mid-age and older women. Osteoporos Int 2013 Mar;24(3):917-927. [doi: 10.1007/s00198-012-2033-7] [Medline: 22673826]
3. Randell RL, Long MD, Cook SF, Wrennall CED, Chen W, Martin CF, et al. Validation of an internet-based cohort of inflammatory bowel disease (CCFA partners). Inflamm Bowel Dis 2014 Mar;20(3):541-544 [FREE Full text] [doi: 10.1097/01.MIB.0000441348.32570.34] [Medline: 24451221]
4. Pastorino S, Richards M, Hardy R, Abington J, Wills A, Kuh D, National Survey of HealthDevelopment ScientificData Collection Teams. Validation of self-reported diagnosis of diabetes in the 1946 British birth cohort. Prim Care Diabetes 2015 Oct;9(5):397-400 [FREE Full text] [doi: 10.1016/j.pcd.2014.05.003] [Medline: 25037637]
5. Jackson CA, Mishra GD, Tooth L, Byles J, Dobson A. Moderate agreement between self-reported stroke and hospital-recorded stroke in two cohorts of Australian women: a validation study. BMC Med Res Methodol 2015 Jan 23;15:7 [FREE Full text] [doi: 10.1186/1471-2288-15-7] [Medline: 25613556]
6. Eichler GS, Cochin E, Han J, Hu S, Vaughan TE, Wicks P, et al. Exploring concordance of patient-reported information on PatientsLikeMe and medical claims data at the patient level. J Med Internet Res 2016 May 12;18(5):e110 [FREE Full text] [doi: 10.2196/jmir.5130] [Medline: 27174602]
7. Vasculitis-Patient Powered Research Network website. URL: https://www.vpprn.org/ [accessed 2020-03-30]
8. Kermani TS, Sreih A, Shaw D, Young K, Burroughs C, Merkel P. Arthritis Rheum. In: Accuracy of self-reported diagnosis of antineutrophil cytoplasmic antibody-associated vasculitis abstract. 2019 Presented at: American College of Rheumatology Annual Meeting; November 8-13 2019; Atlanta URL: https://acrabstracts.org/abstract/accuracy-of-self-reported-diagnosis-of-antineutrophil-cytoplasmic-antibody-associated-vasculitis/
9. Stone JH, Merkel PA, Spiera R, Seo P, Langford CA, Hoffman GS, RAVE-ITN Research Group. Rituximab versus cyclophosphamide for ANCA-associated vasculitis. N Engl J Med 2010 Jul 15;363(3):221-232 [FREE Full text] [doi: 10.1056/NEJMoa0909905] [Medline: 20647199]
10. Pagnoux C, Mahr A, Hamidou MA, Boffa J, Ruivard M, Ducroix J, French Vasculitis Study Group. Azathioprine or methotrexate maintenance for ANCA-associated vasculitis. N Engl J Med 2008 Dec 25;359(26):2790-2803. [doi: 10.1056/NEJMoa0802311] [Medline: 19109574]

XSL•FO
RenderX

11. Guillevin L, Pagnoux C, Karras A, Khouatra C, Aumaître O, Cohen P, French Vasculitis Study Group. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. N Engl J Med 2014 Nov 06;371(19):1771-1780. [doi: 10.1056/NEJMoa1404231] [Medline: 25372085]
12. Solans-Laqué R, Fraile G, Rodriguez-Carballeira M, Caminal L, Castillo MJ, Martínez-Valle F, Spanish Registry of systemic vasculitis (REVAS) from the Autoimmune Diseases Study Group (GEAS) of the Spanish Society of Internal Medicine (SEMI). Clinical characteristics and outcome of Spanish patients with ANCA-associated vasculitides: impact of the vasculitis type, ANCA specificity, and treatment on mortality and morbidity. Medicine (Baltimore) 2017 Feb;96(8):e6083 [FREE Full text] [doi: 10.1097/MD.0000000000006083] [Medline: 28225490]
13. Merkel PA, Lo GH, Holbrook JT, Tibbs AK, Allen NB, Davis JC, Wegener's Granulomatosis Etanercept Trial Research Group. Brief communication: high incidence of venous thrombotic events among patients with Wegener granulomatosis: the Wegener's Clinical Occurrence of Thrombosis (WeCLOT) Study. Ann Intern Med 2005 Apr 19;142(8):620-626. [doi: 10.7326/0003-4819-142-8-200505030-00011] [Medline: 15838068]
14. Allenbach Y, Seror R, Pagnoux C, Teixeira L, Guilpain P, Guillevin L, French Vasculitis Study Group. High frequency of venous thromboembolic events in Churg-Strauss syndrome, Wegener's granulomatosis and microscopic polyangiitis but not polyarteritis nodosa: a systematic retrospective study on 1130 patients. Ann Rheum Dis 2009 Apr;68(4):564-567. [doi: 10.1136/ard.2008.099051] [Medline: 19015208]
15. Pagnoux C, Carette S, Khalidi NA, Walsh M, Hiemstra TF, Cuthbertson D, french Vasculitis Study Group (FVSG), European Vasculitis Society (EUVAS)Vasculitis Clinical Research Consortium (VCRC). Comparability of patients with ANCA-associated vasculitis enrolled in clinical trials or in observational cohorts. Clin Exp Rheumatol 2015;33(2 Suppl 89):S-77 [FREE Full text] [Medline: 26016754]
16. Rhee RL, Hogan SL, Poulton CJ, McGregor JAG, Landis JR, Falk RJ, et al. Trends in long-term outcomes among patients with antineutrophil cytoplasmic antibody-associated vasculitis with renal disease. Arthritis Rheumatol 2016 Jul;68(7):1711-1720 [FREE Full text] [doi: 10.1002/art.39614] [Medline: 26814428]
17. Whatmough S, Fernandez S, Sweeney N, Howell L, Dhaygude A. Comparing outcomes of biopsy-proven anti-neutrophil cytoplasmic autoantibody-associated glomerulonephritis patients treated with cyclophosphamide in the 20th and 21st centuries: a 23-year study. Clin Kidney J 2019 Feb;12(1):42-48 [FREE Full text] [doi: 10.1093/ckj/sfy084] [Medline: 30746129]
18. Curtin R, Presser S, Singer E. The effects of response rate changes on the index of consumer sentiment. Public Opin Q 2000;64(4):413-428. [doi: 10.1086/318638] [Medline: 11171024]
19. Singer E, Van Hoewyk J, Maher MP. Experiments with incentives in telephone surveys. Public Opin Q 2000;64(2):171-188. [doi: 10.1086/317761] [Medline: 10984332]
20. Abdou NI, Kullman GJ, Hoffman GS, Sharp GC, Specks U, McDonald T, et al. Wegener's granulomatosis: survey of 701 patients in North America. Changes in outcome in the 1990s. J Rheumatol 2002 Feb;29(2):309-316. [Medline: 11838848]

Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by R Richesson, D Jayne; comments to author 01.04.20; revised version received 30.04.20; accepted 14.05.20; published 20.07.20.

Please cite as:
Springer JM, Kermani TA, Sreih A, Shaw DG, Young K, Burroughs CM, Merkel PA
Clinical Characteristics of an Internet-Based Cohort of Patient-Reported Diagnosis of Granulomatosis With Polyangiitis and Microscopic Polyangiitis: Observational Study
J Med Internet Res 2020;22(7):e17231
URL: https://www.jmir.org/2020/7/e17231
doi:10.2196/17231
PMID:32459634

©Jason Michael Springer, Tanaz A Kermani, Antoine Sreih, Dianne G Shaw, Kalen Young, Cristina M Burroughs, Peter A Merkel. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Worldwide Prevalence of Hearing Loss Among Smartphone Users: Cross-Sectional Study Using a Mobile-Based App

Marcin Masalski[1,2], MD, PhD, BEng; Krzysztof Morawski[1], MD, PhD

[1]Department of Otolaryngology Head and Neck Surgery, Faculty of Medicine, Wroclaw Medical University, Wroclaw, Poland

[2]Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wroclaw University of Technology, Wroclaw, Poland

Corresponding Author:
Marcin Masalski, MD, PhD, BEng
Department of Otolaryngology Head and Neck Surgery
Faculty of Medicine
Wroclaw Medical University
Wybrzeze Ludwika Pasteura 1
Wroclaw, 50-367
Poland
Phone: 48 515086252
Email: marcin.masalski@pwr.edu.pl

Abstract

Background: In addition to the aging process, risk factors for hearing loss in adults include, among others, exposure to noise, use of ototoxic drugs, genetics, and limited access to medical care. Differences in exposure to these factors are bound to be reflected in the prevalence of hearing loss. Assessment of hearing loss can easily be carried out on a large scale and at low cost using mobile apps.

Objective: This study aimed to conduct a worldwide assessment of the differences in hearing loss prevalence between countries in a group of mobile device users.

Methods: Hearing tests were conducted using the open-access Android-based mobile app Hearing Test. The app is available free of charge in the Google Play store, provided that consent to the use of the results for scientific purposes is given. This study included hearing tests carried out on device models supported by the app with bundled headphones in the set. Calibration factors for supported models were determined using the biological method. The tests consisted of self-determining the quietest audible tone in the frequency range from 250 Hz to 8 kHz by adjusting its intensity using the buttons. The ambient noise level was optionally monitored using a built-in microphone. Following the test, the user could compare his hearing threshold against age norms by providing his or her age. The user's location was identified based on the phone’s IP address.

Results: From November 23, 2016 to November 22, 2019, 733,716 hearing tests were conducted on 236,716 mobile devices across 212 countries. After rejecting the tests that were incomplete, performed with disconnected headphones, not meeting the time criterion, repeated by the same user, or carried out regularly on one device, 116,733 of 733,716 tests (15.9%) were qualified for further analysis. The prevalence of hearing loss, defined as the average threshold at frequencies 0.5 kHz, 1 kHz, 2 kHz, and 4 kHz above 25 dB HL in the better ear, was calculated at 15.6% (95% CI 15.4-15.8). Statistically significant differences were found between countries (P<.001), with the highest prevalences for Bangladesh, Pakistan, and India (>28%) and the lowest prevalences for Taiwan, Finland, and South Korea (<11%).

Conclusions: Hearing thresholds measured by means of mobile devices were congruent with the literature data on worldwide hearing loss prevalence. Uniform recruitment criteria simplify the comparison of the hearing loss prevalence across countries. Hearing testing on mobile devices may be a valid tool in epidemiological studies carried out on a large scale.

(J Med Internet Res 2020;22(7):e17238) doi:10.2196/17238

KEYWORDS

hearing loss; epidemiology; mobile-based; hearing test; pure-tone audiometry

XSL•FO
RenderX

Introduction

The worldwide prevalence of hearing loss is estimated to be between 4.0% and 18.1% [1-5], depending on the methodology, in particular the degree of hearing impairment used in the definition of hearing loss. Methodological differences also include the selection of data sources and calculation methods. Despite the uncertainty of the estimation, the scale of the burden is significant. Hearing loss is also an important problem because of its side effects. Due to communication disorders, hearing loss leads to decreased social activity, lower self-esteem, and consequently, stigmatization, social exclusion, and depression [6,7].

Apart from age, risk factors for hearing loss in adults include, but are not limited to, exposure to noise, use of ototoxic drugs, genetic conditions, infectious diseases, and limited access to medical care [5,8-12]. In view of increasing exposure to risk factors for hearing loss, in particular the aging population and increased exposure to noise in developing countries, the number of people with hearing loss has increased over the years [4,5,12].

Epidemiological analysis of hearing loss allows the identification of causes and planning of remedies [4,12]. Activities in the field of prevention and treatment of hearing loss are highly recommended, as it is estimated that 50% of hearing loss cases can be prevented and a significant portion of the remaining cases can be treated [4].

Worldwide geo-epidemiological investigations of hearing loss have been conducted in the form of meta-analyses of available data from different studies as a consequence of political, financial, cultural, and geographical limitations in the planning of uniform research on such a large scale [1-3,13]. Owing to the differences in study settings, in particular the differences in the definition of hearing loss, age groups, and method of participant recruitment, direct comparison of results is not possible, whereas indirect comparison is burdened with uncertainty resulting from sparse data adjustment [1-3,13].

Self-performed hearing tests on mobile devices can be easily carried out on a large scale and at low cost while maintaining a uniform qualification criterion. They are broadly accessible due to widespread use of smartphones, and they do not require engaging qualified personnel, while their results are comparable to pure-tone audiometry [14-23]. The difference in hearing threshold between the self-test in the Hearing Test app and pure-tone audiometry is estimated at 2.6 dB (SD 8.3 dB) [24], and the absolute difference is estimated at 8.8 dB [17].

The aim of this study was to assess the worldwide prevalence of hearing loss among Android users by means of the Hearing Test app [24,25] and compare the prevalence across countries.

Methods

Ethical Concerns

This was a cross-sectional study of hearing loss prevalence carried out by means of the open-access, mobile-based Hearing Test app [26]. Subject consent to use the results for scientific purposes was required the first time the app was started. The consent to research was issued by the Bioethics Committee at the Medical University in Wroclaw. Evaluation of the app, rationale behind the app, and comparison with other studies concerning mobile and web-based pure-tone hearing screening have been presented in prior articles [24,25,27,28].

Recruitment

The participants were recruited via the Hearing Test app that runs on Android systems. The app is offered free of charge in the Google Play store, where the user can become familiar with the app's features and download it to a phone or tablet. Thus, eligibility criteria for participants only included access to an Android device that is supported by the app. The test could be carried out by the phone owner or with his or her assistance.

Device Calibration

The study included tests carried out on mobile devices that supported the Hearing Test app for which calibration coefficients had been determined. All the tests in the study were carried out on bundled headphones that were supplied by the manufacturer in a set with the device. Devices of the same model used common calibration coefficients that were determined by the biological method [27,28] on the basis of at least 16 measurements carried out by subjects with normal hearing on different devices of the model [25]. This method is characterized by the standard error of determining the calibration coefficients below 5 dB and within-model variability at 4 dB [25]. The headphone connection status was monitored by the app, whereas the usage of bundled headphones was confirmed by the user. Tests taken using headphones other than bundled headphones were not taken into account.

Hearing Threshold

The hearing test consisted of self-determining the quietest audible sound by adjusting its intensity using the “I can hear” and “I can't hear” buttons. The intensity of a test tone could be changed many times. The quietest audible sound was confirmed with the “Barely audible” button. A modulated test tone was used, with a 100% modulation depth, 2 Hz modulation frequency, and an intensity changed in steps of 5 dB. When the intensity of the tone exceeded a 40 dB hearing level (HL), a narrowband masking noise was generated contralaterally at an intensity of 40 dB HL. For test tones above 60 dB HL, the intensity of the masking noise was increased to 60 dB HL. By default, the tests were carried out for frequencies from 250 Hz to 8 kHz; however, this range could be changed in the settings. The time to determine the hearing threshold was measured separately for each frequency. The hearing threshold determined in this way has been compared in previous works with pure-tone audiometry. A difference was found at the level of 2.6 dB (SD 8.3) [24], and the absolute difference was found at 8.8 dB [17].

Ambient Noise

Ambient noise was monitored during the measurement after obtaining the user's consent to access the microphone resources. An equivalent continuous A-weighted sound pressure level LAeq was registered in accordance with the Android specification [29] that sets a reference point at 90 dB SPL.

XSL•FO
RenderX

Age

At the end of the test, the users could optionally enter their age to compare the obtained hearing threshold against the age-based norm. By design, the app was developed for adults. Therefore, the user chose his or her age by selecting a value from 18-90 years. Providing the age was not obligatory for this study, to minimize the amount of potentially incorrect data. All the entered age values were recorded in the database to eliminate tests for which several different values were given. The age was also determined from the note, where it could optionally be given. User gender was not collected.

Geolocalization

The user's country was determined using the geoPlugin service [30] on the basis of the IP address.

Statistical Analysis

Power

Comparison of hearing loss prevalence between countries was conducted by adopting the definition of hearing loss as the average hearing threshold at frequencies 0.5 kHz, 1 kHz, 2 kHz, and 4 kHz above 25 dB HL in the better ear. The minimum number of tests for the country to be included in the analysis was calculated based on the standard deviation of the average hearing threshold determined in preliminary measurements at 17 dB. Assuming a level of statistical significance of .05, test power of .8, and an effect size of 5.0 dB, a sample size of 90 tests was obtained.

Data Exclusion

The tests in this study were unsupervised. Therefore, prior to analysis, they were verified for duration. A very short duration suggests unreliable results that can be manifested by an increased measurement error. To eliminate these tests, the standard deviation of the hearing threshold grand average was analyzed in relation to the test duration. Time threshold was set at the level at which stabilization of the standard deviation was observed. Incomplete tests, test carried out without headphones connected, and tests repeated on the same device were excluded as well.

Data Analysis

Both the subject's age and level of ambient noise during the test belong to data for which acquisition required additional action on the part of the user. Preliminary analyses have shown that these are sparse data compared to the hearing threshold, geolocation, and device model. Therefore, the country-specific hearing loss prevalence was calculated based on all the tests, whereas the bias was estimated and discussed basing on subgroup analysis. The tests with age were used to characterize the population and analyze an age-related hearing threshold, while tests with noise were used to calculate the ambient noise effect. Distributions of hearing thresholds in the subgroups were tested for equivalence using the Kolmogorov-Smirnov test. Confidence intervals for the hearing threshold median and the hearing loss prevalence were determined by means of bootstrapping.

Results

Data Collection and Exclusion

In the period from November 23, 2016 to November 22, 2019, 733,716 tests were carried out on 236,716 devices in 212 countries. The hearing threshold measurement at the fundamental frequencies 0.25 kHz, 0.5 kHz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz was performed in 728,674 of 733,716 (99.3%) tests. After examining headphone connection status and rejecting the tests carried out with disconnected headphones, 637,169 of 733,716 (86.8%) tests were obtained (Figure 1).

Time limits were introduced to discard tests with a duration that actually prevented a correct measurement. The grand standard deviation of the hearing threshold in relation to the duration of the measurement at a single frequency is presented in Figure 2. Stabilization of the deviation is observed at longer durations. It has been assumed that tests in which the measurement at any frequency was carried out quicker than 6.7 seconds are subject to significant error. This threshold was set heuristically by fitting the data with a lognormal distribution and assuming its value at a cumulative probability density level of 0.99 (Figure 2). The time requirements were met by 239,752 of 733,716 (32.6%) tests.

Many devices were used for more than one test. To eliminate tests repeated by the same person or carried out in bulk (eg, in outpatient clinics or as part of screening programs) only one test from each device was qualified for the analysis. The last test was selected, at first considering the test with an age provided. Finally, 116,733 of 733,716 (15.9%) tests were obtained for the analysis. Age was provided for 8194 of 116,733 (7.0%) tests, and monitoring for ambient noise was present in 30,119 of 116,733 (25.8%) tests, whereas only 2556 of 116,733 (2.2%) tests included both age and noise data.

XSL•FO
RenderX

Figure 1. Flow diagram.

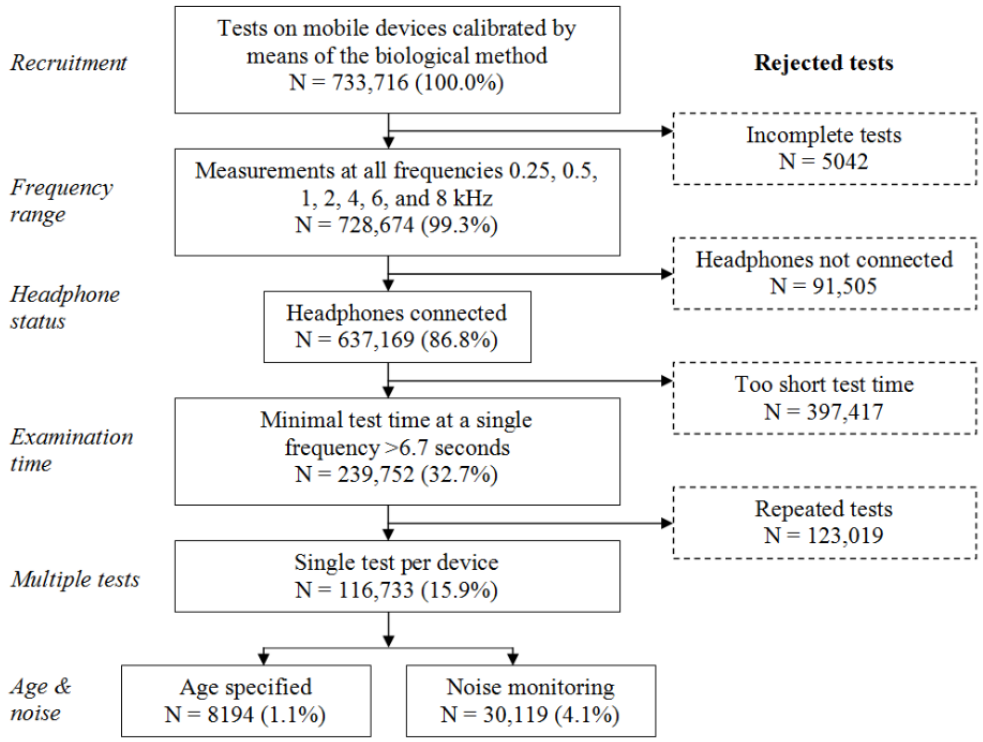
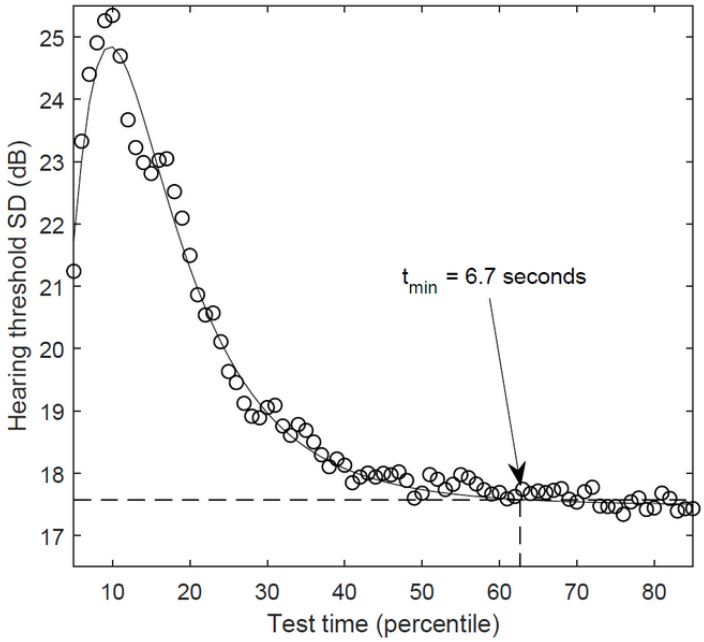


Figure 2. Grand standard deviation of the hearing threshold in relation to the duration of the measurement at a single frequency.



Age

The subject's age was provided for 8194 of 116,733 (7.0%) tests that qualified for the analysis. The age characteristics of the research group is shown in Table 1. The age group with the highest number of tests was 30-39 years, while the fewest subjects were in the oldest groups (80-89 years and >89 years). The median age was 39 years, and the average age was 40.0 years.

Tests with age provided were compared with those without age provided to estimate the bias resulting from the generalization of the results for the whole group. There were no statistically significant differences between the distributions of the hearing thresholds at the level of P=.05.

XSL•FO
RenderX

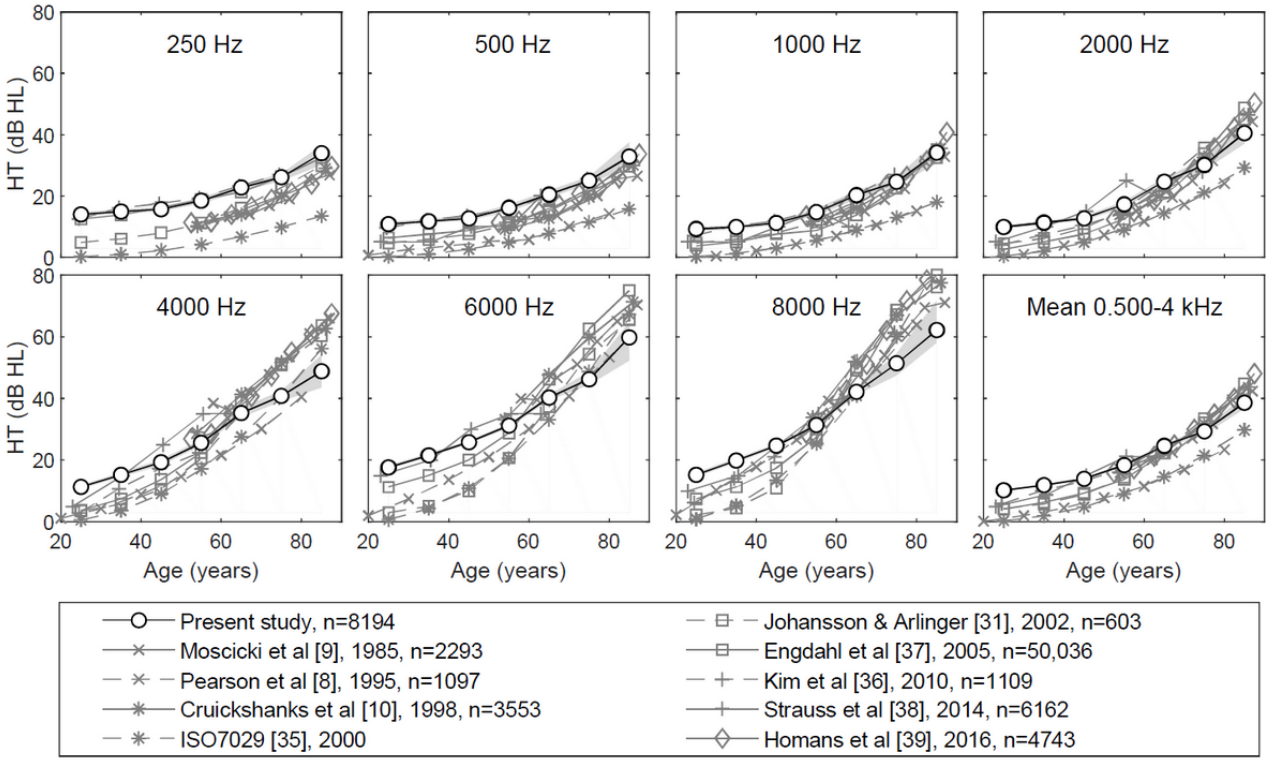
Table 1. Ages of the participants (n=8194).

Age characteristics	Participant values
Age (years), mean (SD)	40.0 (15.1)
Age (years), median	39
Age (years), n (%)	
<20	842 (10.3)
20-29	1597 (19.5)
30-39	1813 (22.1)
40-49	1725 (21.1)
50-59	1227 (15.0)
60-69	655 (8.0)
70-79	253 (3.1)
80-89	62 (0.8)
>89	20 (0.2)

Hearing Threshold

Hearing threshold in relation to age was analyzed based on all 7332 of 116,733 (6.3%) tests for which the age provided was within the range of 20 to 89 years. Tests completed by subjects aged 18, 19, or 90 years were omitted from the analysis to obtain equal age ranges (20-29 to 80-89 years). Moreover, the ages of 18 and 90 years were characterized by a higher number of tests in comparison with the adjacent age groups, suggesting the presence of additional data, such as results for subjects younger than 18 years or older than 90 years, or with incorrectly provided data. The median thresholds by age group and frequency are given in Figure 3 and in Multimedia Appendix 1. A comparison with the prior work presented in Figure 3 is provided in the Discussion section.

Figure 3. Hearing threshold by age group and frequency. Dashed lines indicate screening studies including screening only for noise exposure [31], and continuous lines show studies without screening. The grey area illustrates the 99% CI.



Country Prevalence of Hearing Loss

The prevalence of hearing loss was determined based on 733,716 tests conducted in 212 countries. The required sample size of 90 tests was reached for 74 countries. The prevalence of hearing loss, defined as a mean hearing threshold >25 dB HL in the better ear at 0.5-4 kHz, is presented in Figure 4, Figure 5, and

XSL•FO
RenderX

Multimedia Appendix 2. The highest prevalences were obtained for Pakistan, Bangladesh, and India at 37.8% (95% CI 31.4-44.2), 32.2% (95% CI 23.6-40.8), and 28.5% (95% CI 27.1-29.9), respectively, whereas the lowest prevalences were obtained for Taiwan, Finland, and South Korea at 9.6% (95% CI 7.2-12.1), 9.8% (95% CI 5.7-13.9), and 10.2% (95% CI 9.5-10.9), respectively. The global prevalence of hearing loss was calculated at 15.6% (95% CI 15.4-15.8).

Figure 4. The prevalence of hearing loss by country. Hearing loss was defined as an average hearing threshold >25 dB hearing level (HL) in the better ear at frequencies of 0.5 kHz, 1 kHz, 2 kHz, and 4 kHz. Whiskers indicate 95% CI. Countries with >1000 tests or with boundary values of hearing loss prevalence are included. The full list is available in Multimedia Appendix 2. ppm: number of individuals per million people in the population.

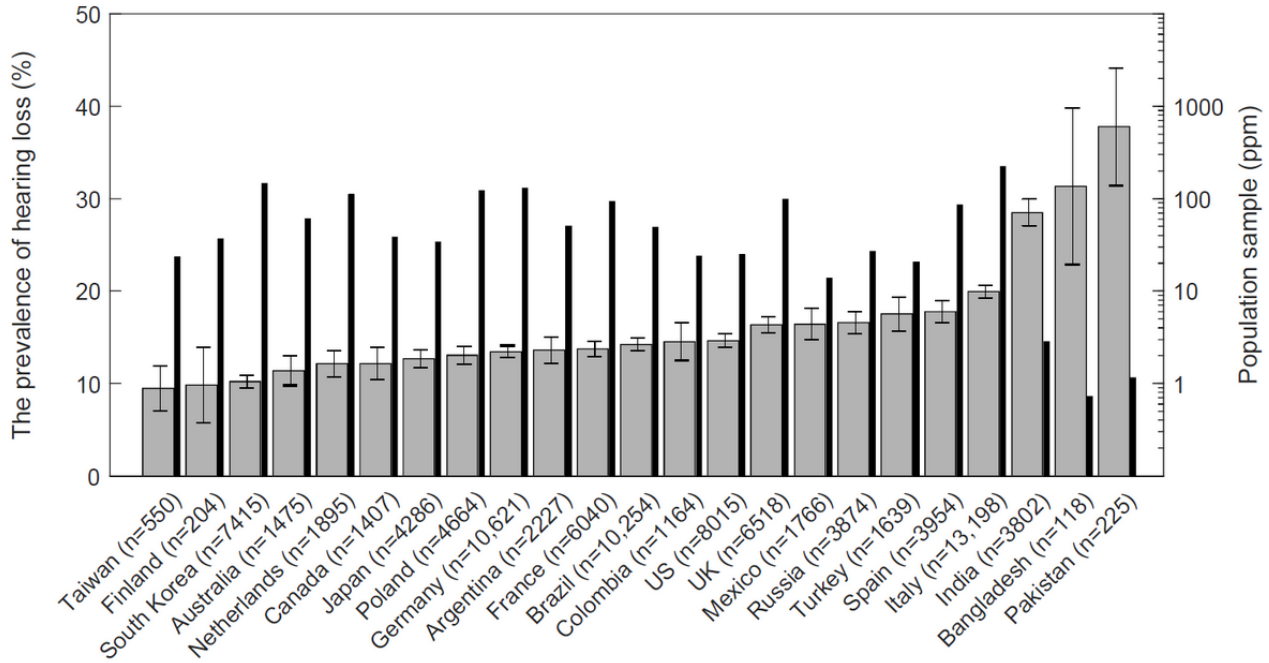
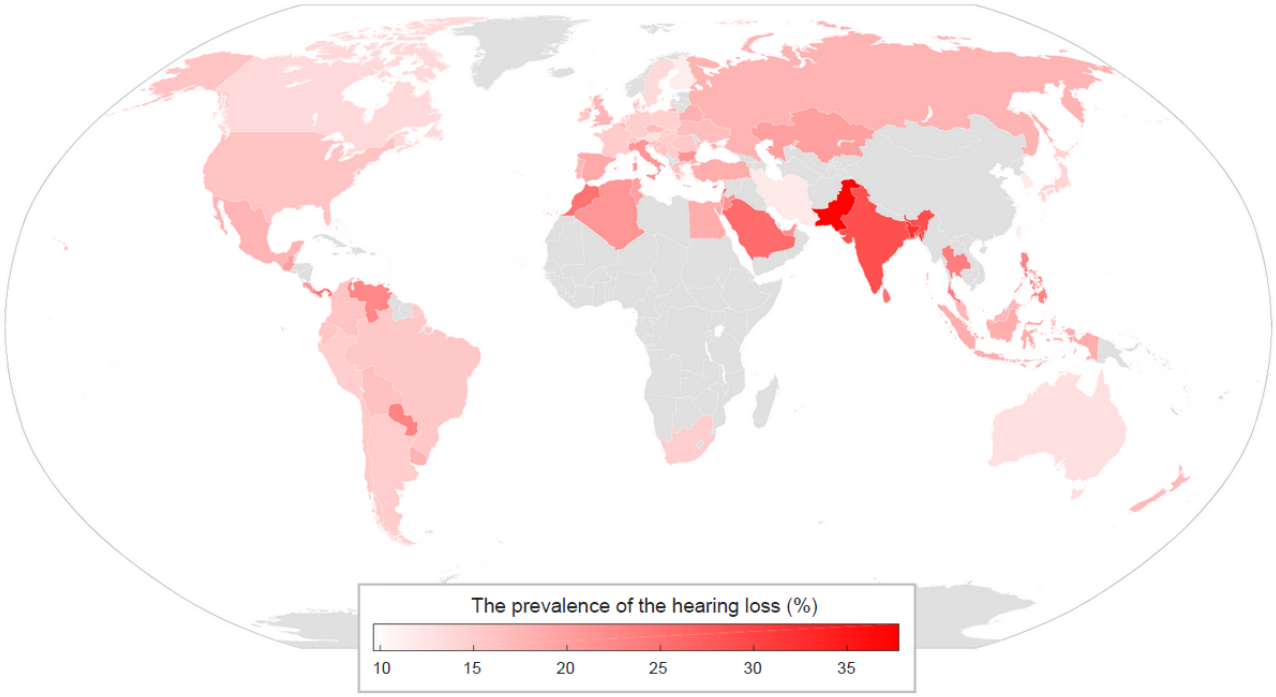


Figure 5. Global prevalence map of hearing loss.



Effect of Ambient Noise

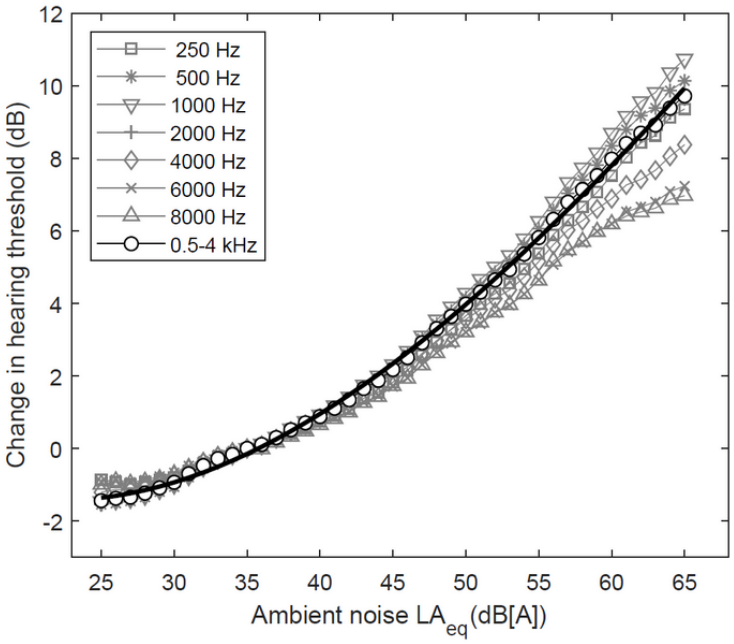
The effect of ambient noise was estimated based on 30,119 of 116,733 (25.8%) tests conducted with ambient noise monitoring. The difference in the hearing threshold in relation to the ambient noise level is presented in Figure 6. Hearing thresholds were adjusted for ambient noise by assuming the reference level LAeq=35 dB(A) [32]. The average decrease in the hearing threshold was obtained at a level of 2.53 dB (SD 0.74 dB). Additionally, the distributions of the hearing threshold in the

XSL•FO
RenderX

groups with and without ambient noise monitoring turned out to be significantly different at the level of $P<.001$, with the monitoring group having a hearing threshold that was lower by 1.25 dB (SD 0.12 dB). This produces an overall effect of ambient noise of 3.78 dB (SD 0.75 dB), which corresponds to a 4.16% (SD 1.46%) reduction in the country-specific prevalence.

The highest ambient noise levels were found in Pakistan, Bangladesh, and India, which corresponded to 10.7%, 8.6%, and 6.2% decreases, respectively, in the prevalences of hearing loss. Nevertheless, these countries still had the highest prevalences of hearing loss.

Figure 6. The difference in hearing threshold in relation to ambient noise level.



Effect of the Device Model

Differences in hearing thresholds between device models were found to be significant at the level of $P=.01$. These differences result directly from the biological calibration method that is based on determination of the reference sound level in relation to subjects with normal hearing. The standard error of the biological calibration estimated at 4.2 dB for 16 independent measurements [25] justifies the discrepancies. After adjusting for the device model, the mean hearing threshold by country changed by a mere −0.07 dB (SD 0.96 dB).

Discussion

Principal Findings

This paper presents the prevalence of hearing loss worldwide based on the 116,733 hearing tests conducted by Android users on mobile devices. The global prevalence of hearing loss was 15.6% (95% CI 15.4-15.8). Statistically significant differences were found across countries ($P<.001$). Of the 212 countries, 74 countries exceeded the number of tests required for the assumed sample size. Amongst these countries, the highest prevalences were found in Pakistan, Bangladesh, and India at 37.8% (95% CI 31.4-44.2), 32.2% (95% CI 23.6-40.8), and 28.5% (95% CI 27.1-29.9), respectively, whereas the lowest prevalences were found in Taiwan, Finland, and South Korea at 9.6% (95% CI 7.2-12.1), 9.8% (95% CI 5.7-13.9), and 10.2% (95% CI 9.5-10.9), respectively. The risk of hearing loss in Pakistan was 74.6% (95% CI 38.2-110.7) higher than in Taiwan. The absolute difference was 28.2% (95% CI 14.4-41.8).

Of the 733,716 tests, 116,733 (15.9%) were selected for the analysis. The number of tests rejected due to incompleteness, short duration, lack of headphones, or repetitions on a single device seems reasonable, especially as prior to performing a proper test, users often become familiar with the app by conducting a trial test.

In addition to Finland, relatively low hearing loss values were obtained for the other two Scandinavian countries, Sweden and Norway, at 12.1% (95% CI 8.4-15.9) and 12.7% (95% CI 8.1-17.4), respectively. The prevalences of hearing loss for the other two Asian Tigers (ie, Singapore and Hong Kong) were slightly higher, at 14.6% (95% CI 10.0-19.2) and 15.3% (95% CI 10.3-20.4), respectively. The high prevalence of hearing loss in India, which, along with Pakistan and Bangladesh, is one of the most populated countries in the world, is supported by the findings of Garg et al [33]. Similar hearing loss prevalences were found in most Western and Central European countries, which were slightly higher than Australia (11.4%, 95% CI 9.8-13.0) and Canada (12.2%, 95% CI 10.4-13.9) but lower than that in the United States (14.6%, 95% CI 13.9-15.4). Similar prevalences of hearing loss were obtained among the countries of the former Soviet Union (ie, Russia, Belarus, Ukraine, and Kazakhstan), at levels between 15.4% (95% 12.9-17.9) for Ukraine and 18.9% (95% CI 14.3-23.6) for Kazakhstan.

XSL•FO
RenderX

In Italy, 20.0% (95% CI 19.3-20.7) of the subjects had hearing loss, and this value was definitely higher than in the other Western and Central European countries. The credibility of this result is questionable due to the much higher use of apps in this country, being the highest in the world and reaching 222.3 people per million population (Multimedia Appendix 2). Attention should also be paid to the low Iranian result of 10.3% (95% CI 7.8-12.7), which significantly differs from other countries of North Africa and the Middle East. On the other hand, according to a study by the World Health Organization [5] in the region of North Africa and the Middle East, the prevalence of hearing loss is lower than in developed countries.

The required sample size was not reached in a significant number of African countries. In the poorest countries, the limiting factor is access to mobile devices and the internet. The same situation may also occur in countries with significant social stratification, contributing to the lower prevalences of hearing loss resulting from a lack of data from the poorer part of the population. The lack of data from China is related to the lack of active Google Play service in this country.

The prevalence of hearing loss in countries has been correlated with the infant mortality rate [34], which is often used as an indicator of the health status in a country. A Cronbach alpha coefficient of 0.76 (95% CI 0.52-0.90) has been obtained.

Comparison With Prior Work

Age-Related Hearing Loss

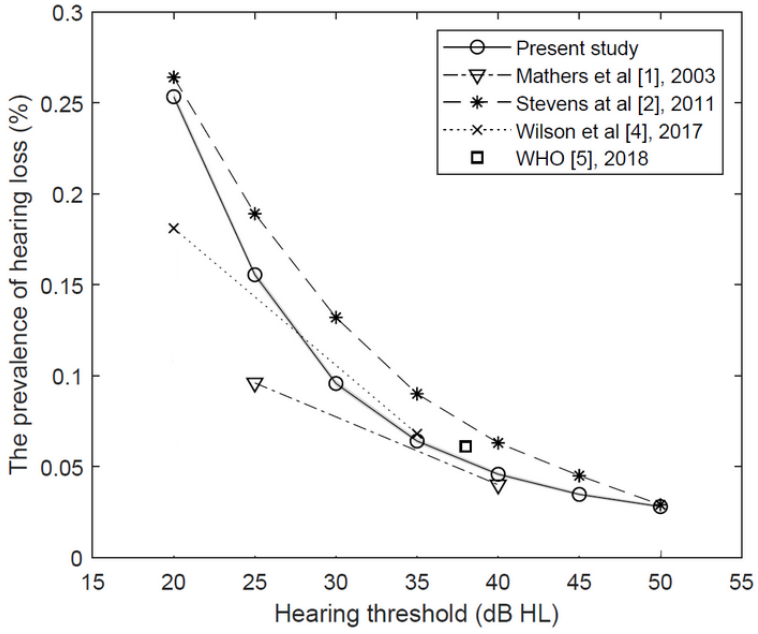
Age related hearing loss assessed by means of the mobile app was compared with the results from other studies [8-10,31,35-39] (Figure 3). The hearing threshold in subjects screened negatively for noise exposure [8,31,35-37], otologic disorders [8,35-37], ototoxic drugs [8,36], and asymmetric hearing [8,36,37] tends to be lower than that obtained in this study and prior unscreened studies [9,10,37-39]. However, substantial differences have been found between studies; although the threshold in the study by Kim et al [36] was higher than those in the studies by Engdahl et al [37] and Homans et al [39], it should have actually been lower because Kim et al conducted screening, while the other 2 studies did not [37,39]. In this study, for younger age groups (20-49 years), especially at low frequencies (250 Hz, 500 Hz), hearing thresholds were higher than in most previous studies. The results obtained at 250 Hz are consistent with those found by Kim et al [36] and Engdahl et al [37], while at 500 Hz, the results were consistent only with those found by Kim et al [36]. At frequencies above 4 kHz for older age groups (>70 years), the thresholds were lower than the values reported in the literature. Apart from the two youngest age groups (20-39 years) and the oldest age group (80-89 years), agreement with most studies was achieved for the average hearing threshold at the frequency range 0.5-4 kHz.

Increased hearing thresholds in younger age groups may be associated with a tendency for hearing-impaired subjects to take a test more willingly than people with normal hearing as well as with increased ambient noise, which mainly affects soft sounds. Lower values for the oldest groups might be related to an incorrect reaction to the masking noise. The effect of ambient noise and the masking noise is discussed in the Limitations section.

Hearing Loss in the World

The prevalence of hearing loss among users of the Hearing Test app may differ from population values, at least as regards the age structure. However, the value was compared with worldwide estimates. Figure 7 presents the worldwide prevalence of hearing loss with respect to the criterion based on the average hearing threshold in the better ear at frequencies 0.5-4 kHz. Despite statistically significant differences between the obtained results and other studies [1,2,4,5], the data are consistent.

Figure 7. Worldwide prevalence of hearing loss with respect to the criterion based on the average hearing threshold in the better ear at frequencies 0.5-4 kHz. Separate thresholds for adults (40 dB hearing level [HL]) and children (30 dB HL) in the study by the World Health Organization [5].



XSL•FO
RenderX

Country-Specific Studies

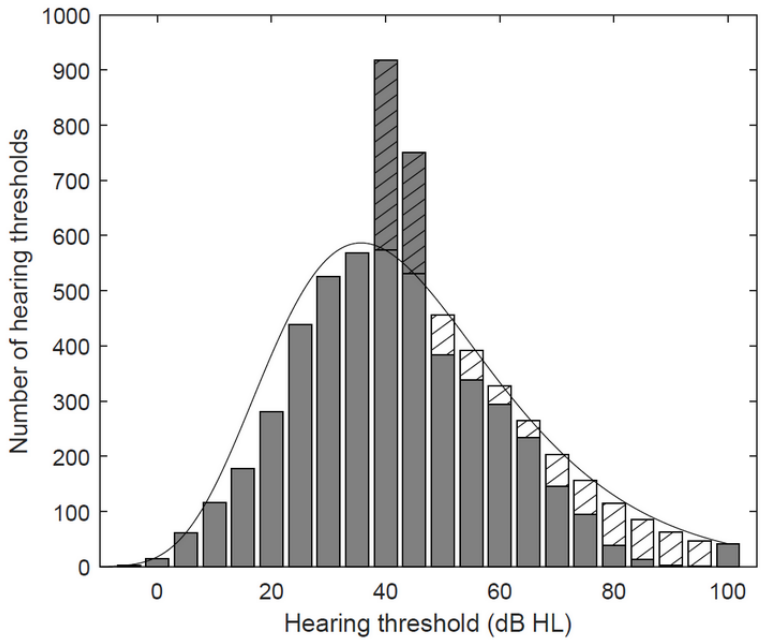
The prevalence of hearing loss was also compared with other studies. Calculations were performed by adopting the criteria for hearing loss and matching age ranges (Multimedia Appendix 3). The prevalence of hearing loss was lower in previous studies [11,16,40-45], pursuant to at least one criterion in research [9,10,46,47], and was higher in other articles [39,48-52]. There is large variation within the data presented in the literature, which may result not only from age structure differences and heterogeneous hearing loss definitions but also from research methodology, especially trial-specific recruitment methods or exclusion criteria. Therefore, drawing conclusions regarding country differences in the prevalence of hearing loss by means of literature data is challenging, as it requires the use of estimation methods in the presence of sparse data [2,3,13]. In this aspect, for country comparison, mobile data seem more valuable than a meta-analysis, although other confounding factors such as cultural background, propensity to new technologies, economic status, or app promotion should be underlined here as well.

Limitations

Masking Noise

Hearing thresholds for older groups at a frequency ≥4 kHz are lower than those reported in the literature (Figure 3). A histogram for these groups reveals an increased number of hearing thresholds at intensities of 40 dB HL and 45 dB HL (Figure 8). This bias could be related to the masking noise. Contralateral masking noise was switched on when the signal level exceeded 40 dB HL. Thus, some of the subjects could react incorrectly to the masking noise instead of the test signal. The higher the mean threshold in a group at a given frequency, the more subjects might react incorrectly to the masking noise, thus leading to an understatement of the threshold. The lapses are rare for a low mean, and the bias is not detectable until the mean threshold reaches 40 dB. Therefore, the bias is observable in a limited range at 4 kHz (older groups only) and is unnoticeable at lower frequencies. Consequently, its impact on the average hearing threshold at 0.5-4 kHz is negligible, and the prevalence of hearing loss largely coincides with data reported in the literature. Nonetheless, in the future, changes should be introduced to the app that will prevent improper reaction to the masking noise.

Figure 8. Hearing thresholds for subjects ≥60 years old at frequencies ≥4 kHz. The striped areas indicate the bias related to the masking noise that was enabled for stimuli >40 dB hearing level (HL). A lognormal distribution was assumed.



Ambient Noise

Ambient noise was recorded using a built-in microphone with a working range of 72-102 dB SPL, as guaranteed by the Android specification [29]. Since the measurements were carried out outside this range, they should be treated as approximate. Nevertheless, the results obtained are comparable with data reported in the literature. Elevation of the hearing threshold as a result of an increase in ambient noise from 35 dB LAeq to 60dB LAeq was estimated at 8.0 dB (Figure 6), which agrees with the elevation of about 7.5 dB presented by Na et al [53].

A relatively low elevation of the hearing threshold related to a considerable increase in ambient noise may be associated with the measurement method applied in this study and in the study by Na et al [53]. Instead of being ready for the test signal during the whole test, the subject decides for herself or himself about the duration of the test signal presentation and can extend it during an event of louder noise. Subjects who spent sufficient time performing the test were able to assess the quietest signals during a temporary decrease in ambient noise.

The effect of ambient noise on hearing test results was determined assuming that the increase in hearing threshold caused by ambient noise is conditioned by the masking of the

XSL•FO
RenderX

test signal by the noise. However, for significant hearing losses, ambient noise values above the normative level need not mask the loud test signal. Therefore, it must be assumed that the effect is present only for soft test signals and the estimated level for the effect of 3.78 dB (SD 0.75 dB) found in the present study should actually be treated as its upper limit. For these reasons, the results broadly correspond to pure-tone audiometry (Figure 3) despite ambient noise often exceeding the normative value for the audiometry.

Calibration

Mobile devices were calibrated using the biological method (ie, in relation to subjects with normal hearing). This method is less accurate than laboratory calibration. However, it enables semi-automatic calibration, which is crucial for as many as 1336 models. Because the tests were conducted on diverse device models in each country, the error associated with calibration has limited impact on differences between country-specific hearing loss prevalences.

Conclusions

This paper presents global country-specific prevalences of hearing loss based on self-tests carried out by Android users. Unsupervised self-tests require additional quality control based on the test duration and connection status of the headphones. Moreover, ambient noise and calibration method introduce additional bias. Despite this, hearing thresholds measured by means of mobile devices were congruent with data reported in the literature, whereas uniform recruitment criteria facilitate the comparison between countries. Hearing tests on mobile devices may be a valid tool in epidemiological studies carried out on a large scale.

Conflicts of Interest

MM is the author and owner of the Hearing Test application, by means of which the measurements described in this paper were conducted.

Multimedia Appendix 1
Median hearing threshold by age groups and by frequencies.
[PNG File , 31 KB - jmir_v22i7e17238_app1.png]

Multimedia Appendix 2
The prevalence of the hearing loss by countries.
[PNG File , 105 KB - jmir_v22i7e17238_app2.png]

Multimedia Appendix 3
Prevalence of the hearing loss in comparison with other studies.
[PNG File , 280 KB - jmir_v22i7e17238_app3.png]

References

1. Mathers C, Smith A, Concha M. Global burden of hearing loss in the year 2000. Global burden of Disease 2000. Geneva, Switzerland: World Health Organization; 2003. URL: https://www.who.int/healthinfo/statistics/bod_hearingloss.pdf [accessed 2020-06-24]
2. Stevens G, Flaxman S, Brunskill E, Mascarenhas M, Mathers CD, Finucane M, Global Burden of Disease Hearing Loss Expert Group. Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. Eur J Public Health 2013 Feb;23(1):146-152 [FREE Full text] [doi: 10.1093/eurpub/ckr176] [Medline: 22197756]
3. GBD 2015 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
4. Wilson BS, Tucci DL, Merson MH, O'Donoghue GM. Global hearing health care: new findings and perspectives. The Lancet 2017 Dec 02;390(10111):2503-2515. [doi: 10.1016/S0140-6736(17)31073-5] [Medline: 28705460]
5. Addressing the rising prevalence of hearing loss. Geneva, Switzerland: World Health Organization; 2018:1-28.
6. Arlinger S. Negative consequences of uncorrected hearing loss--a review. Int J Audiol 2003 Jul;42 Suppl 2:2S17-2S20. [Medline: 12918624]
7. Heine C, Browning CJ. Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions. Disabil Rehabil 2002 Oct 15;24(15):763-773. [doi: 10.1080/09638280210129162] [Medline: 12437862]
8. Pearson JD, Morrell CH, Gordon-Salant S, Brant LJ, Metter EJ, Klein LL, et al. Gender differences in a longitudinal study of age-associated hearing loss. J Acoust Soc Am 1995 Feb;97(2):1196-1205. [doi: 10.1121/1.412231] [Medline: 7876442]
9. Mościcki EK, Elkins EF, Baum HM, McNamara PM. Hearing loss in the elderly: an epidemiologic study of the Framingham Heart Study Cohort. Ear Hear 1985;6(4):184-190. [Medline: 4043571]

XSL•FO
RenderX

10. Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, et al. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. Am J Epidemiol 1998 Nov 01;148(9):879-886. [doi: 10.1093/oxfordjournals.aje.a009713] [Medline: 9801018]
11. Nash SD, Cruickshanks KJ, Klein R, Klein BEK, Nieto FJ, Huang GH, et al. The prevalence of hearing impairment and associated risk factors: the Beaver Dam Offspring Study. Arch Otolaryngol Head Neck Surg 2011 May;137(5):432-439 [FREE Full text] [doi: 10.1001/archoto.2011.15] [Medline: 21339392]
12. Olusanya BO, Neumann KJ, Saunders JE. The global burden of disabling hearing impairment: a call to action. Bull World Health Organ 2014 May 01;92(5):367-373 [FREE Full text] [doi: 10.2471/BLT.13.128728] [Medline: 24839326]
13. Roth TN, Hanebuth D, Probst R. Prevalence of age-related hearing loss in Europe: a review. Eur Arch Otorhinolaryngol 2011 Aug;268(8):1101-1107 [FREE Full text] [doi: 10.1007/s00405-011-1597-8] [Medline: 21499871]
14. Foulad A, Bui P, Djalilian H. Automated audiometry using apple iOS-based application technology. Otolaryngol Head Neck Surg 2013 Nov;149(5):700-706. [doi: 10.1177/0194599813501461] [Medline: 23963611]
15. Kam ACS, Sung JKK, Lee T, Wong TKC, van Hasselt A. Clinical evaluation of a computerized self-administered hearing test. Int J Audiol 2012 Aug;51(8):606-610. [doi: 10.3109/14992027.2012.688144] [Medline: 22676699]
16. Yousuf Hussein S, Wet Swanepoel D, Biagio de Jager L, Myburgh HC, Eikelboom RH, Hugo J. Smartphone hearing screening in mHealth assisted community-based primary care. J Telemed Telecare 2016 Oct;22(7):405-412. [doi: 10.1177/1357633X15610721] [Medline: 26468215]
17. Renda L, Selçuk Ö, Eyigör H, Osma Ü, Yılmaz MD. Smartphone Based Audiometric Test for Confirming the Level of Hearing; Is It Useable in Underserved Areas? J Int Adv Otol 2016 Apr;12(1):61-66 [FREE Full text] [doi: 10.5152/iao.2016.1421] [Medline: 27340985]
18. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. Int J Audiol 2014 Dec;53(12):841-849. [doi: 10.3109/14992027.2014.920965] [Medline: 24998412]
19. Sandström J, Swanepoel DW, Carel Myburgh H, Laurent C. Smartphone threshold audiometry in underserved primary health-care contexts. Int J Audiol 2016;55(4):232-238. [doi: 10.3109/14992027.2015.1124294] [Medline: 26795898]
20. Chu Y, Cheng Y, Lai Y, Tsao Y, Tu T, Young ST, et al. A Mobile Phone-Based Approach for Hearing Screening of School-Age Children: Cross-Sectional Validation Study. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e12033 [FREE Full text] [doi: 10.2196/12033] [Medline: 30932870]
21. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Patient Uptake, Experience, and Satisfaction Using Web-Based and Face-to-Face Hearing Health Services: Process Evaluation Study. J Med Internet Res 2020 Mar 20;22(3):e15875 [FREE Full text] [doi: 10.2196/15875] [Medline: 32196459]
22. Sandström J, Swanepoel D, Laurent C, Umefjord G, Lundberg T. Accuracy and Reliability of Smartphone Self-Test Audiometry in Community Clinics in Low Income Settings: A Comparative Study. Ann Otol Rhinol Laryngol 2020 Jun;129(6):578-584. [doi: 10.1177/0003489420902162] [Medline: 31965808]
23. Bright T, Pallawela D. Validated Smartphone-Based Apps for Ear and Hearing Assessments: A Review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e13 [FREE Full text] [doi: 10.2196/rehab.6074] [Medline: 28582261]
24. Masalski M, Grysiński T, Kręcicki T. Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry. JMIR Mhealth Uhealth 2018 Jan 10;6(1):e10 [FREE Full text] [doi: 10.2196/mhealth.7800] [Medline: 29321124]
25. Masalski M, Kipiński L, Grysiński T, Kręcicki T. Hearing Tests on Mobile Devices: Evaluation of the Reference Sound Level by Means of Biological Calibration. J Med Internet Res 2016 May 30;18(5):e130 [FREE Full text] [doi: 10.2196/jmir.4987] [Medline: 27241793]
26. Masalski M. Hearing Test App.: Google Play URL: https://play.google.com/store/apps/details?id=mobile.eaudiologia&hl=en [accessed 2020-05-16]
27. Masalski M, Grysiński T, Kręcicki T. Biological calibration for web-based hearing tests: evaluation of the methods. J Med Internet Res 2014 Jan 15;16(1):e11 [FREE Full text] [doi: 10.2196/jmir.2798] [Medline: 24429353]
28. Masalski M, Kręcicki T. Self-test web-based pure-tone audiometry: validity evaluation and measurement error analysis. J Med Internet Res 2013 Apr 12;15(4):e71 [FREE Full text] [doi: 10.2196/jmir.2222] [Medline: 23583917]
29. Audio: Implementation: Configuring Preprocessing Effects. Android Interfaces and Architecture.: Google LLC and the Open Handset Alliance URL: https://source.android.com/devices/audio/implement-pre-processing [accessed 2020-06-24] [WebCite Cache ID 6rpnPjpR5]
30. geoPlugin. France: GEOPLUGIN, SAS URL: https://www.geoplugin.com/ [accessed 2020-06-24]
31. Johansson MSK, Arlinger SD. Hearing threshold levels for an otologically unscreened, non-occupationally noise-exposed population in Sweden. Int J Audiol 2002 Apr;41(3):180-194. [doi: 10.3109/14992020209077182] [Medline: 12033636]
32. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. Reading, UK: British Society of Audiology; 2011. URL: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf [WebCite Cache ID 67yJElAP1]

XSL•FO
RenderX

33. Garg S, Kohli C, Mangla V, Chadha S, Singh MM, Dahiya N. An Epidemiological Study on Burden of Hearing Loss and Its Associated Factors in Delhi, India. Ann Otol Rhinol Laryngol 2018 Sep;127(9):614-619. [doi: 10.1177/0003489418781968] [Medline: 29938517]
34. Infant Mortality Rate. The World Factbook. Washington, DC: Central Intelligence Agency; 2017. URL: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2091rank.html [accessed 2020-06-24]
35. ISO 7029:2000 Acoustics — Statistical distribution of hearing thresholds as a function of age. Geneva, Switzerland: International Organization for Standardization; 2000. URL: https://www.iso.org/standard/26314.html [accessed 2000-06-24]
36. Kim S, Lim EJ, Kim HS, Park JH, Jarng SS, Lee SH. Sex Differences in a Cross Sectional Study of Age-related Hearing Loss in Korean. Clin Exp Otorhinolaryngol 2010 Mar;3(1):27-31 [FREE Full text] [doi: 10.3342/ceo.2010.3.1.27] [Medline: 20379399]
37. Engdahl B, Tambs K, Borchgrevink HM, Hoffman HJ. Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag Hearing Loss Study. Int J Audiol 2005 Apr;44(4):213-230. [doi: 10.1080/14992020500057731] [Medline: 16011050]
38. Strauss S, Swanepoel DW, Becker P, Eloff Z, Hall JW. Noise and age-related hearing loss: a study of 40 123 gold miners in South Africa. Int J Audiol 2014 Mar;53 Suppl 2:S66-S75. [doi: 10.3109/14992027.2013.865846] [Medline: 24564695]
39. Homans NC, Metselaar RM, Dingemanse JG, van der Schroeff MP, Brocaar MP, Wieringa MH, et al. Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. Laryngoscope 2017 Mar;127(3):725-730. [doi: 10.1002/lary.26150] [Medline: 27377351]
40. Lohi V, Ohtonen P, Aikio P, Sorri M, Mäki-Torkko E, Hannula S. Hearing impairment is common among Saami adults in Northern Finland. Int J Circumpolar Health 2017;76(1):1398004 [FREE Full text] [doi: 10.1080/22423982.2017.1398004] [Medline: 29132251]
41. Jun HJ, Hwang SY, Lee SH, Lee JE, Song J, Chae S. The prevalence of hearing loss in South Korea: data from a population-based study. Laryngoscope 2015 Mar;125(3):690-694. [doi: 10.1002/lary.24913] [Medline: 25216153]
42. Mattos LC, Veras RP. The prevalence of hearing loss in an elderly population in Rio de Janeiro: a cross-sectional study. Brazilian Journal of Otorhinolaryngology 2007 Sep;73(5):654-659 [FREE Full text] [doi: 10.1016/s1808-8694(15)30126-9] [Medline: 18094807]
43. Agrawal Y, Platz EA, Niparko JK. Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. Arch Intern Med 2008 Jul 28;168(14):1522-1530. [doi: 10.1001/archinte.168.14.1522] [Medline: 18663164]
44. Helzner EP, Cauley JA, Pratt SR, Wisniewski SR, Zmuda JM, Talbott EO, et al. Race and sex differences in age-related hearing loss: the Health, Aging and Body Composition Study. J Am Geriatr Soc 2005 Dec;53(12):2119-2127. [doi: 10.1111/j.1532-5415.2005.00525.x] [Medline: 16398896]
45. Louw C, Swanepoel DW, Eikelboom RH, Hugo J. Prevalence of hearing loss at primary health care clinics in South Africa. Afr Health Sci 2018 Jun;18(2):313-320 [FREE Full text] [doi: 10.4314/ahs.v18i2.16] [Medline: 30602958]
46. Lin FR, Thorpe R, Gordon-Salant S, Ferrucci L. Hearing loss prevalence and risk factors among older adults in the United States. J Gerontol A Biol Sci Med Sci 2011 May;66(5):582-590 [FREE Full text] [doi: 10.1093/gerona/glr002] [Medline: 21357188]
47. Sindhusake D, Mitchell P, Smith W, Golding M, Newall P, Hartley D, et al. Validation of self-reported hearing loss. The Blue Mountains Hearing Study. Int J Epidemiol 2001 Dec;30(6):1371-1378. [doi: 10.1093/ije/30.6.1371] [Medline: 11821349]
48. Lin C, Yang Y, Guo YL, Wu C, Chang C, Wu J. Prevalence of hearing impairment in an adult population in Southern Taiwan. Int J Audiol 2007 Dec;46(12):732-737. [doi: 10.1080/14992020701448986] [Medline: 18049962]
49. Lin FR, Niparko JK, Ferrucci L. Hearing loss prevalence in the United States. Arch Intern Med 2011 Nov 14;171(20):1851-1852 [FREE Full text] [doi: 10.1001/archinternmed.2011.506] [Medline: 22083573]
50. Fisher DE, Li C, Hoffman HJ, Chiu MS, Themann CL, Petersen H, et al. Sex-specific predictors of hearing-aid use in older persons: The age, gene/environment susceptibility - Reykjavik study. Int J Audiol 2015;54(9):634-641 [FREE Full text] [doi: 10.3109/14992027.2015.1024889] [Medline: 25816699]
51. Brennan-Jones CG, Taljaard DS, Brennan-Jones SEF, Bennett RJ, Swanepoel DW, Eikelboom RH. Self-reported hearing loss and manual audiometry: A rural versus urban comparison. Aust J Rural Health 2016 Apr;24(2):130-135. [doi: 10.1111/ajr.12227] [Medline: 26311193]
52. Zhan W, Cruickshanks KJ, Klein BEK, Klein R, Huang G, Pankow JS, et al. Generational differences in the prevalence of hearing impairment in older adults. Am J Epidemiol 2010 Jan 15;171(2):260-266 [FREE Full text] [doi: 10.1093/aje/kwp370] [Medline: 20008889]
53. Na Y, Joo HS, Yang H, Kang S, Hong SH, Woo J. Smartphone-based hearing screening in noisy environments. Sensors (Basel) 2014 Jun 12;14(6):10346-10360 [FREE Full text] [doi: 10.3390/s140610346] [Medline: 24926692]

Abbreviations

HL: hearing level.

XSL•FO
RenderX

Edited by G Eysenbach; submitted 02.12.19; peer-reviewed by Y Shu, Y Chu; comments to author 10.03.20; revised version received 28.04.20; accepted 15.05.20; published 23.07.20.

Please cite as:
Masalski M, Morawski K
Worldwide Prevalence of Hearing Loss Among Smartphone Users: Cross-Sectional Study Using a Mobile-Based App
J Med Internet Res 2020;22(7):e17238
URL: http://www.jmir.org/2020/7/e17238/
doi:10.2196/17238
PMID:32706700

©Marcin Masalski, Krzysztof Morawski. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Divide in Perceived Benefits of Online Health Care and Social Welfare Services: National Cross-Sectional Survey Study

Tarja Heponiemi[1], PhD; Vesa Jormanainen[1], MD; Lars Leemann[1], PhD; Kristiina Manderbacka[1], PhD; Anna-Mari Aalto[1], PhD; Hannele Hyppönen[1], PhD

Finnish Institution for Health and Welfare, Helsinki, Finland

Corresponding Author:
Tarja Heponiemi, PhD
Finnish Institution for Health and Welfare
PO Box 30
Helsinki,
Finland
Phone: 358 295247434
Email: tarja.heponiemi@thl.fi

Abstract

Background: The number of online services in health care is increasing rapidly in developed countries. Users are expected to take a more skilled and active role in taking care of their health and prevention of ill health. This induces risks that users (especially those who need the services the most) will drop out of digital services, resulting in a digital divide or exclusion. To ensure wide and equal use of online services, all users must experience them as beneficial.

Objective: This study aimed to examine associations of (1) demographics (age, gender, and degree of urbanization), (2) self-rated health, (3) socioeconomic position (education, experienced financial hardship, labor market position, and living alone), (4) social participation (voting, satisfaction with relationships, and keeping in touch with friends and family members), and (5) access, skills, and extent of use of information and communication technologies (ICT) with perceived benefits of online health care and social welfare services. Associations were examined separately for perceived health, economic, and collaboration benefits.

Methods: We used a large random sample representative of the Finnish population including 4495 (56.77% women) respondents aged between 20 and 97 years. Analyses of covariance were used to examine the associations of independent variables with perceived benefits.

Results: Access to online services, ICT skills, and extent of use were associated with all examined benefits of online services. ICT skills seemed to be the most important factor. Poor self-rated health was also consistently associated with lower levels of perceived benefits. Similarly, those who were keeping in touch with their friends and relatives at least once a week perceived online services more often beneficial in all the examined dimensions. Those who had experienced financial hardship perceived fewer health and economic benefits than others. Those who were satisfied with their relationships reported higher levels of health and collaboration benefits compared with their counterparts. Also age, education, and degree of urbanization had some statistically significant associations with benefits but they seemed to be at least partly explained by differences in access, skills, and extent of use of online services.

Conclusions: According to our results, providing health care services online has the potential to reinforce existing social and health inequalities. Our findings suggest that access to online services, skills to use them, and extent of use play crucial roles in perceiving them as beneficial. Moreover, there is a risk of digital exclusion among those who are socioeconomically disadvantaged, in poor health, or socially isolated. In times when health and social services are increasingly offered online, this digital divide may predispose people with high needs for services to exclusion from them.

(J Med Internet Res 2020;22(7):e17616) doi:10.2196/17616

KEYWORDS

digital divide; online services; benefits; skills; access; sociodemographics; participation; Finland

XSL•FO
RenderX

Introduction

Background

The number of online services in health care is increasing rapidly in Finland and other developed countries. The new online services are implemented and adopted to support self-management and self-service, and thus users are expected to take a more skilled and active role in maintaining their well-being, health, and especially long-term illnesses. To ensure that users increase participation in online services, it is important that they find that online services are beneficial for them [1].

There are risks that users (especially those people who need the services most) will drop out of online services resulting in a digital divide or exclusion. The concept of digital divide has traditionally referred to the gap between persons who have and do not have access to new forms of information technology [2]. Limits to access can have motivational, physical, skills, or use grounds [3]. Nonuse of online digital services can have psychological (motivation, confidence, and experienced attractiveness of use), physical (opportunity to use), or practical (lack of skills) causes [4]. As the internet has diversified and diffused widely over the past 10 years, it is assumed that internet nonuse is, at present, far more common among vulnerable groups and that lack of access has become a less important reason for internet nonuse compared with lack of skills [5].

The digital divide or exclusion affects the poorest and most marginalized groups from the outset [6]. The internet reinforces existing social differences, and internet nonusers are reported to be older and less educated and more likely to be unemployed, disabled, and socially isolated [7]. Sociodemographic, economic, and geographical factors have been found to be key determinants of digital exclusion [8]. Among urban adults, the discriminating factors between computer users and nonusers include age, education, employment, and income, self-rated health status, hospital visits, organization membership, and participation in voluntary work [9].

A previous study from the Netherlands showed that the elderly benefit less from internet use than the young [5]. Moreover, those less educated and poorer gained fewer economic benefits from the internet than their counterparts, and single people gained fewer social benefits than those living with others [5]. A study from the United Kingdom found that more educated people benefit most from internet use [10]; surprisingly, the elderly benefited more than the young. In Germany, it was found that enabling and caring support through the internet was associated with job search benefits among the unemployed [11].

Social and digital exclusion are interlinked: differential access to technology contributes to socioeconomic stratification or inclusion and vice versa [12]. The debate has moved from a single digital divide to multiple digital divides that include not only global challenges but also local contextual problems in terms of availability of content, bandwidth, and skills, among other issues [13]. Helsper's [14] corresponding fields model suggests that digital and social exclusion influence one another and the resources a person has offline influence their ability to use digital online solutions. Consequently, client and patient resources influence adoption and sustained use of new digital online health services. Therefore, socioeconomic factors, social participation, health status, and well-being have potential effects on digital divide or exclusion. Only by studying these together can we identify the separate and combined influences of different types of social exclusion on different types of digital exclusion. The model characterizes access and skills as important social impact mediators in this context. Without access, the internet or other information and communication technologies (ICT) cannot be used, which makes access the most basic mediator between offline and digital fields of exclusion. Similarly, certain skills on a basic technical and operational level are required for the use of ICT and the internet [14].

Aceto et al [15] have suggested that ICT in health care decreases costs, enables discovery, and improves patient outcomes. Moreover, ICT is proposed to extensively transform health care to being more proactive, preventive, and person-centered instead of being reactive and hospital-centered [15]. To fully take advantage of all the suggested benefits of ICT, patients must use online services and find them beneficial. We need to understand why people choose not to take advantage of the opportunities ICT offers for managing their health and define the characteristics of people who are most likely to benefit and those at risk of exclusion.

Research focus should shift from access to digital technology (first-level divide) to skills and use of technology (second-level divide) and who benefits the most or least from internet use (third-level divide) [16,17]. Third-level digital divide refers to gaps in capacity to translate individual internet access and use into favorable offline outcomes [5]. Research related to the third-level divide and its social determinants is scarce [17], and the role of access, skills, and extent of use in this context must be identified. Previous studies [4,7,8] have focused on general use of the internet, but information about determinants of using the internet for health care and social welfare services is lacking and must be identified because these vital services are increasingly moving online.

Objectives

This study aimed to examine direct associations of demographics (age, gender, degree of urbanization), self-rated health, socioeconomic position (education, experienced financial hardship, labor market position, and living alone) and social participation (voting, satisfaction with relationships, and keeping in touch with friends and family members) with perceived benefits of online health care and social welfare services (health, economic, and collaboration benefits). The study data are based on a large random sample representative of the Finnish population. Furthermore, we examined the effects of access, skills, and the extent of ICT use and whether these could account for those associations.

Methods

Sampling

The data collection and questionnaire formulation have been reported in more detail elsewhere [18]. A random sample of 10,000 people representative of the adult population (aged 20

XSL•FO
RenderX

years and older) living in Finland was collected from the Population Register Center of Finland. For those aged 75 years and older, double picking probability was used in order to guarantee sufficient group size. The questionnaire (Multimedia Appendix 1) was sent by postal mail to all sampled persons in 2017 with instructions on how to respond online. Reminders were sent 3 times to those who had not responded. The response rate was 44.95% (4495/10,000), with 56.77% (2552/4495) being women. The respondents were older, more often women, and had more education than the eligible population [18], and we addressed this by applying inverse probability weighting based on age, gender, marital status, education level, living region, and degree of urbanization of the residential municipality. This method has been used previously in many Finnish population studies, and it has been found to perform well in correcting possible effects of nonresponse on representativeness of the results [19].

Setting

Finland is among the leading countries in digitalization of health care and social welfare services. The country has launched nationwide integrated data services for health care and social welfare services, the Kanta services, in several steps from May 2010 onward [20,21]. Kanta services are targeted to citizens, health care and social welfare service providers and professionals, and community pharmacies. Kanta services include electronic prescribing (Prescription Centre), patient-accessible electronic health records (My Kanta Pages), health records (Patient Data Repository), and the client data archive for social welfare services. It is mandatory for all health care providers, public or private, to subscribe to and use the national Kanta services. Implementation and adoption of Kanta services for social welfare started in May 2018 and is currently ongoing. In addition to national comprehensive undertakings, at least two other nationwide digital service projects (Health Village 2.0 and Digisote [former ODA]) have been developed, implementing a large number of new online health and welfare services. In particular, these services have focused on supporting clients to self-manage their health and well-being online. Moreover, service providers such as hospital districts and primary health care centers provide regional and local electronic online services such as booking services, electronic consultations, and web-based question-answer services, etc [22].

Measurements

Dependent Variables

Benefits of online health and social care services were measured with 3 scales on perceived health, economic, and collaboration benefits. Health benefits were measured with 6 items such as “online services help citizens self-manage their health/well-being.” Cronbach alpha for this scale was α=.90. Economic benefits included 4 items (α=.87) such as “online services provide useful reminders (eg, the time of reception, laboratory tests, renewing prescriptions etc).” Collaboration benefits included 3 items (α=.86) such as “online services support the collaboration and information flow between the patient/client and carer.” All items were rated on a 5-point Likert scale ranging from 1 (completely disagree) to 5 (completely agree), and the mean for each scale was calculated.

Independent Variables

Demographics included age in years, gender (male or female), and degree of urbanization of the residential municipality (urban, semiurban, or rural area). Health status was measured by asking how respondents would describe their state of health at present (with answer alternatives good, fairly good, average, fairly poor, poor). For analyses, it was coded 0 (above average health) or 1 (average health or poorer).

Indicators of socioeconomic position were education, classified into 3 categories (low, intermediate, or high) by setting the self-reported years of education in proportion to the tertiles of education years in each respondents’ age and gender group (using 10-year age groups); financial hardship, assessed by asking whether respondent had been forced to bargain for food, medication, or physician visits due to shortage of money during the previous 12 months (none or yes to at least one of these); current labor market position (currently working, old age pension, disability pension, unemployed, student, or other); and living alone (no or yes).

Indicators of social participation were voted in last municipal elections (no or yes), satisfaction with one’s social relationships (no or yes), and keeping in contact with friends and relatives outside the own household at least once a week (no or yes).

ICT-related variables included access to, skills to use, and extent of use related to online services in health and social care. Access to online services was measured by 7 items (α=.84) asking availability of online services, access to computer and internet, quality of data connections, possibility of getting online services in their own language, accessibility of services despite obstacles (eg, disabilities), ease of finding online services, and possibility of taking care of things online on behalf of someone else. Skills to use online services were measured with 5 items (α=.86) assessing technical skills to use online services, support received for using online services, possibility of receiving help when facing technical problems, experienced clearness and length of terms of use, and perceived ease of using online services. The response options for access and skills scales ranged between 1 (completely disagree) and 5 (completely agree), and the mean of the items was calculated for both scales. The extent of use of online services in health and social care was measured with 16 items asking whether the respondent had used described service/functionality during the past year (response options no, yes traditionally, or yes online). The functionalities/services were related to self-care (3 items assessing, for example, whether the respondent had used online symptom checkers or followed own personal health data online), use for collaboration (8 items such as whether respondent has viewed the patient data recorded by professionals about them, received laboratory results, renewed prescriptions, or asked advice from professionals), and seeking out services (5 items assessing, for example, whether the respondent had searched for information about available services, booked appointment, or compared quality of services). The sum of functionalities/services a respondent had used online was calculated, resulting in a range of 0 to 16 functionalities (extent of use).

XSL•FO
RenderX

Ethical Issues

Ethical approval for the study was received from the research ethics committee of the Finnish Institute for Health and Welfare. Data were collected with no direct identification information concerning the respondents, and therefore no individuals can be identified from the data.

Statistical Analysis

Associations of independent variables with outcome variables related to perceived benefits (health, economic, and collaboration) were examined using analysis of covariance (in separate analyses). Analyses were conducted in 3 steps. In the first step, analyses included demographics (age, gender, and degree of urbanization) and health status (model A). In the second step, variables related to socioeconomic position (education, experienced financial hardship, labor market position, living alone) and variables related to social participation (voting, satisfaction with relationships, and keeping in touch with friends and family members) were added to the model (model B). In model C, ICT-related variables concerning access, skills, and the extent of use were added. Skills and access were examined in separate analyses to avoid multicollinearity because they correlated with each other. Analyses were conducted in these steps to examine the independent effect of each of the demographic and socioeconomic variables on each of the perceived benefits and whether ICT-related variables would partly account for possible associations of the independent variables from models A and B with outcome variables. Analyses were conducted using SPSS Statistics version 25 (IBM Corp). Methods suitable for weighted data were used (ie, complex samples general linear model for analyses of variance and complex samples descriptives and frequencies for descriptive statistics).

Results

Characteristics of Respondents

The characteristics of the study population are presented in Table 1. The mean age of the respondents was 51.7 years, more than 70% (3201/4494, 71.23%) lived in urban regions, and about half (2160/4495, 48.08%) were employed. Those respondents having low education comprised the largest group. Approximately two-thirds (3054/4495, 67.94%) of the respondents had used at least one of the online health or social services and functionalities under study.

XSL•FO
RenderX

Table 1. The basic background characteristics of the weighted study sample (n=4495).

Characteristics	Value
Gender[a], n (%)	
Women	2300 (51.18)
Men	2194 (48.82)
Degree of urbanization, n (%), (n=4494)	
Urban	3201 (71.23)
Semiurban	658 (14.64)
Rural	635 (14.13)
Self-rated health, n (%)	
Above average	2884 (64.83)
Average or poorer	1564 (35.17)
Education, n (%), (n=4260)	
Low	1888 (44.32)
Intermediate	1279 (30.02)
High	1093 (25.66)
Financial hardship, n (%)	
No	3613 (81.7)
Yes	807 (18.3)
Labor market position, n (%), (n=4494)	
Working	2160 (55.34)
Old age pension	823 (21.09)
Disability pension	158 (4.05)
Unemployed	273 (7.00)
Student	257 (6.58)
Other	232 (5.94)
Living alone, n (%)	
No	3174 (72.19)
Yes	1223 (27.81)
Voted in the last municipal elections, n (%)	
No	1142 (26.20)
Yes	3217 (73.80)
Satisfaction with relationships, n (%)	
No	984 (22.68)
Yes	3355 (77.32)
Keeping in touch with friends and relatives, n (%)	
No	805 (18.09)
Yes	3645 (81.91)
Age[b], mean (SE)	51.7 (0.32)
ICT[c]-related factors	
Access[d], mean (SE)	3.90 (0.02)
Skills[e], mean (SE)	3.51 (0.02)
Extent of use[d], mean (SE)	2.62 (0.05)

XSL•FO
RenderX

[a]Because this table presents weighted results, the percentages of genders vary from those given for the actual respondents (eg, in the abstract and sampling section).

[b]Range: 20-97.

[c]ICT: information and communication technologies.

[d]Range: 1-5.

[e]Range: 0-16.

Perceived Health Benefits

Table 2 shows results regarding perceived health benefits. In model A, age and self-rated health were significantly associated with perceived health benefits. Older respondents and those with poor health perceived fewer health benefits from online services. The association of health to perceived benefits remained significant after all adjustments, whereas the association regarding age attenuated to nonsignificant after adjusting for ICT-related variables. In model B, education, financial hardship, satisfaction with relationships, and keeping in touch with friends and relatives were all significantly associated with health benefits. Compared with other respondents, perceived health benefits were higher among those who were satisfied with their relationships and were keeping in touch with their friends and relatives, whereas health benefits were lower among those who had experienced financial hardship. Those having high education (estimated mean [EM] 3.64 [SE 0.04]) or intermediate education (EM 3.62 [SE 0.04]) perceived more health benefits compared with those having low education (EM 3.48 [SE 0.04]). Other associations persisted after adding the ICT-related variables, whereas education did not remain significant after that. In model C, higher levels of access, skills, and extent of use were all associated with higher levels of perceived health benefits. R^2 of model C was .16.

Table 2. Association of explanatory factors with perceived health benefits (analysis of covariance, n=4495).

Variable	Model A		Model B		Model C	
	F score (*df*)	*P* value	*F* score (*df*)	*P* value	*F* score (*df*)	*P* value
Gender	0.84 (1)	.36	1.06 (1)	.30	2.61 (1)	.11
Age	37.93 (1)	<.001	11.48 (1)	.001	1.70 (1)	.19
Urbanization	1.09 (2)	.34	1.11 (2)	.33	1.79 (2)	.17
Self-rated health	53.33 (1)	<.001	16.24 (1)	<.001	1.94 (1)	.001
Education	—	—	1.70 (2)	<.001	2.74 (2)	.07
Financial hardship	—	—	5.97 (1)	.02	6.17 (1)	.01
Labor market position	—	—	1.02 (5)	.40	0.77 (5)	.57
Living alone	—	—	0.00 (1)	.97	0.37 (1)	.54
Voting	—	—	3.06 (1)	.08	8.15 (1)	.004
Relationships	—	—	8.52 (1)	.004	6.75 (1)	.009
Keeping in touch	—	—	17.91 (1)	<.001	9.27 (1)	.002
Access	—	—	—	—	51.43 (1)	<.001
Skills	—	—	—	—	109.76 (1)	<.001
Extent of use	—	—	—	—	43.02 (1)	<.001

Perceived Economic Benefits

Table 3 shows the results regarding perceived economic benefits. In model A, age, degree of urbanization, and health status were significantly associated with perceived economic benefits. Older respondents and those with poor health perceived fewer economic benefits compared with their counterparts. Those living in urban areas perceived more economic benefits (EM 4.1 [SE 0.02]) than those living in semiurban (EM 3.9 [SE 0.04]) or rural areas (EM 4.0 [SE 0.04]). The association of benefits with health remained significant after all adjustments, whereas the association with age attenuated to nonsignificant after model B adjustments and with urbanization after adding ICT-related variables. In model B, education, financial hardship, and keeping touch with friends and relatives were all significantly associated with perceived economic benefits. Those having high (EM 3.94 [SE 0.05]) or intermediate (EM 3.96 [SE 0.04]) education perceived more economic benefits compared with those having low education (EM 3.83 [SE 0.04]). Perceived economic benefits were higher among those who were keeping in touch with their friends and relatives and lower among those who had experienced financial hardship. The associations of economic benefits with financial hardship and keeping in touch with friends and relatives remained significant in all analyses, whereas the association with education attenuated to nonsignificant after adjusting for ICT-related variables. In model C, higher levels of access, skills, and extent of use were all associated with higher levels of perceived economic benefits. R^2 of model C was .14.

XSL•FO
RenderX

Table 3. The association of explanatory factors with perceived economic benefits (analysis of covariance, n=4495).

Variable	Model A		Model B		Model C	
	F score (*df*)	*P* value	*F* score (*df*)	*P* value	*F* score (*df*)	*P* value
Gender	0.90 (1)	.34	0.95 (1)	.33	0.05 (1)	.82
Age	17.37 (1)	<.001	0.54 (1)	.46	2.10 (1)	.15
Urbanization	4.13 (2)	.02	3.74 (2)	.02	2.57 (2)	.08
Self-rated health	51.42 (1)	<.001	11.65 (1)	.001	8.07 (1)	.005
Education	—	—	7.67 (2)	<.001	1.93 (2)	.15
Financial hardship	—	—	9.85 (1)	.002	9.64 (1)	.002
Labor market position	—	—	2.00 (5)	.08	0.87 (5)	.50
Living alone	—	—	1.01 (1)	.32	0.04 (1)	.84
Voting	—	—	0.07 (1)	.79	2.21 (1)	.14
Relationships	—	—	0.59 (1)	.44	0.48 (1)	.49
Keeping in touch	—	—	15.39 (1)	<.001	6.84 (1)	.009
Access	—	—	—	—	58.32 (1)	<.001
Skills	—	—	—	—	93.14 (1)	<.001
Extent of use	—	—	—	—	54.10 (1)	<.001

Perceived Collaboration Benefits

Table 4 shows the results regarding perceived collaboration benefits. In model A, age and health status were significantly associated with collaboration benefits. Older respondents and those with poor self-rated health perceived fewer collaboration benefits. The association between health and perceived collaboration benefits remained significant after all adjustments, whereas the association with age attenuated to nonsignificant after adjusting for ICT-related variables. In model B, satisfaction with relationships and keeping in touch with friends and relatives were both significantly associated with collaboration benefits. Perceived collaboration benefits were higher among those who were satisfied with their relationships and those who were keeping in touch with their friends and relatives. These associations remained significant even after adjusting for ICT-related variables. In model C, higher levels of access, skills, and extent of use were all associated with higher levels of perceived collaboration benefits. R^2 of model C was .15.

Table 4. The association of explanatory factors with perceived collaboration benefits (analysis of covariance, n=4495).

Variable	Model A		Model B		Model C	
	F score (*df*)	*P* value	*F* score (*df*)	*P* value	*F* score (*df*)	*P* value
Gender	0.20 (1)	.65	0.00 (1)	.99	0.38 (1)	.54
Age	24.46 (1)	<.001	4.55 (1)	.03	0.00 (1)	.99
Urbanization	2.63 (2)	.07	3.04 (2)	.048	3.29 (2)	.04
Self-rated health	48.81 (1)	<.001	14.48 (1)	<.001	1.26 (1)	.001
Education	—	—	2.10 (2)	.12	0.54 (2)	.58
Financial hardship	—	—	2.46 (1)	.12	2.23 (1)	.14
Labor market position	—	—	0.57 (5)	.72	0.30 (5)	.91
Living alone	—	—	1.78 (1)	.18	0.46 (1)	.50
Voting	—	—	1.73 (1)	.19	5.43 (1)	.02
Relationships	—	—	7.68 (1)	.006	6.86 (1)	.009
Keeping in touch	—	—	25.71 (1)	<.001	15.05 (1)	<.001
Access	—	—	—	—	56.14 (1)	<.001
Skills	—	—	—	—	111.77 (1)	<.001
Extent of use	—	—	—	—	38.90 (1)	<.001

XSL•FO
RenderX

Additional Analyses

Given that ICT-related variables attenuated the association of age, education, and urbanization to nonsignificant, we additionally examined the associations between these variables and access, skills, and extent of use in separate analyses of variance. All associations were statistically highly significant. Older respondents reported lower levels of access (F_1=842.8, P<.001), skills (F_1=1069.5, P<.001), and extent of use (F_1=266.2, P<.001) compared with younger respondents. Those who were living in urban areas reported higher levels of access (F_2=37.2, P<.001), skills (F_2=35.7, P<.001), and extent of use (F_2=29.9, P<.001) compared with those living in semiurban or rural areas. Those with a low education level reported lower levels of access (F_2=47.5, P<.001), skills (F_2=39.4, P<.001), and extent of use (F_2=36.1, P<.001) compared with those having intermediate or high education.

Discussion

Principal Findings

This study examined the associations of demographics, health, socioeconomic position, social participation, and ICT variables with perceived benefits of using online services. We found that access to online services, ICT skills, and extent of use were consistent factors associated with all examined benefits (ie, with perceived health, economic, and collaboration benefits). ICT skills seemed to be the most important factor. Moreover, self-rated poor health was consistently associated with lower levels of perceptions of each of the benefits examined. Experienced financial hardship seemed to be the most important indicator of socioeconomic position, given that poor people who reported having been forced to bargain for basic needs such as food, medication, or physician visits due to financial constraints perceived fewer health and economic benefits from online services. Dimensions of social participation were also highly relevant to perceived benefits. Those who were keeping in touch with their friends and relatives at least once a week perceived online services beneficial regarding all dimensions examined. Those who were satisfied with their relationships reported higher levels of health and collaboration benefits compared with other respondents.

Older respondents perceived less of each of the benefits examined, but this seemed to be at least partly explained by their lower access to online services, poorer skills when using these services, and lower extent of use. Those living in urban areas perceived more economic benefits than others, but this seemed to be at least partly explained by their higher access, skills, and use. Gender was unrelated to perceived benefits.

In addition to experienced financial hardship, there were also significant associations of other dimensions of socioeconomic position with the examined benefits. Those with a low level of education perceived lower levels of health and economic benefits. However, these significant associations also attenuated to nonsignificant after adjusting for ICT variables suggesting that the effects of education on perceived benefits from online services may also be largely explained by low access, poor skills, and low use among those with a lower level of education.

Comparison With Previous Results

This study showed the importance of ICT access, skills, and the extent of use for perceived benefits of online services. Previous studies have mainly focused on the determinants of these variables but not on their role for the outcomes such as perceived benefits [9,23]. Access and digital skills have been identified as being vital for the success of personal health record use and for ensuring that the records will not become limited to those who are already linked to the internet with high levels of health literacy and computer skills [24]. Internet use has previously been associated with economic, social, and institutional benefits [5]. Access and skills have also been associated with the adoption, engagement, and sustained use of patient portals [25-27].

Our findings suggest that people with economic disadvantage, poor health, and a low level of participation see fewer benefits from online services even when differences in access, skills, and use are accounted for. This corresponds to a previous finding showing that those with below-average earnings gained fewer economic benefits from the internet use than those with average earnings [5]. It is possible that people from these disadvantaged groups need more personal and individually tailored health and social services, which evokes an urge for getting services provided in an interpersonal encounter. Online services may be more suited for providing care for uncomplicated problems in a standardized manner and prevention of health or welfare problems. The health and welfare problems of disadvantaged groups may tend to be more complex, including multimorbidity, and potentially include social problems. Thus, they may require multisectoral and multidisciplinary services and simultaneous, coordinated action of different professionals.

Our results suggest that older people experience fewer benefits from online services, but this association is partly due to their lower levels of access, poorer skills, and lower use of online services. Previous studies have mainly focused on use and found mixed results [8,9]. One finding suggests that the young benefit more from internet use than older people [5], but another study had opposite findings [10]. Skills and access were not adjusted for in these studies, however, and our results suggest that they are of importance. For example, in Italy researchers found that a substantial proportion of those aged 50 years or older were not digitally skilled [28]. Because skills seem to be crucial for older people to benefit from online services, training should be provided for them [29]. A study found that older people were satisfied with ICT training that included working in couples, practicing with the device, choosing what to learn, and practicing material that facilitates communication and learning [30].

Differences in education level could be partly explained by differential access, skills, and use. Previously, lower education has been found to be associated with fewer benefits from internet use [5,10] and lower use of governmental e-services [8]. Computer nonusers have been found to have less education, higher unemployment, lower annual income, and poorer health and be less likely to have memberships in community organizations or perform volunteer work than their counterparts [9].

XSL•FO
RenderX

Previous studies have found mixed results regarding findings related to urbanization level. Rural participants are less likely to have regular access to the internet or manage personal health information online and email health care providers in the United States [31]. However, it has been found that rural patients engage digital rehabilitation means more than urban patients [32]. In Sweden, primary care choice reform in 2010 seemed to have negative effects on geographic equity; the effects do not appear to be very large but they concern some indicators of socioeconomically less advantaged areas [33]. The introduction of electronic online services may face special challenges in rural areas [34]. It has been suggested that rural online service implementation only leads to sustainable adoption (ie, it sticks) when the implementation carefully considers and aligns the content (the clicks), the preexisting structures in the context (the bricks), and the interventions in the implementation process (the tricks) [35].

Our findings that gender was unrelated to experienced benefits are supported by an earlier study from the Netherlands, where there were no gender differences related to who benefits from internet use [5]. Correspondingly, in Finland there were no gender differences in the use of governmental online services [8].

Deursen and Helsper [5] suggested that economic benefits of digital services are related to economic resources such as education, whereas social benefits are more related to social resources. Our findings give only partial support to this, given that satisfaction with relationships and keeping in touch with friends and relatives were associated with perceived collaboration benefits but also associated with other forms of benefits. Correspondingly, education was associated with economic benefits but also with health benefits and did not remain significant after adjusting for ICT variables. A previous study has found support for this idea showing that differences in economic outcomes were related to economic resources such as education and income, whereas differences in social outcomes were related to social resources such as marital status [5].

Strengths and Limitations

Previous studies have mainly focused on first-level digital divide or exclusion showing that demographics, health, socioeconomic position, and social participation are associated with the use of internet and online services. However, it has been suggested that shifting research focus from the first-level divide to the third-level divide and examining the benefits from online services and the inequalities in these outcomes is important [16,17]. Research related to social determinants of benefits is especially lacking [17]. Focusing on aspects of the third-level divide or exclusion and factors related to them is a strength of this study.

This study used a large random sample representative of the Finnish population. However, it relied on self-reported measures, which may lead to problems associated with an inflation of the strengths of associations and with common method variance. To minimize problems with self-reports, we used measures that showed good reliability. Our study was cross-sectional, thus we cannot draw any conclusions about causality. Moreover, although we controlled for many factors, we cannot rule out the possibility of residual confounding. In addition, there are many possible determinants affecting the use and benefits of online services that we did not examine. For example, we did not examine attitudes toward ICT and motivation to use it, satisfaction with care provider, or personality. Future studies should examine these factors and the different types of health internet users (ie, a recent study found 6 different types: learners, pragmatists, skeptics, worriers, delegators, and adigitals [36]).

Finland is among the forerunners in the digitalization of health care and social welfare services [33]: the national digital Kanta services are unique in the world, and tax-financed universal health care is provided for all residents. Therefore, generalizing our findings to countries with other types of health care systems or online services should be done with caution.

In our sample the respondents were older, had more education, and were more often women than the eligible population [18]. One reason for this was that we used double picking probability for those 75 years of age or older to guarantee a sufficient group size for older people, given that they are an important group suggested to be at risk of digital divide and exclusion. To tackle possibly biased results, we used inverse probability weighting based on age, gender, marital status, education level, living region, and degree of urbanization of the residential municipality.

Conclusions

Our findings suggest that access to electronic online services, skills to use them, and the extent of use play crucial roles in perceiving benefits from these services. Skills seem to be the most central element. Moreover, health, financial hardship, and participation-related variables seem to be important as well. Thus, it seems that there is a significant risk of digital exclusion among economically disadvantaged people, those with poor health, and those who are socially isolated. In times when services are increasingly provided by electronic means, this digital divide may predispose people who do not perceive benefit from online services to be excluded from those services.

To ensure that the population, particularly those who are older, in poor health, socially isolated, or of low socioeconomic position, can equally benefit from online health services, access, skills, and possibilities to use these services must be provided. These groups are at the highest risk of digital exclusion and have the highest need for health and social welfare services. Promoting use would also increase benefits for organizations by enabling more users for online services.

Providing services online has the potential to widen social and health inequalities among the population. Widespread expansion calls for rigorous consideration of interventions aimed at tackling the negative effects that can arise from providing health and social care services online and promoting equal opportunities and capabilities among the population [37]. In accordance with Öberg et al [38], targeting training toward vulnerable groups such as senior citizens and people with poor health, lower levels of education, or social isolation may help to ensure that online health services are accessible and can reach a wide population and improve client involvement in their own care. Moreover, organizations should consider offering

XSL•FO
RenderX

instruction and support services to improve patient engagement [29].

Acknowledgments

This study was supported by the Strategic Research Council at the Academy of Finland (project 327145) and the Ministry of Social Affairs and Health (project 112241). The authors wish to thank all the respondents of the questionnaires.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire items.
[DOCX File , 19 KB - jmir_v22i7e17616_app1.docx]

References

1. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008 Nov(175):1-1422. [Medline: 19408968]
2. National Telecommunications and Information Administration. 1999. Falling through the net: defining the digital divide URL: http://www.ntia.doc.gov/ntiahome/fttn99/fttn.pdf [accessed 2020-04-23]
3. van Dijk JA. Digital divide research: achievements and shortcomings. Poetics 2006 Aug;34(4-5):221-235. [doi: 10.1016/j.poetic.2006.05.004]
4. van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Inform Soc 2003 Sep;19(4):315-326. [doi: 10.1080/01972240309487]
5. van Deursen AJ, Helsper EJ. The third-level digital divide: who benefits most from being online? Commun Inform Technol Ann Dig Distinctions Inequalities Studies Media Comm 2015;10:29-53. [doi: 10.1108/s2050-206020150000010002]
6. Castells M. The Internet Galaxy: Reflections on the Internet, Business, and Society. New York: Oxford University Press; 2011.
7. Helsper EJ, Reisdorf BC. A quantitative examination of explanations for reasons for internet nonuse. Cyberpsychol Behav Soc Netw 2013 Feb;16(2):94-99. [doi: 10.1089/cyber.2012.0257] [Medline: 23249243]
8. Taipale S. The use of e-government services and the internet: the role of socio-demographic, economic and geographical predictors. Telecomm Pol 2013 May;37(4-5):413-422. [doi: 10.1016/j.telpol.2012.05.005]
9. Cresci MK, Yarandi HN, Morrell RW. The digital divide and urban older adults. Comp Inform Nursing 2010;28(2):88-94. [doi: 10.1097/ncn.0b013e3181cd8184]
10. Blank G, Lutz C. Benefits and harms from internet use: a differentiated analysis of Great Britain. New Med Soc 2016;20(2):618-640. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
11. Fieseler C, Meckel M, Müller S. With a little help of my peers. The supportive role of online contacts for the unemployed. CompHum Beh 2014 Dec;41:164-176. [doi: 10.1016/j.chb.2014.09.017]
12. Warschauer M. Technology and Social Inclusion: Rethinking the Digital Divide. Cambridge: MIT Press; 2004.
13. United N. E-government for leaving no one behind. In: United Nations E-Government Survey 2018: Gearing E-Government to Support Transformation Towards Sustainable and Resilient Societies. New York: United Nations, Department of Economic and Social Affairs; 2018:27-46.
14. Helsper EJ. A corresponding fields model for the links between social and digital exclusion. Commun Theor 2012 Oct 15;22(4):403-426. [doi: 10.1111/j.1468-2885.2012.01416.x]
15. Aceto G, Persico V, Pescapé A. The role of information and communication technologies in healthcare: taxonomies, perspectives, and challenges. J Netw Comp Applic 2018 Apr;107:125-154. [doi: 10.1016/j.jnca.2018.02.008]
16. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002;7:4. [doi: 10.5210/fm.v7i4.942]
17. Scheerder A, van Deursen A, van Dijk J. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Informat 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
18. Hyppönen H, Aalto AM. Citizen experiences of e-healthe-welfare services. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-Health and E-Welfare of Finland Checkpoint. Helsinki: National Institute for Health and Welfare; 2019:148-167.
19. Härkänen T, Kaikkonen R, Virtala E, Koskinen S. Inverse probability weighting and doubly robust methods in correcting the effects of non-response in the reimbursed medication and self-reported turnout estimates in the ATH survey. BMC Public Health 2014 Nov 06;14:1150 [FREE Full text] [doi: 10.1186/1471-2458-14-1150] [Medline: 25373328]

XSL•FO
RenderX

20. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. Finnish J EHealth EWelfare 2018;10(4):381-395. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
21. Jormanainen V, Parhiala K, Niemi A, Erhola M, Keskimäki I, Kaila M. Half of the Finnish population accessed their own data: comprehensive access to personal health information online is a corner-stone of digital revolution in Finnish health and social care. FinJeHeW 2019 Nov 02;11(4):310. [doi: 10.23996/fjhw.83323]
22. Reponen J, Kangas M, Hämäläinen P, Haverinen J, Keränen N. Availability and use of e-health in Finland. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-Health and E-Welfare of Finland Checkpoint. Helsinki: National Institute for Health and Welfare; 2018:52-86.
23. Andreasson K. Digital Divides: The New Challenges and Opportunities of E-Inclusion. Boca Raton: CRC Press; 2015.
24. Kahn JS, Aulakh V, Bosworth A. What it takes: characteristics of the ideal personal health record. Health Aff (Millwood) 2009;28(2):369-376 [FREE Full text] [doi: 10.1377/hlthaff.28.2.369] [Medline: 19275992]
25. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
26. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
27. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
28. Di Giacomo D, Ranieri J, D'Amico M, Guerra F, Passafiume D. Psychological barriers to digital living in older adults: computer anxiety as predictive mechanism for technophobia. Behav Sci (Basel) 2019 Sep 11;9(9):1 [FREE Full text] [doi: 10.3390/bs9090096] [Medline: 31514364]
29. Crouch E, Gordon NP. Prevalence and factors influencing use of internet and electronic health resources by middle-aged and older adults in a US health plan population: cross-sectional survey study. JMIR Aging 2019 Mar 26;2(1):e11451 [FREE Full text] [doi: 10.2196/11451] [Medline: 31518256]
30. López Seguí F, de San Pedro M, Aumatell Verges E, Simó Algado S, Garcia Cuyàs F. An intergenerational information and communications technology learning project to improve digital skills: user satisfaction evaluation. JMIR Aging 2019 Aug 09;2(2):e13939 [FREE Full text] [doi: 10.2196/13939] [Medline: 31518276]
31. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. J Rural Health 2018 Feb;34 Suppl 1:s30-s38. [doi: 10.1111/jrh.12228] [Medline: 28075508]
32. Munsell M, De Oliveira E, Saxena S, Godlove J, Kiran S. Closing the digital divide in speech, language, and cognitive therapy: cohort study of the factors associated with technology usage for rehabilitation. J Med Internet Res 2020 Feb 07;22(2):e16286 [FREE Full text] [doi: 10.2196/16286] [Medline: 32044752]
33. Isaksson D, Blomqvist P, Winblad U. Free establishment of primary health care providers: effects on geographical equity. BMC Health Serv Res 2016 Jan 23;16:28 [FREE Full text] [doi: 10.1186/s12913-016-1259-z] [Medline: 26803298]
34. Sørensen JFL. Attitudes toward telehealth use among rural residents: a Danish survey. J Rural Health 2008;24(3):330-335. [doi: 10.1111/j.1748-0361.2008.00178.x] [Medline: 18643814]
35. Hage E, Roo JP, van Offenbeek MAG, Boonstra A. Implementation factors and their effect on e-Health service adoption in rural communities: a systematic literature review. BMC Health Serv Res 2013;13:19 [FREE Full text] [doi: 10.1186/1472-6963-13-19] [Medline: 23311452]
36. Powell J, Deetjen U. Characterizing the digital health citizen: mixed-methods study deriving a new typology. J Med Internet Res 2019 Mar 05;21(3):e11279 [FREE Full text] [doi: 10.2196/11279] [Medline: 30835238]
37. Latulippe K, Hamel C, Giroux D. Social health inequalities and ehealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
38. Öberg U, Isaksson U, Jutterström L, Orre CJ, Hörnsten A. Perceptions of persons with type 2 diabetes treated in Swedish primary health care: qualitative study on using ehealth services for self-management support. JMIR Diabetes 2018 Mar 12;3(1):e7 [FREE Full text] [doi: 10.2196/diabetes.9059] [Medline: 30291075]

Abbreviations

ICT: information and communication technologies
EM: estimated mean

XSL•FO
RenderX

Edited by G Eysenbach; submitted 27.12.19; peer-reviewed by K Saranto, H Hannerz, V Minichiello; comments to author 11.02.20; revised version received 24.02.20; accepted 22.03.20; published 07.07.20.

Please cite as:
Heponiemi T, Jormanainen V, Leemann L, Manderbacka K, Aalto AM, Hyppönen H
Digital Divide in Perceived Benefits of Online Health Care and Social Welfare Services: National Cross-Sectional Survey Study
J Med Internet Res 2020;22(7):e17616
URL: https://www.jmir.org/2020/7/e17616
doi:10.2196/17616
PMID:32673218

©Tarja Heponiemi, Vesa Jormanainen, Lars Leemann, Kristiina Manderbacka, Anna-Mari Aalto, Hannele Hyppönen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relationships Between the Usage of Televisions, Computers, and Mobile Phones and the Quality of Sleep in a Chinese Population: Community-Based Cross-Sectional Study

Yao Jie Xie[1], PhD; Daphne SK Cheung[1,2], PhD; Alice Y Loke[1], PhD; Bernice L Nogueira[3], PhD; Karry M Liu[3], PhD; Angela YM Leung[1,2,4], PhD; Alice SM Tsang[5], PhD; Cindy SU Leong[3], DHSc; Alex Molassiotis[1,4], PhD

[1]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]Centre for Gerontological Nursing, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]School of Health Sciences and Sports, Macao Polytechnic Institute, Macau, Macao

[4]World Health Organization Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[5]Kiang Wu Nursing College of Macau, Macau, Macao

Corresponding Author:
Yao Jie Xie, PhD
School of Nursing
The Hong Kong Polytechnic University
FG424, PolyU
Hong Kong
China (Hong Kong)
Phone: 852 63135399
Email: grace.yj.xie@polyu.edu.hk

Abstract

Background: No study has comprehensively investigated the association between the usage of typical screen-based electronic media devices and sleep quality in a Chinese population with individuals in a wide range of ages.

Objective: This study aimed to understand the characteristics of television (TV) viewing, computer usage, and mobile phone usage in a representative Chinese population in Macau and to examine their roles in predicting the variations in sleep quality.

Methods: This cross-sectional study was an analysis of 1500 Macau residents aged 15 to 90 years based on a community-based health needs assessment study entitled, “Healthy Living, Longer Lives.” Data collection was conducted in 7 districts of Macau from 2017 to 2018 through face-to-face interviews. The durations of daily TV viewing, computer usage, and mobile phone usage were recorded in a self-administered questionnaire. The Chinese version of the Pittsburgh Sleep Quality Index (PSQI) was used to assess the sleep quality.

Results: The prevalence of TV, computer, and mobile phone usage was 78.4% (1176/1500), 51.6% (769/1490), and 85.5% (1276/1492), respectively. The average daily hours of usage were 1.75 (1.62), 1.53 (2.26), and 2.85 (2.47) hours, respectively. Females spent more time watching TV (P=.03) and using mobile phones (P=.02) and less time on the computer (P=.04) as compared to males. Older adults were more likely to watch TV while young people spent more time using the computer and mobile phones (P for all trends<.001). The mean PSQI global score was 4.79 (2.80) among the participants. Females exhibited significantly higher PSQI scores than males (5.04 vs 4.49, respectively; P<.001). No linear association was observed between the PSQI score and the amount of time spent on the 3 electronic devices (P=.58 for PSQI-TV, P=.05 for PSQI-computer, and P=.52 for PSQI-mobile phone). Curve estimation showed significant quadratic curvilinear associations in PSQI-TV (P=.003) and PSQI-computer (P<.001) among all the participants and in PSQI-mobile phone among youths (age, 15-24 years; P=.04). After adjustment of the gender, age, body mass index, demographics, and lifestyle factors, more than 3 hours of TV viewing and 4 hours of computer usage or mobile phone usage was associated with 85% (95% CI 1.04-1.87; P=.008), 72% (95% CI 1.01-2.92; P=.045), and 53% (95% CI 1.06-2.22; P=.03) greater odds of having poor sleep quality (PSQI score>5), respectively.

Conclusions: The mobile phone was the most popular screen-based electronic device used in the Macau population, especially among young people. “J” shape associations were observed between sleep quality and the duration of TV viewing, computer usage, and mobile phone usage, indicating that the extreme use of screen-based electronic devices predicted poorer sleep status, whereas moderate use would be acceptable.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e18095) doi:10.2196/18095

KEYWORDS

electronic device; screen-based; sleep; Chinese; digital; mobile phone

Introduction

The use of screen-based electronic media devices such as computers, televisions (TVs), mobile phones, video games, and other screen-based products have become an integral part of our lives and has continued to increase sharply in the past decade [1,2]. For example, about 80% of the population use smartphones and about 90% of the population watch TV in the United States of America, the United Kingdom, Mainland China, and Hong Kong [2]. However, the pervasive use of screen-based electronic media devices is a likely contributor to widespread sleep problems [3].

Numerous studies across the countries have suggested that prolonged daily use of screen-based electronic devices lead to negative sleep outcomes, including decreased sleep duration, reduced sleep efficiency, increased daytime sleepiness, and sleep disturbances [1,4,5]. A recent systematic review showed that in over 90% of the studies, increased screen-based media consumption by children and adolescents (age, 5-17 years) was associated with a reduction in total sleep duration [3]. Another systematic review of 67 studies in school-aged youth and teenagers found that screen time was associated adversely with sleep health through delayed bedtimes and reduced sleep duration [6]. Regarding the Chinese population, a cross-sectional study on adolescents in Hong Kong showed that prolonged mobile phone use was significantly related to daytime sleepiness, decreased sleep duration, and poor sleep quality [7]. Similar sleep problems were reported among the residents of Macau, which is another special administrative region of China. Nearly 40% of the adults in Macau reported at least one type of sleep disturbance [8]. We believe that most residents of Macau use electronic media devices because the cultural customs and lifestyle of Macau are similar to those of Hong Kong. However, no study has examined the role of screen-based electronic device usage in the development of sleep problems in Macau.

The majority of the previous studies focused on a special age group—in particular, the age groups of children and adolescents [1,3,9-12]. Very few studies have targeted a population of a wide age range. Youths are not the only age group that use screen-based electronic devices. The different pattern of use in adulthood may reflect a different sleep health profile. For example, adults may seek compensatory sleep when spending too much time on screen-based media devices, and thus, media use is associated with sleep onset latency but not tiredness [13]. Many of the previous studies focused only on 1 or 2 screen-based electronic devices [3,10,14,15]. It is also not known whether the same pattern of sleep problems occurs with the usage of different screen-based electronic media devices [1]. In fact, the trend in the modern society is the simultaneous use of multiple screen-based electronic devices and thus, a comprehensive investigation of several screen-based electronic media devices may help us to better understand the overall situation [6].

We conducted this study with the following aims: (1) describe the usage of 3 major screen-based electronic media devices (TV, computer, and mobile phone) and the sleep quality in a representative Chinese population in Macau, (2) examine the patterns of the associations between electronic device usage and sleep quality, and (3) explore how the usage of electronic devices predicts poor sleep status.

Methods

Data Source and Study Design

This study is a report of the partial data from a project named "Healthy life, longer lives (HLLL)," which was conducted from 2017 to 2018 in Macau. HLLL is a health needs assessment study of the Macau civilian, noninstitutionalized population. It was designed to assess the lifestyle behaviors of Macau residents in terms of their dietary habits, sleep patterns, physical activity, electronic device usage, smoking habits, drinking habits, and health information seeking behaviors, and to identify the barriers for a healthy lifestyle in vulnerable groups. The ethical approval for the study was obtained from the Human Research Committee of Hong Kong Polytechnic University (HSEARS20180516001), the Macau Polytechnic Institution (07/PSC/ESS/2017), and the Kiang Wu Nursing College of Macau (REC–2017.3). All participants signed informed consent forms.

Participants and Sampling Method

The inclusion criteria of the participants were (1) aged 15 years or above, (2) can read and write Chinese and speak Cantonese or Putonghua, which are the dialects commonly used in the city, and (3) residing in one of the 7 parishes in Macau. The exclusion criteria were (1) nonpermanent residents, (2) unable to communicate owing to psychological or physical illness, and (3) cannot provide written informed consent.

A multiple-stage sampling frame according to gender and age was adopted to recruit the participants. First, the number of the needed participants in each age group (age [years]: 15-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+) was determined by referring to the distribution of age and gender in the census data of the Macau population [16]. Second, several recruitment strategies were adopted to approach potential participants in each age group. Thirty baccalaureate nursing students from 2 tertiary education institutions in Macau and living in different parishes of Macau were trained for recruitment and data collection. They went to their living districts and the nearby housing districts to recruit participants. They were also asked to invite 1-3 friends or relatives that belonged to the same age groups but with different education levels to participate in the study. To recruit enough adolescents (age, 15-17 years) and older adults (aged 65 years or above), the trained students also went to the public-funded health care and social service centers for adolescents and older adults for recruitment. To ensure the representativeness of the sample, only 1 eligible person in 1 household was considered as a valid participant. To avoid

XSL•FO
RenderX

duplication, we asked the participants to provide the last 3 numbers of their Macau national identification numbers. These Macau identification numbers were stored in a separate data set and the numbers were checked thoroughly, and we did not find any duplicate cases. Upon successful contact with the target person and after obtaining written informed consent, the self-administered questionnaire was given to the participants to complete. If he/she could not read the questionnaire (eg, older adults), the student assistants conducted a face-to-face interview thereafter. According to the feedback from the student assistants, only a small number of older persons aged 75 years and above completed the questionnaires by interview, accounting to less than 1% of the whole sample.

Measures

Screen-Based Electronic Media Device Usage

The screen-based electronic devices analyzed in this study were the TV, computer, and mobile phone, as these are the 3 most frequently used screen-based electronic media devices. Respondents were asked how many days they spent watching TV or using a computer and mobile phone in 1 week and the duration they spent on TV viewing and computer and mobile phone usage on a typical day (minutes or hours/day). Minutes were converted to hours and recorded to the nearest 0.1 hours. The upper bounds of the duration of each electronic device usage was set as 16 hours, because we considered that one person might spend 8 hours for sleep and for other things in a day. The total number of hours spent on the 3 electronic devices in 1 day should not be beyond 24 hours. The student assistants corrected the answers with the participants after receiving the questionnaires to avoid any unreasonable value. The average time reported for using each electronic device every day was calculated by dividing the sum of the hours spent on these devices in 1 week divided by 7 (ie, 7 days in a week).

Sleep Quality

The Chinese version of the Pittsburgh Sleep Quality Index (PSQI) [17,18] was used to measure the sleep quality. The PSQI has 7 components, namely, (1) sleep duration, (2) sleep disturbance, (3) sleep latency, (4) daytime dysfunction due to sleepiness, (5) sleep efficiency, (6) overall sleep quality, and (7) sleep medication. The scale consists of 19 items, and each item is scored from 0 to 3. The PSQI global score was calculated from the 7 component scores. The score ranges from 0 to 21. Higher PSQI global score indicates poorer subjective sleep quality. The Chinese version of PSQI has been identified with good reliability (r=0.82-0.83) [17]. A global PSQI score >5 is considered as poor sleep quality with a sensitivity of 98% and a specificity of 55% [17], indicating good performance when the PSQI is used in the Chinese population.

Demographics and Anthropometric Measures

Information on the following demographic characteristics were solicited: gender (male or female), age (years), and age groups (age [years]: 15-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+), living region (7 parishes of Macau), employment status (ie, employed, retired, student, housewife, unable to work, unemployed), marital status (ie, never married, married, divorced, widowed, others), education level (ie, primary or below, secondary, tertiary or above), and monthly household income (in MOP, conversion rate: 1MOP=US $0.125). Participants reported their weights and heights by kilogram and centimeters, respectively. Body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters. Participants were categorized as underweight (BMI<18.5 kg/m^2), normal weight (18.5 kg/m^2≤BMI<23.0 kg/m^2), overweight (23.0 kg/m^2≤BMI<25.0 kg/m^2), and obese (BMI≥25.0 kg/m^2) according to the World Health Organization (WHO) standard for Asian populations [19]. The former 3 categories were designated as nonobese (BMI<25.0 kg/m^2) in some analyses to compare with the obese group (BMI≥25.0 kg/m^2).

Other Covariates

The following lifestyle factors were considered as potential covariates and measured in the questionnaire: physical activity (insufficient/sufficient according to WHO standards), smoking (currently smoking or not), drinking (currently drinking or not), and unhealthy eating habits in terms of remaining hungry (eg, delaying or skipping meal), dieting (eg, weight losing, fasting), and overeating. The eating habits were measured using the 5-point Likert scale ranging from Never (1), Occasionally (2), Sometimes (3), Frequently (4), to Always (5).

Statistical Analysis

The basic characteristics of the participants were described by the mean and SD for the continuous variables and by count and percentage for the categorical variables. To describe the usage of the device and the sleep quality, the duration of usage (hours), status of usage (yes/no), global score of the PSQI, and poor sleep status were presented by mean (SD) or frequency (n [%]) where appropriate. The device usage and sleep quality according to the different demographical groups were analyzed using one-way analysis of variance (ANOVA) or independent t-test for the continuous variables and chi-square test for the categorical variables. The mean sleep duration (hours) and the average 7 component scores of the PSQI were presented and compared with gender and age groups by ANOVA. To examine the association between screen-based electronic device usage and sleep quality, the average number of hours spent in using electronic devices were categorized into different strata and ANOVA was used to compare the differences in the PSQI scores among the groups. Both linear and curve estimation models were performed to demonstrate the nature of the association between using electronic devices and sleep quality. Binary logistic regression analysis was adopted to determine the predictive role of the electronic device usage on poor sleep status. The dependent variable was poor sleep quality (PSQI score>5). Odds ratios (ORs) and 95% CIs for having poor sleep quality according to different durations of using the computer, mobile phone, and TV were calculated accordingly. SPSS 25.0 software (IBM Corp) was used in all data analyses and the significance level was considered as .05.

XSL•FO
RenderX

Results

General Characteristics of the Participants

A total of 1500 participants were recruited between 2017 and 2018. The mean age of the females was 39.1 years and that of males was 39.3 years. The 4 age groups were as follows: youth (15-24 years), young adults (25-44 years), middle-aged adults (45-64 years), and older adults (65 or above). The proportions of the 4 age groups were as follows: youth, 14.1% (212/1500); young adults, 35.4% (531/1500); middle-aged adults, 37.4% (561/1500); and older adults, 13.1% (196/1500). The majority of the participants were employed (1030/1473, 70.0%), had secondary or above education level (1206/1489, 81.0%), and were married or they had a partner (899/1487, 60.5%). Females had lower monthly household income than males (P<.001). Males were more likely to be overweight or obese than females (289/667, 43.3% vs 269/805, 33.4% respectively; P<.001). The mean BMIs of the males and females were 22.8 kg/m^2 and 22.0 kg/m^2, respectively (P<.001). Most participants did not have drinking and smoking habits, but the number of male drinkers was higher than that of female drinkers (221/675, 32.7% vs 152/821, 18.5%, respectively; P<.001). The number of male smokers was higher than that of the female smokers (97/673, 14.4% vs 23/821, 2.8%, respectively; P<.001). Around half of the male population (332/669, 49.6%) had enough physical activity while only 35.7% (286/801) of the female population was involved in any physical activity. Over half of the participants reported that they had at least one unhealthy eating habit, and males showed a higher percentage of unhealthy eating habits than females (401/669, 59.9% vs 420/819, 51.3%, respectively; P<.001). The basic characteristics of the participants according to gender are shown in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of the participants (N=1500).

Characteristics	Males	Females
Age (years), mean (SD)	39.3 (17.5)	39.1 (16.8)
Length of stay in Macau (years), mean (SD)[a]	36.4 (15.8)	33.1 (15.2)
BMI, mean (SD)[a]	22.8 (3.1)	22.0 (3.3)
Age group (years), n (%)		
Youth (15-24)	102 (15.1)	110 (13.3)
Young adults (25-44)	240 (35.5)	291 (35.3)
Middle-aged adults (45-64)	241 (35.7)	320 (38.8)
Older adults (≥65)	93 (13.8)	103 (12.5)
Marital status, n (%)		
Single/Divorced/Separated	261 (38.7)	327 (40.3)
Married/with partner	414 (61.3)	485 (59.7)
Education level, n (%)		
Primary or lower	125 (18.5)	158 (19.4)
Secondary	334 (49.6)	428 (52.5)
Tertiary or higher	215 (31.9)	229 (28.1)
Employment status, n (%)[a]		
Employed	468 (70.5)	562 (69.5)
Retired	82 (12.3)	90 (11.1)
Student	73 (11.0)	69 (8.5)
Others (housewife, unable to work, unemployed)	41 (6.2)	88 (10.9)
Monthly household income (MOP)[b], n (%)[c]		
≤MOP $9999	26 (3.9)	83 (10.2)
MOP $10,000-MOP $29,999	275 (40.9)	349 (42.8)
MOP $30,000-MOP $59,999	196 (29.2)	217 (26.6)
≥MOP $60,000	66 (9.8)	49 (6.0)
Unknown	109 (16.2)	117 (14.4)
BMI, n (%)[c]		
Underweight (≤18.5 kg/m^2)	38 (5.7)	100 (12.4)
Normal (18.6-22.9 kg/m^2)	340 (51.0)	436 (54.2)
Overweight (23-24.9 kg/m^2)	159 (23.8)	132 (16.4)
Obese (≥25 kg/m^2)	130 (19.5)	137(17.0)
Lifestyle behaviors, n (%)[c]		
Sufficient physical activity[d]	332 (49.6)	286 (35.7)
Unhealthy eating habits[e]	401 (59.9)	420 (51.3)
Drinking	221 (32.7)	152 (18.5)
Smoking	97 (14.4)	23 (2.8)

[a]*P*<.001.

[b]MOP: Macau Pataca; 1MOP = US $0.125.

[c]*P*<.001.

[d]According to WHO (2018), sufficient physical activity means doing at least 150 minutes of moderate-intensity aerobic physical activity or at least 75

XSL•FO
RenderX

minutes of vigorous-intensity aerobic physical activity throughout the week; *P*=.001.

[e]Unhealthy eating habit means remaining hungry (eg, delaying or skipping meal), dieting (eg, weight losing, fasting), or overeating.

Screen-Based Electronic Media Device Usage Among Participants

Among the 1500 participants, the prevalence of watching TV and using computer and mobile phones was 78.4% (1176/1500), 51.6% (769/1490), and 85.5% (1276/1492), respectively. Multimedia Appendix 1 shows the duration of screen-based electronic device usage and sleep quality according to the demographics and lifestyle factors of the participants. The average number of daily hours of watching TV, using a computer, and using a mobile phone was 1.75 (1.62) hours, 1.53 (2.26) hours, and 2.85 (2.47) hours, respectively. Females spent more time watching TV (*P*=.03) and using mobile phones (*P*=.02) and less time on computers (*P*=.04) than men. The use of the 3 electronic devices showed trends according to age, with older adults being more likely to watch TV and young people spending more time using computers and mobile phones (*P* for all trends<.001). The youth spent the longest time on computers and mobile phones at 2.7 hours and 4.4 hours, respectively. Participants who were couples, had lower educational level, were retired or unemployed, and had lower monthly household income had longer TV viewing times, whereas participants who were single, had a higher education level, who were students, and who were employed reported longer computer and mobile phone usage (*P*<.001). Significant differences were also observed across the different BMI groups. Obese people had longer TV viewing time (*P*=.04) but shorter computer usage (*P*=.008) and mobile phone usage (*P*<.001) as compared to the nonobese people. Regarding the common lifestyle factors, participants with enough physical activity (*P*=.01) and drinking habits (*P*=.003) spent shorter durations on TV viewing, while those who had unhealthy eating habits spent a longer time on TV viewing (*P*<.001) and shorter time on computer (*P*=.001) and mobile phone (*P*<.001) usage. The number and percentage of the users of the 3 electronic devices according to demographics and lifestyle factors is shown in Multimedia Appendix 2.

Sleep Quality of the Participants

The mean (SD) of the PSQI global score was 4.79 (2.80). Females had a significantly higher PSQI global score than males (5.04 vs 4.49; respectively; *P*<.001). The prevalence of poor sleeping quality in females was around 10% higher than that in males (308/794, 38.8% vs 181/642, 28.2%, respectively; *P*<.001). No statistically significant differences in the PSQI global scores were detected according to the age groups (*P*=.55), marital status (*P*=.22), education level (*P*=.07), and employment status (*P*=.06). The highest average PSQI global score was found in obese participants (5.36 (2.96); *P*<.001); 40.4% had poor sleep quality (*P*=.009). Those with unhealthy eating habits (*P*<.001), that is, remaining hungry, dieting, or overeating and smokers (*P*=.002) were more likely to have higher PSQI global scores. However, participants with different physical activity levels (*P*=.97) and drinking habits (*P*=.72) showed no significant difference in sleep quality (Multimedia Appendix 1). The sleep duration (hours) and the 7 component scores of the PSQI by gender and age groups are shown in Multimedia Appendix 3. Males had better subjective sleep quality (*P*=.007) and lower sleep disturbances (*P*=.01) and sleep latency (*P*<.001) than females. Youths had higher daytime dysfunction (*P*=.01) and sleep latency (*P*=.04) but lower sleep disturbances (*P*=.004) than older adults.

Associations Between Screen-Based Electronic Device Usage and PSQI Scores

The associations between electronic device usage and sleep quality were examined by first comparing the mean global scores of PSQI with the stratified duration of electronic device usage. The results are shown in Table 2. Significant differences were found among TV viewing strata and computer usage strata. Those with 1.5-2.5 hours of viewing (PSQI: 4.37; *P*<.001) and 2.0-2.4 hours computer usage (PSQI: 4.25; *P*=.009) had the lowest PSQI scores. However, no significant difference was observed between PSQI scores among the different age groups of mobile phone users (*P*=.23). No linear association was found between sleep quality and the duration of usage of all 3 electronic devices (*P*=.58 for PSQI-TV, *P*=.05 for PSQI-computer, and *P*=.52 for PSQI-mobile phone). Thus, we conducted the curve estimation and observed significant quadratic curvilinear associations of sleep quality with the durations of using TV (*P*=.003) and computer (*P*<.001) but not the mobile phone (*P*=.12). We further analyzed the association in each age group for the mobile phone usage and found a significant quadratic association between mobile phone use time and sleep quality among youths (*P*=.04). We further examined the model fitting through the scatterplot of residuals by fit values for the linear model and quadratic model, respectively. We found that these linear models would be inadequate because their coefficients did not make practical sense, and residuals were not independent of the fit values. The quadratic models did not have these problems, and hence we considered the quadratic models to be the more reasonable models that appropriately described the associations of sleep quality with TV viewing, computer usage, and mobile phone usage. The quadratic associations could be demonstrated using the following equations:

*PSQI global score = 5.033 – 0.268*duration of TV viewing + 0.040*duration of TV viewing²*

*PSQI global score = 5.044 – 0.336*duration of computer usage + 0.036*duration of computer usage²*

*PSQI global score = 4.644 – 0.013*duration of smartphone usage + 0.011*duration of smartphone usage² (age group 15-24 years)*

The association looks like a “J” shape. Both the shorter and longer use of TV, computer, and mobile phone were associated with increased PSQI scores (Figure 1-3).

XSL•FO
RenderX

Table 2. Univariate analysis of the duration of electronic media device usage with sleep quality (PSQI[a] global score).

Duration of device usage	Total (n)	PSQI global scores, mean (SD)	*P* value
Television[b]			<.001
Nonuser	308	5.05 (2.88)	
<1.5 h	394	4.91 (2.79)	
1.5-2.5 h	345	4.37 (2.53)	
2.5-3 h	213	4.42 (2.96)	
>3 h	160	5.13 (2.89)	
Computer[c]			.009
Nonuser	685	5.02 (2.91)	
<2 h	427	4.70 (2.69)	
2-4 h	164	4.25 (2.66)	
>4 h	143	4.64 (2.60)	
Mobile phone[d]			.27
Nonuser	202	4.88 (2.82)	
<2 h	552	4.81 (2.74)	
2-4 h	365	4.57 (2.80)	
>4 h	305	4.98 (2.87)	

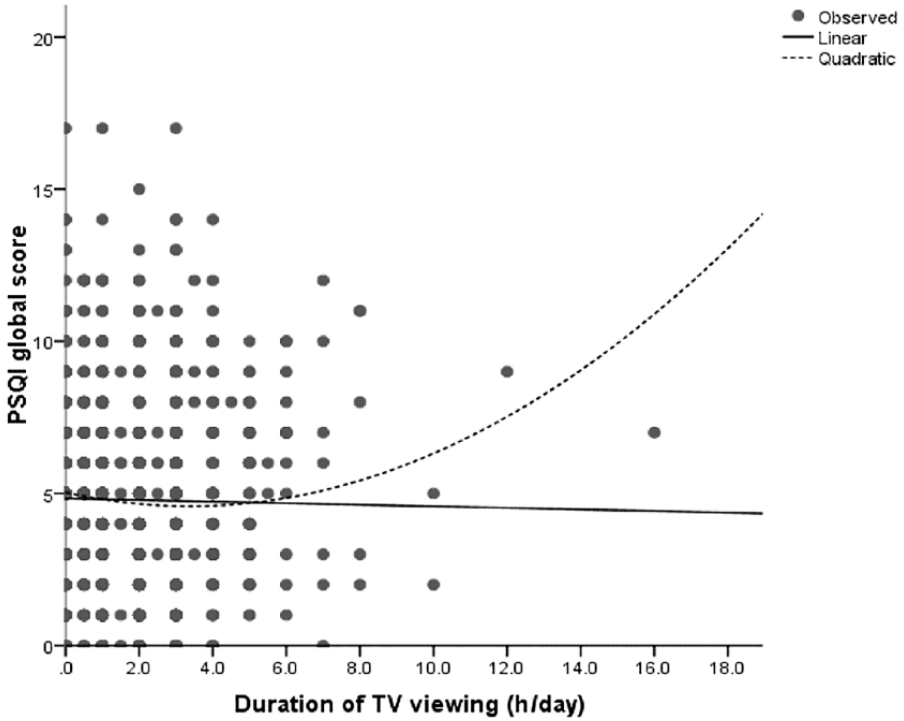
[a]PSQI: Pittsburgh Sleep Quality Index.

[b]No linear association; *P*=.58; quadratic: *P*=.003.

[c]No linear association; *P*=.05; quadratic: *P*<.001.

[d]No linear association; *P*=.52; not quadratic: *P*=.12.

Figure 1. Curve estimation of duration of TV viewing with PSQI global score. PSQI: Pittsburgh Sleep Quality Index; TV: television.



XSL•FO
RenderX

Figure 2. Curve estimation of duration of computer usage with PSQI global score. PSQI: Pittsburgh Sleep Quality Index.

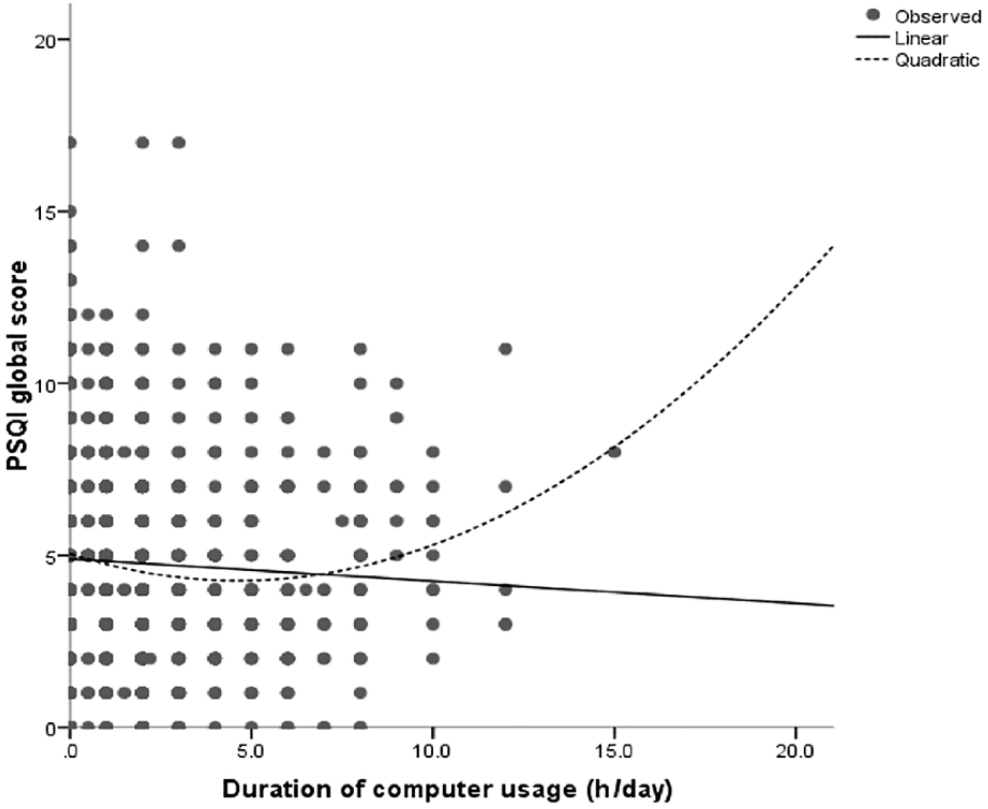
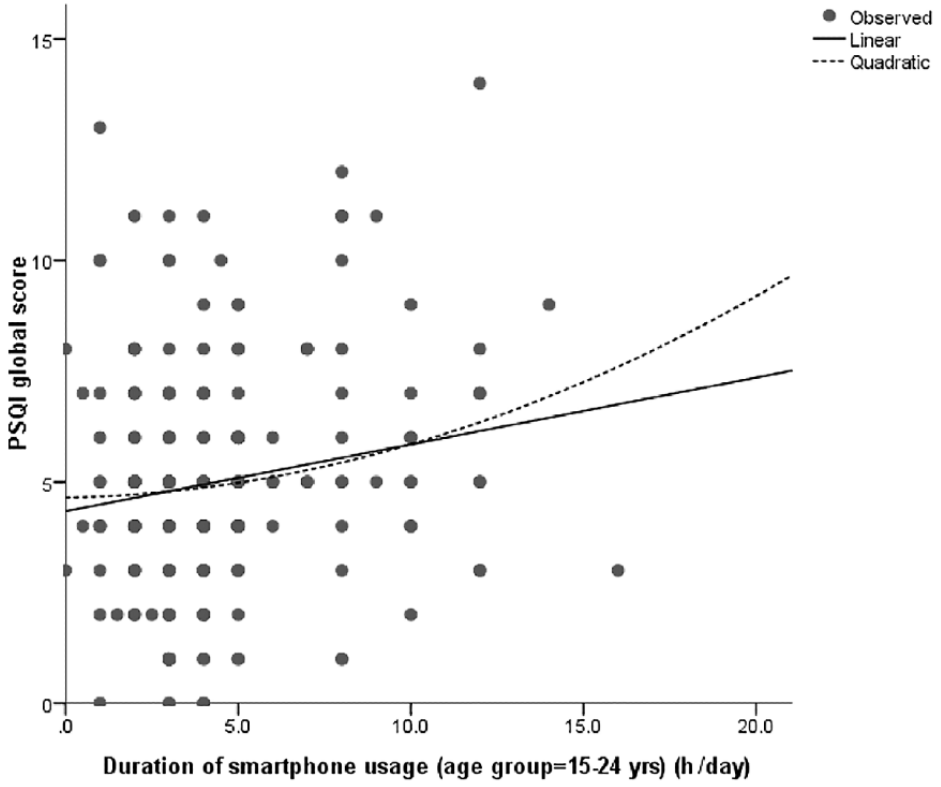


Figure 3. Curve estimation of the duration of mobile phone usage with PSQI global score for the age group of 15-24 years. PSQI: Pittsburgh Sleep Quality Index.



XSL•FO
RenderX

Role of Screen-Based Electronic Device Usage in Predicting Poor Sleep Quality

According to the stratified analyses and curve estimations, the lowest PSQI score occurred in the participants with a moderate duration of using electronic devices. We thereby used 1.5-3 hours of TV viewing and 2-4 hours of computer and mobile phone usage as the reference groups. We conducted binary logistic regression for analyzing the association of poor sleep status with electronic device usage. As shown in Table 3, except for those using mobile phones less than 2 hours, all other groups showed statistically significant associations. The crude ORs ranged from 1.43 to 1.78. After adjustment for gender, age, BMI, demographics, and lifestyle factors, the strength of the associations was slightly changed but they still showed statistical significance. The OR of poor sleep quality was 38% (95% CI 1.03-1.84; *P*=.03) and 85% (95% CI 1.18-2.91; *P*=.008) in the <1.5-hours and >3-hours TV viewing groups, respectively. In the computer usage groups, the OR was 71% (95% CI 1.10-2.64; *P*=.02) higher for <2-hours group and 72% (95% CI 1.01-2.92; *P*=.04) higher for the >4-hours group. As for the mobile phone usage, the adjusted OR was 1.53 (95% CI 1.06-2.22; *P*=.02) for the group with daily mobile phone use longer than 4 hours.

Table 3. Multiple logistic regression models for testing the association between the duration of daily electronic media device usage and poor sleep quality.

	Model 1[a]		Model 2[b]		Model 3[c]	
Device, Usage group	OR[d] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Television						
1.5-3 h	1		1		1	
<1.5 h	1.55 (1.22-1.97)	<.001	1.50 (1.15-1.94)	.003	1.38 (1.03-1.84)	.03
>3 h	1.78 (1.23-2.57)	.002	1.89 (1.29-2.77)	.001	1.85 (1.18-2.91)	.008
Computer						
2-4 h	1		1		1	
<2 h	1.63 (1.12-2.37)	.01	1.66 (1.12-2.46)	.01	1.71 (1.10-2.64)	.02
>4 h	1.66 (1.02-2.72)	.04	1.74 (1.03-2.93)	.04	1.72 (1.01–2.92)	.04
Mobile phone						
2-4 h	1		1		1	
<2 h	1.13 (0.86-1.47)	.39	1.15 (0.86-1.53)	.35	1.10 (0.80-1.52)	.57
>4 h	1.43 (1.04-1.96)	.03	1.45 (1.05-2.01)	.03	1.53 (1.06-2.22)	.02

[a]Simple model.

[b]Adjusted for gender, age, and body mass index.

[c]Adjusted for gender, age, body mass index, demographics, and lifestyle factors.

[d]OR: odds ratio.

Discussion

Principal Findings

Mobile phone and TV were the most popular electronic media devices in our study population, but the usage was polarized by the age groups. Almost all the youth and young adults had mobile phones, while nearly two-thirds of the older adults did not use a mobile phone. On the contrary, over 90% of the older population treated TV viewing as a pastime, while relatively fewer young people watched TV. Another typical finding was that, unlike most previous studies, we found a quadratic curvilinear association between sleep quality and electronic device usage time; this J-shape occurred among all the participants for TV viewing and computer usage and among the youths for mobile phone usage.

Our study found that the prevalence of TV viewing in Macau was approximately 10% lower than that reported in some other countries and regions, such as the United States, the United Kingdom, Japan, Mainland China, and Hong Kong [2]. About half of the participants used computers daily, which was consistent with that reported in other Chinese populations in Mainland China and Hong Kong [2]. Regarding mobile phone usage, our study population exhibited a slightly higher prevalence than that reported in other countries and regions [2]. In our study, the longest number of hours was spent daily on mobile phone use, followed by TV viewing and computer use. Internet access is the key for people choosing electronic media devices. A smartphone can facilitate access to the internet rapidly and conveniently, which might be the reason for most participants to spend more time on the mobile phones instead of the traditional TV [20]. The percentage of users and viewing time of mobile phones and computers decreased with increasing age, which was consistent with that reported in previous studies in other countries [6,21]. Although the older population spent lesser time on mobile phones and computers, an increasing trend in using these devices was reported in a Chinese population [7].

The sleep duration in our study population was similar to that reported in previous studies conducted in Macau, Hong Kong, and Mainland China [7,22,23], which was around 7.3-7.8 hours.

XSL•FO
RenderX

Females had poorer sleep quality than males, and they reported that they took a long time to fall asleep and were more likely to experience sleep disturbances. These findings were similar to that reported in a local study conducted among adolescents in 2012 [24] and that reported in a study on older adults in 2016 [8]. We found that sleep latency increased with age, which is consistent with the findings from a large cross-cultural study [25] conducted in China, Ghana, India, Mexico, and Russia, as well as in many other studies [26-28]. A recent study conducted among 0.5 million Chinese individuals suggested that increased sleep problems were associated with the growing use of TV, internet, mobile phones, and social media in the nighttime, which results in reduced sleep duration and increased prevalence of insomnia [22].

Most previous studies suggested an inverse linear relationship between electronic device usage and sleep problems [1]. However, our study found a significant quadratic curvilinear association, which was a "J" shape pattern in all 3 types of electronic device usage. The higher PSQI global score increased towards the end of the longer and shorter durations of electronic device usage. This pattern also appeared in the logistic regression analysis for poor sleep quality, in which people with a moderate duration of usage had the lowest odds to have a poor sleep status. Several mechanisms can be highlighted as the reasons for longer electronic device usage leading to poorer sleep quality. The electronic devices might affect sleep quality directly through reduced sleep duration and persistent electromagnetic radiation [29]. Bright light exposure from the electronic devices might suppress the release of melatonin and disrupt the circadian rhythm [10,29,30]. Long-time use of electronic devices may lead to physical discomforts such as muscular pain and headache, which could result in sleep problems [31]. All these reasons could explain how sleep quality was affected by long-time usage of screen-based electronic devices. Our study also found that moderate use of screen-based electronic devices resulted in the best sleep quality; the J-shape pattern indicating the mild use of these devices was better than that of nonusers or of those who used these devices for very short durations. A study conducted in the United Kingdom suggested that a moderate amount of digital-screen engagement such as watching TV, using a computer, or using a smartphone could help maintain the mental well-being and increase mood regulation and relaxation, which could improve sleep quality [32]. These screen-based electronic devices are widely used in modern society for recreation and relaxation. A study in Belgium showed that 36.7% of the adolescents considered TV viewing as a sleep aid for relaxation [33]. Another study in a Chinese population identified that watching TV and surfing the internet could improve happiness [34]. Some studies indicated that moderate use of smartphones could be considered as a coping strategy that diverts attention from stressful situations, which helps to regulate the individuals' negative emotions [35,36]. Negative emotions and stressful situations can increase sleep fragmentation and decrease sleep duration, thereby resulting in a poor sleep status [37]. Based on our findings and the evidence from these previous studies, we suggest that moderate and appropriate use of screen-based electronic devices could be helpful in emotion regulation and stress relief, which enhances sleep quality thereafter.

Limitations of Our Study

Our study has several limitations. First, the measurement of the electronic device usage did not specify the times, places, and contents watched. Moreover, the usage of other portable or screen-based devices such as video game machines were not evaluated in our study. Further, we did not measure the details of the smartphone usage pattern, for example, the time spent on making traditional phone calls and that spent on using apps based on interactive operations. Nor did we measure the details about when the phone calls were made (day or night), the frequency of the phone calls, and the content of the phone calls. Such information would have allowed for a more comprehensive and insightful investigation into the relationship between electronic device usage and sleep quality. Second, underestimation or overestimation might occur from self-reporting. The participants could tend to report a shorter or longer screening time for social desirability, which might affect the strength of the associations. However, we believe that this would not overturn the nature of the associations that we found because the biases in the responses were random. The large sample size would alleviate the effects of the self-reporting bias on the results.

Future Research and Conclusions

Our study was the first to conduct a comprehensive investigation of the usage of 3 typically used screen-based electronic media devices, namely, TV, computer, and mobile phones and its relationships with sleep quality in a representative Chinese population with a wide range of ages. Most young people use mobile phones frequently in their daily lives, indicating that digital life is an integral part of the modern society. The most interesting finding in our study is that a typical "J" shape association was observed between sleep quality and TV viewing, computer usage, and mobile phone usage. The extreme use of screen-based electronic devices predicted poor sleep status, whereas moderate use was acceptable. These findings can be used to raise awareness among the general public on how to maintain the healthy use of screen-based electronic devices. As a public health issue, making age-specific guidelines for screen-based electronic device usage with regard to quantity and timing should be developed [29]. Future experimental and observational studies that elucidate how the pattern (frequency, duration, contents, location, time of the day, etc) of electronic device usage alters sleep and circadian rhythms across the life course and leads to poor health outcomes are needed. These studies could provide evidence for tailor-made primary prevention strategies in the future.

XSL•FO
RenderX

Acknowledgments

This study was supported by the Macau Galaxy Entertainment Foundation. The authors wish to acknowledge the invaluable contribution of the participants in the study. We would like to specially thank Lee Sara, Lau Kwong Lok, Ng Wai Lok, Sit Chi Hin, Suen Lai Ngan, To Sze Pan, Wong Kin Woon, Yeung Sin Yu Eunice, Yim Cheuk Hin, and students from the Macau Polytechnic Institution and Kiang Wu Nursing College for their support and assistance in data collection, data cleaning, and preliminary data analysis.

Authors' Contributions

YJX conceived and designed the study; YJX, DC, and BLN provided administrative, technical, or material support to the study; AYL, BLN, KML, CL, AYML, and ASMT helped in data collection; YJX analyzed the data and drafted the manuscript; and YJX, AM, DC, AYL, and AYML made critical revisions to the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Average daily hours of the electronic media device usage, the PSQI global score, and poor sleep quality among the participants according to demographics and lifestyle factors.
[DOC File , 101 KB - jmir_v22i7e18095_app1.doc]

Multimedia Appendix 2
Media device usage of the participants (N=1500).
[DOC File , 80 KB - jmir_v22i7e18095_app2.doc]

Multimedia Appendix 3
Seven component scores of sleep quality by gender and age groups.
[DOC File , 40 KB - jmir_v22i7e18095_app3.doc]

References

1. Hysing M, Pallesen S, Stormark KM, Jakobsen R, Lundervold AJ, Sivertsen B. Sleep and use of electronic devices in adolescence: results from a large population-based study. BMJ Open 2015 Feb 02;5(1):e006748 [FREE Full text] [doi: 10.1136/bmjopen-2014-006748] [Medline: 25643702]
2. Simon K. We Are Social. 2018. Global digital report 2018 URL: https://wearesocial.com/blog/2018/01/global-digital-report-2018 [accessed 2020-06-10]
3. LeBourgeois MK, Hale L, Chang A, Akacem LD, Montgomery-Downs HE, Buxton OM. Digital Media and Sleep in Childhood and Adolescence. Pediatrics 2017 Nov;140(Suppl 2):S92-S96 [FREE Full text] [doi: 10.1542/peds.2016-1758J] [Medline: 29093040]
4. Twenge JM, Krizan Z, Hisler G. Decreases in self-reported sleep duration among U.S. adolescents 2009-2015 and association with new media screen time. Sleep Med 2017 Nov;39:47-53. [doi: 10.1016/j.sleep.2017.08.013] [Medline: 29157587]
5. de Souza SCS, Campanini MZ, de Andrade SM, González AD, de Melo JM, Mesas AE. Watching television for more than two hours increases the likelihood of reporting poor sleep quality among Brazilian schoolteachers. Physiol Behav 2017 Oct 01;179:105-109. [doi: 10.1016/j.physbeh.2017.05.029] [Medline: 28576672]
6. Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015 Jun;21:50-58 [FREE Full text] [doi: 10.1016/j.smrv.2014.07.007] [Medline: 25193149]
7. Mak YW, Wu CST, Hui DWS, Lam SP, Tse HY, Yu WY, et al. Association between screen viewing duration and sleep duration, sleep quality, and excessive daytime sleepiness among adolescents in Hong Kong. Int J Environ Res Public Health 2014 Oct 28;11(11):11201-11219 [FREE Full text] [doi: 10.3390/ijerph111111201] [Medline: 25353062]
8. Nogueira BOCL, Li L, Meng L, Ungvari GS, Forester BP, Chiu HFK, et al. Prevalence of Sleep Disturbances and Their Associations With Demographic and Clinical Characteristics and Quality of Life in Older Adults in Macao. Perspect Psychiatr Care 2018 Jan;54(1):46-53. [doi: 10.1111/ppc.12199] [Medline: 27966223]
9. Punamäki RL, Wallenius M, Nygård CH, Saarni L, Rimpelä A. Use of information and communication technology (ICT) and perceived health in adolescence: the role of sleeping habits and waking-time tiredness. J Adolesc 2007 Aug;30(4):569-585. [doi: 10.1016/j.adolescence.2006.07.004] [Medline: 16979753]
10. Weaver E, Gradisar M, Dohnt H, Lovato N, Douglas P. The effect of presleep video-game playing on adolescent sleep. J Clin Sleep Med 2010 Apr 15;6(2):184-189 [FREE Full text] [Medline: 20411697]

XSL•FO
RenderX

11. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
12. Van den Bulck J. Television viewing, computer game playing, and Internet use and self-reported time to bed and time out of bed in secondary-school children. Sleep 2004 Feb 01;27(1):101-104. [doi: 10.1093/sleep/27.1.101] [Medline: 14998244]
13. Basner M, Spaeth AM, Dinges DF. Sociodemographic characteristics and waking activities and their role in the timing and duration of sleep. Sleep 2014 Dec 01;37(12):1889-1906 [FREE Full text] [doi: 10.5665/sleep.4238] [Medline: 25325472]
14. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
15. Nuutinen T, Roos E, Ray C, Villberg J, Välimaa R, Rasmussen M, et al. Computer use, sleep duration and health symptoms: a cross-sectional study of 15-year olds in three countries. Int J Public Health 2014 Aug;59(4):619-628. [doi: 10.1007/s00038-014-0561-y] [Medline: 24848704]
16. The Statistics and Census Service. The Statistics and Census Service, Government of Macao Special Administrative Region. 2017 May 05. Detailed Results of 2016 Population By-Census URL: https://www.dsec.gov.mo/Statistic/Demographic/GlobalResultsOfBy-Census/2016%E4%B8%AD%E6%9C%9F%E4%BA%BA%E5%8F%A3%E7%B5%B1%E8%A8%88.aspx?lang=en-US [accessed 2020-06-10]
17. Tsai P, Wang S, Wang M, Su C, Yang T, Huang C, et al. Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CPSQI) in primary insomnia and control subjects. Qual Life Res 2005 Oct;14(8):1943-1952. [doi: 10.1007/s11136-005-4346-x] [Medline: 16155782]
18. Buysse DJ, Reynolds CF, Monk TH, Hoch CC, Yeager AL, Kupfer DJ. Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh Sleep Quality Index (PSQI). Sleep 1991 Aug;14(4):331-338. [Medline: 1947597]
19. World Health Organization. Regional Office for the Western Pacific, International Diabetes Institute, International Association for the Study of Obesity, International Obesity Task Force. The Asia-Pacific perspective : redefining obesity and its treatment. Sydney: Health Communications Australia; 2000.
20. Liebowitz SJ, Zentner A. Clash of the Titans: Does Internet use Reduce Television Viewing? Review of Economics and Statistics 2012 Feb;94(1):234-245. [doi: 10.1162/rest_a_00148]
21. van Deursen AJ, Bolle CL, Hegner SM, Kommers PA. Modeling habitual and addictive smartphone behavior. Computers in Human Behavior 2015 Apr;45:411-420. [doi: 10.1016/j.chb.2014.12.039]
22. Chen Y, Kartsonaki C, Clarke R, Guo Y, Yu C, Bian Z, China Kadoorie Biobank Study (members listed at end of report). Characteristics and correlates of sleep duration, daytime napping, snoring and insomnia symptoms among 0.5 million Chinese men and women. Sleep Med 2018 Apr;44:67-75 [FREE Full text] [doi: 10.1016/j.sleep.2017.11.1131] [Medline: 29530372]
23. Pan J, Chou M, Zhang J, Liu Y. Sleep patterns, insomnia, and daytime sleepiness between Guangdong and Macau Chinese adolescents: a cross-cultural comparison study. Biological Rhythm Research 2012 Oct;43(5):527-539. [doi: 10.1080/09291016.2011.614791]
24. IP CH, Pan JY, Liu YP. Sleep quality, sleep pattern and self-rated depression and anxiety among junior high school students in Macau. Chinese Mental Health Journal 2012;26(10):761-765 [FREE Full text]
25. Endeshaw Y. Self-Reported Sleep Problems Across the Ages-An Intercontinental Study. J Gerontol Geriatric Res 2012;01(05):1-17. [doi: 10.4172/2167-7182.1000112]
26. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004 Nov 01;27(7):1255-1273. [doi: 10.1093/sleep/27.7.1255] [Medline: 15586779]
27. Gaudreau H, Carrier J, Montplaisir J. Age-related modifications of NREM sleep EEG: from childhood to middle age. J Sleep Res 2001 Sep;10(3):165-172 [FREE Full text] [doi: 10.1046/j.1365-2869.2001.00252.x] [Medline: 11696069]
28. Lauer CJ, Riemann D, Wiegand M, Berger M. From early to late adulthood. Changes in EEG sleep of depressed patients and healthy volunteers. Biol Psychiatry 1991 May 15;29(10):979-993. [doi: 10.1016/0006-3223(91)90355-p] [Medline: 2065140]
29. Cain N, Gradisar M. Electronic media use and sleep in school-aged children and adolescents: A review. Sleep Med 2010 Sep;11(8):735-742. [doi: 10.1016/j.sleep.2010.02.006] [Medline: 20673649]
30. Khalsa SBS, Jewett ME, Cajochen C, Czeisler CA. A phase response curve to single bright light pulses in human subjects. J Physiol 2003 Jun 15;549(Pt 3):945-952 [FREE Full text] [doi: 10.1113/jphysiol.2003.040477] [Medline: 12717008]
31. Thomée S, Dellve L, Härenstam A, Hagberg M. Perceived connections between information and communication technology use and mental symptoms among young adults - a qualitative study. BMC Public Health 2010 Feb 12;10:66 [FREE Full text] [doi: 10.1186/1471-2458-10-66] [Medline: 20152023]
32. Przybylski AK, Weinstein N. A Large-Scale Test of the Goldilocks Hypothesis. Psychol Sci 2017 Feb;28(2):204-215. [doi: 10.1177/0956797616678438] [Medline: 28085574]

XSL•FO
RenderX

33. Eggermont S, Van den Bulck J. Nodding off or switching off? The use of popular media as a sleep aid in secondary-school children. J Paediatr Child Health 2006;42(7-8):428-433. [doi: 10.1111/j.1440-1754.2006.00892.x] [Medline: 16898880]
34. Wei X, Huang S, Stodolska M, Yu Y. Leisure Time, Leisure Activities, and Happiness in China. Journal of Leisure Research 2017 Dec 13;47(5):556-576. [doi: 10.18666/jlr-2015-v47-i5-6120]
35. Hoffner CA, Lee S. Mobile Phone Use, Emotion Regulation, and Well-Being. Cyberpsychol Behav Soc Netw 2015 Jul;18(7):411-416. [doi: 10.1089/cyber.2014.0487] [Medline: 26167841]
36. Toda M, Ezoe S, Tatsuya T. Mobile Phone Use and Stress-Coping Strategies of Medical Students. International Journal of Cyber Behavior, Psychology and Learning 2014 Oct;4(4):41-46 [FREE Full text] [doi: 10.4018/ijcbpl.2014100104]
37. Vandekerckhove M, Cluydts R. The emotional brain and sleep: an intimate relationship. Sleep Med Rev 2010 Aug;14(4):219-226. [doi: 10.1016/j.smrv.2010.01.002] [Medline: 20363166]

Abbreviations

ANOVA: analysis of variance
BMI: body mass index
HLLL: healthy life, longer lives
OR: odds ratio
PSQI: Pittsburgh Sleep Quality Index
TV: television
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.02.20; peer-reviewed by H Wang, H Yu; comments to author 03.03.20; revised version received 19.04.20; accepted 27.04.20; published 07.07.20.

Please cite as:
Xie YJ, Cheung DSK, Loke AY, Nogueira BL, Liu KM, Leung AYM, Tsang ASM, Leong CSU, Molassiotis A
Relationships Between the Usage of Televisions, Computers, and Mobile Phones and the Quality of Sleep in a Chinese Population: Community-Based Cross-Sectional Study
J Med Internet Res 2020;22(7):e18095
URL: https://www.jmir.org/2020/7/e18095
doi:10.2196/18095
PMID:32369439

©Yao Jie Xie, Daphne SK Cheung, Alice Y Loke, Bernice L Nogueira, Karry M Liu, Angela YM Leung, Alice SM Tsang, Cindy SU Leong, Alex Molassiotis. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Relationship Between Internet Use and Mental Health Among Older Adults in England: Longitudinal Observational Study

Sabrina Sze Man Lam[1], MSc; Stephen Jivraj[1], PhD; Shaun Scholes[1], PhD

University College London, London, United Kingdom

Corresponding Author:
Stephen Jivraj, PhD
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 20 7679 1721
Email: stephen.jivraj@gmail.com

Abstract

Background: There is uncertainty about the impact of internet use on mental health in older adults. Moreover, there is very little known specifically about the impact of particular purposes of internet use.

Objective: This study aims to investigate the longitudinal relationship between two distinct concepts of mental health with the frequency of internet use among older adults: the moderating role of socioeconomic position (SEP) and the association between specific purposes of internet use.

Methods: Longitudinal fixed and random effects (27,507 person-years) models were fitted using waves 6-8 of the English Longitudinal Study of Ageing to examine the relationship between different aspects of internet use (frequency and purpose) and two mental health outcomes (depression and life satisfaction). The potential moderating effect of SEP on these associations was tested using interaction terms.

Results: Infrequent internet use (monthly or less vs daily) was predictive of deteriorating life satisfaction (β=−0.512; *P*=.02) but not depression. Education and occupational class had a moderating effect on the association between frequency of internet use and mental health. The associations were stronger in the highest educational group in both depression (*P*=.09) and life satisfaction (*P*=.02), and in the highest occupational group in life satisfaction (*P*=.05) only. Using the internet for communication was associated with lower depression (β=−0.24; *P*=.002) and better life satisfaction (β=.97; *P*<.001), whereas those using the internet for information access had worse life satisfaction (β=−0.86; *P*<.001) compared with those who did not.

Conclusions: Policies to improve mental health in older adults should encourage internet use, especially as a tool to aid communication.

(J Med Internet Res 2020;22(7):e15683) doi:10.2196/15683

KEYWORDS

internet; socioeconomic factors; mental health; life satisfaction; depression; effect modifier

Introduction

As population aging brings major challenges for society and health care services, promoting healthy aging is important to prevent the burden and financial costs related to aging-associated diseases. Mental health is a crucial component of healthy aging. Poor mental health can profoundly affect the quality of life of older adults and is associated with an increased risk of mortality and morbidity [1,2].

Risk factors for poor mental health in older adults could be potentially impacted by internet use; maintaining and building social relationships through email or social networking sites may increase feelings of social support and connectedness and relieve feelings of isolation and loneliness [3-7]. The internet is used at least once every 3 months by an increasing proportion of the English population aged 50 years and older (63% and 51% of men and women, respectively, in 2010 compared with 84% and 78% of men and women, respectively, in 2016) [8].

XSL•FO
RenderX

However, internet use remains lower in older adults than in younger adults [9].

Evidence is mixed on the impact of internet use on mental health in older adults. Forsman and Nordmyr [10] reported a balanced positive association in a review of 18 quantitative and 14 qualitative studies but found that 40% of the outcomes they reviewed showed no association. Claims for positive impacts according to some researchers are often exaggerated because of training effects, inappropriate generalization of results, and misattribution of causality [11]. Training effects and limited generalizability are typically leveled in experimental studies where a small intervention group of selective participants is given support such as training to use the internet [12-14]. However, recent evidence suggests that internet use can improve mental health. Using data from the US Health and Retirement Study (HRS) over a 6-year period, Cotten et al [15] found that prior internet usage reduced the probability of depression by one-third between 2002 and 2008 in a sample of 3000 retired adults aged 50 years and older. Using data over a 10-year period, Lissitsa and Chachashvili-Bolotin [16] found that internet adoption among an Israeli sample of adults aged 65 years and older was associated with increased life satisfaction, with the strength of the estimated association being as strong as that found for being married.

To our knowledge, no literature has longitudinally examined the associations between internet usage and multiple subdimensions of mental health in the same sample. The dimensions of mental health include hedonic and eudemonic well-being [17]. According to Vanhoutte [18], hedonic well-being refers to pleasure maximization and suffering minimization, with the cognitive and affective aspects of well-being forming subdimensions of hedonic well-being. Cognitive well-being is an evaluative process in which an individual assesses the quality of their life; in contrast, affective well-being refers to positive and negative moods and emotions that are not simply the opposite of one another [18,19]. Eudemonic well-being is conceptually distinctive in that it refers to a sense of purpose or control [20]. Separating these dimensions is important (through, eg, separate measurement scales for life satisfaction, depression, and quality of life) because they capture varied aspects of mental health; these could, theoretically, be influenced by internet use in different ways. For example, Matthews and Nazroo [21] explored the cross-sectional relationship between internet use and several dimensions of mental health using data from persons aged 50 years and older in the 2012-2013 English Longitudinal Study of Ageing (ELSA). Frequent internet usage was found to be associated with lower levels of depression (affective well-being) and better quality of life (eudemonic well-being), but no association was found with life satisfaction (hedonic well-being) [21]. In contrast, Heo et al [22] found that more frequent internet use was associated with better life satisfaction and greater psychological well-being among HRS participants aged 65 years and older.

The current literature is limited on the purposes of internet use which have a positive (or negative) influence on mental health in older adults. The internet has myriad uses, for example, using the internet for communication purposes may improve mental health among older adults because it enables them to stay in touch with family and friends who previous generations may have lost contact with [23,24]. On the other hand, certain types of information access may undermine mental health. For example, using the internet for job searching was associated with more depressive symptoms, lower quality of life, and lower life satisfaction among ELSA participants [21]. Using the internet for finding new people and for entertainment was associated with decreased well-being among older Australians [25].

A possible confounder or moderator of the internet use and mental health association is socioeconomic position (SEP). Internet use is more prevalent in higher SEP groups [26-28], with use for information access and for entertainment being higher for persons with higher and lower SEP, respectively. SEP could also moderate this association; however, this has not been investigated in the literature to date. It could be the case that in groups with the highest levels of internet use (eg, the wealthiest and better educated), nonusers are at a greater disadvantage in terms of their mental health.

This study aimed to further the debate by using a longitudinal sample, analyzing the associations between internet usage and the hedonic dimension of well-being (life satisfaction) and the affective dimension of wellbeing (depression) over a 4-year period, analyzing different purposes of internet usage (eg, communication, entertainment, and information access), and examining whether SEP moderates these relationships. This study hypothesized that (1) more frequent internet use predicts better mental health longitudinally; (2) the association is moderated by SEP, being stronger in the higher SEP groups; and (3) mental health is most strongly associated with internet use for the purposes of communication (positively) and information access (negatively). Our findings will have implications for understanding the importance of internet use and its impact on mental health, which can inform policies to encourage wider and more frequent internet use in older adults to improve mental health.

Methods

Data and Sample

The ELSA is a nationally representative sample of the English population aged 50 years and older. Established in 2002, the sample has been followed up every 2 years [29]. Data were collected through computer-assisted personal interviews, self-completion questionnaires, and nurse assessments (every 4 years) [29]. The original sample was selected from the Health Survey for England (1998, 1999, and 2001) [30]. This analysis used data from waves 6 (2012-2013), 7 (2014-2015), and 8 (2016-2017). The analytical sample (n=9169 persons aged 50 years and older at wave 6, with 27,507 person-years of follow-up between waves 6 and 8) comprised sample members first recruited at waves 1 (n=5659), 3 (n=888), 4 (n=1796), and 6 (n=826). ELSA participants provided signed consent, and the London Multicentre Research Ethics Committee (MREC 01/2/91) granted ethical approval for the study.

XSL•FO
RenderX

Time-Varying Outcomes

As an aspect of the affective dimension of hedonic well-being, mental health was measured using the validated 8-item version of the Center for Epidemiologic Studies Depression Scale (CES-D) [31]. Items were answered using a dichotomous response (*yes* or *no*), which were summed to result in a total score ranging from 0 to 8, where a score of 0 indicates no depressive symptoms. As an aspect of the cognitive dimension of hedonic well-being, life satisfaction was measured using the Satisfaction with Life Scale [32,33], which has been explicitly tested on older persons and consists of 5 items. The individual item responses ranged from *strongly agree* to *strongly disagree* on a 7-point Likert scale, which was summed to result in a total score ranging from 5 (lowest satisfaction) to 35 (highest satisfaction).

Time-Varying Exposure

Data on the frequency of internet use were collected through self-completion questionnaires. The responses were recoded as *daily*, *weekly*, *monthly or less*, or *never*.

Time-Constant Exposure

The purposes of internet use were assessed only for participants who reported using the internet at least once every 3 months. For this study, 12 purposes of internet use at wave 6 were grouped into 6 as follows: (1) communication (sending or receiving emails; use of social networking sites; and creating, uploading, or sharing content), (2) entertainment (news, newspaper or blog websites, streaming or downloading live or on demand television or radio, music, electronic books, and games), (3) information access (finding information about goods and services; searching for information for learning, research, and fact finding; and looking for a job or sending a job application), (4) managing finances, (5) electronic commerce (shopping and buying goods or services over the internet or selling goods or services over the internet), and (6) other. The questions on the purposes of internet use changed in waves 6, 7, and 8; therefore, we could not use it as a time-varying variable in the modeling. Our analyses focused on the 5 categories mentioned earlier.

Time-Varying Covariates

Participants' age was measured in a single year (coded to 99 to indicate those aged 91 years or older, whose individual age was not released). We tested nonlinear age and mental health associations using second-order polynomial terms. Couple status was measured according to whether a participant was cohabiting (married or unmarried) or not. Working status (currently working) was derived from economic activity status by dichotomizing those who were employees or self-employed from those who were seeking work, sick and not seeking work, retired, or unoccupied. Health status (presence of a limiting long-term illness) was measured through whether a participant self-reported an illness or disability that limited their activities in any way.

Time-Invariant Covariates

SEP at wave 6 was measured using 3 different domains [34]: occupational class, wealth, and education. On the basis of degree of control and autonomy, occupational class (based on the household reference person's last occupation) was measured using the 3-category version of the National Statistics Socio-Economic Classification: managerial and professional, intermediate, and routine and manual occupations [35]. Wealth was measured using quintiles of total net nonpension household wealth, a summary measure of the value of financial, physical, and housing wealth minus any debt [35]. It represents the permanent economic status of older adults [36], as most participants in aging cohort studies are retired. The highest educational qualification was grouped into 3 categories: degree level, below degree level, and no qualifications. Education among the ELSA cohort members is generally determined earlier in the life course, whereas wealth and occupational class are more likely to change during mid to later life.

Statistical Analysis

To examine the longitudinal association between internet use and mental health, we used a series of fixed and random effects linear models. A nonlinear model for depression did not satisfy the condition of the fixed effects model, and the results from the random effects logistic regression model were substantively similar (not shown here). Therefore, the CES-D score was treated as a continuous dependent variable in both the fixed and random effects models. The fixed effects models control for all measured and unmeasured time-invariant characteristics by using each participant as their own control over survey waves. We adjusted for time-varying age, age squared, working status, couple status, and health status to address our first hypothesis.

H1: More frequent internet use predicts better mental health longitudinally.

Internet use was entered into the models as a four-category independent variable, with daily use as the reference. Random effects models do not control for the correlation between measured and unmeasured characteristics and, therefore, analyze both within- and between-participant variance. We adjusted for the same time-varying variables as for the fixed effects model and additionally adjusted for time-invariant SEP and sex to address our second hypothesis.

H2: SEP moderates the association between internet use and mental health.

An assessment of moderation was performed by including an interaction term between the frequency of internet usage and each marker of SEP; these were retained if statistically significant (likelihood ratio test, $P<.10$). Separate random effect models tested the association between the purpose of internet use and mental health to address our final hypothesis.

H3: There is a positive association between communication internet use and mental health and a negative association between information internet use and mental health.

The analysis not presented here showed similar results when limiting the analytical sample to those who used the internet compared with the results for all participants presented here.

To address attrition from wave 6 onward and item nonresponse within a survey wave, we fitted multiple imputation models by chained equations on all variables (outcomes, exposures, and

XSL•FO
RenderX

covariates) with missing values. The purpose of internet use was imputed conditionally on whether a participant reported internet usage at wave 6. All statistical analyses were conducted using cross-sectional survey weights to minimize bias arising from sample attrition from the beginning of the ELSA study to participation in wave 6. The use of imputation and survey weights did not substantively alter the main findings.

Results

Descriptive Results

Table 1 presents the means, SDs, or percentages across person-waves for the key variables by frequency of internet usage. Overall, 56% of older adults used the internet daily, 12% weekly, 6% monthly or less, and 26% had never used the internet. The most common purpose was information access (68%), followed by communication (66%), electronic commerce (52%), entertainment (45%), and managing finances (39%). Those who used the internet more frequently had better life satisfaction scores; lower depression scores; and were more likely to be younger, currently working, be in a couple, have no limiting long-term illness, male, have a degree, be in managerial and professional occupations, and be in the wealthiest quintile compared with those who used the internet less frequently or not at all. For example, the mean age of a daily internet user was 66 (SD 7.64) years, whereas the mean age of a participant who had never used the internet was 75 (SD 9.30) years. Overall, 49% of daily users were male, compared with 38% of those who had never used the internet.

XSL•FO
RenderX

Table 1. Participant characteristics across person-waves by frequency of internet use (data are unweighted).

Measures	Daily	Weekly	Monthly	Never	Total
Life satisfaction					
Mean (SD)[a]	26.12 (6.10)	25.10 (6.34)	24.44 (6.63)	24.72 (6.65)	25.56 (6.34)
Participants, N	11,033	2294	1164	4847	19,338
Depression					
Mean (SD)[a]	1.02 (1.61)	1.31 (1.77)	1.52 (1.95)	1.76 (2.04)	1.28 (1.80)
Participants, N	11,136	2373	1203	5290	20,002
Age (years)					
Mean (SD)[a]	65.81 (7.64)	68.66 (8.43)	69.64 (8.83)	74.86 (9.30)	68.78 (9.12)
Participants, N	11,233	2388	1218	5349	20,188
Currently working					
Participants, n (%)[a]	4304 (38.32)	629 (26.34)	281 (23.07)	519 (9.70)	5733 (28.40)
Participants, N	11,233	2388	1218	5349	20,188
In a couple					
Participants, n (%)[a]	8599 (76.55)	1686 (70.60)	846 (69.46)	2897 (54.16)	14,028 (69.49)
Participants, N	11,233	2388	1218	5349	20,188
Limiting long-term illness					
Participants, n (%)[a]	3058 (27.22)	832 (34.84)	484 (39.74)	2558 (47.82)	6932 (34.34)
Participants, N	11,232	2387	1218	5347	20,184
Internet purpose[b]					
Entertainment, n (%)	66.0	33.9	16.8	0	44.7
Communication, n (%)	91.3	69.4	37.8	0	65.6
Information access, n (%)	92.5	79.0	50.5	0	68.4
Electronic commerce, n (%)	76.5	40.2	19.8	0	52.0
Finances, n (%)	59.3	24.3	8.0	0	39.0
Participants, N	10,627	2093	780	5349	18,849
Sex[b]					
Female, n (%)	5772 (51.38)	1406 (58.88)	770 (63.22)	3314 (61.96)	11,262 (55.79)
Participants, N	11,233	2388	1218	5349	20,188
Education[b]					
Degree qualification, n (%)	3068 (27.31)	317 (13.27)	122 (10.02)	221 (4.13)	3728 (18.56)
Below degree qualification, n (%)	6936 (61.75)	1590 (66.58)	809 (66.42)	2557 (47.80)	11,892 (59.19)
No qualifications, n (%)	1183 (10.53)	469 (19.64)	285 (23.40)	2533 (47.35)	4470 (22.25)
Participants, N	11,187	2376	1216	5311	20,090
Occupational class[b]					
Managerial and professional, n (%)	5417 (48.22)	719 (30.11)	353 (28.98)	919 (17.18)	7408 (36.87)
Intermediate, n (%)	2979 (26.52)	651 (27.26)	316 (25.94)	1277 (23.87)	5223 (26.00)
Routine and manual, n (%)	2799 (24.92)	994 (41.62)	544 (44.66)	3138 (58.67)	7475 (37.21)
Participants, N	11,195	2364	1213	5334	20,106
Wealth quintile[b]					
Least affluent, n (%)	1046 (9.31)	302 (12.65)	194 (15.93)	1367 (25.56)	2909 (14.69)

XSL•FO
RenderX

Measures	Daily	Weekly	Monthly	Never	Total
2, n (%)	1565 (13.93)	478 (20.02)	221 (18.14)	1272 (23.78)	3536 (17.86)
3, n (%)	2063 (18.37	528 (22.11)	291 (23.89)	1329 (24.85)	4211 (21.27)
4, n (%)	2809 (25.01)	581 (24.33)	284 (23.32)	830 (15.52)	4504 (22.75)
Most affluent, n (%)	3484 (31.02)	461 (19.30)	206 (16.91)	490 (9.16)	4641 (23.44)
Participants, N	10,967	2350	1196	5288	19,801

[a]Descriptive statistics for time-varying variables calculated using data from waves 6-8.

[b]Descriptive statistics for time-invariant variables calculated at wave 6 only.

Frequency of Internet Use and Mental Health (Fixed Effects Models for Within-Person Change)

Table 2 presents the fixed effects panel modeling estimates for the frequency of internet usage on depression and life satisfaction. The model estimates from the full models are shown in Multimedia Appendix 1. The coefficients represent estimates of change within participants between waves 6 and 8, and therefore positive estimates represent an increase in the depression score (a poorer health outcome) and an increase in the life satisfaction score (a positive health outcome). The frequency of internet use was not predictive of within-person change in depression when taking into account time-varying age, working, being in a couple, and health status. Daily internet usage did predict better life satisfaction over time compared with using the internet on, at most, a monthly basis when controlling for the same time-varying variables (β for monthly or less=−0.512; *P*=.02). On the basis of model estimates, Figure 1 shows that those using the internet daily were predicted to have more than a half-point increase in life satisfaction scores between waves 6 and 8. There was no significant within-person change in life satisfaction for weekly or never users (vs daily users).

Table 2. Fixed effects model coefficients for frequency of internet use on mental health outcomes.

Frequency of internet use[a]	Depression[b]		Life satisfaction[c]	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Daily	Reference	N/A[d]	Reference	N/A
Weekly	0.030 (−0.822 to 0.143)	.59	−0.230 (−0.495 to 0.034)	.09
Monthly or less	0.111 (−0.060 to 0.281)	.20	−0.512 (−0.956 to −0.067)	.02
Never	0.096 (−0.101 to 0.293)	.33	−0.472 (−0.955 to 0.012)	.06

[a]Models adjusted for time-varying age, age squared, working status, couple status, and health status.

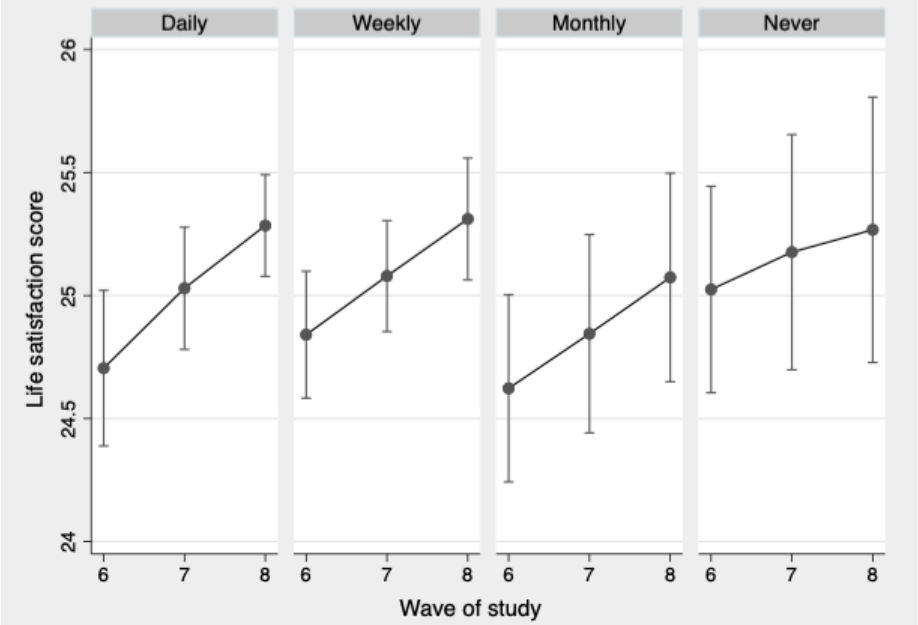
[b]Higher scores represent deteriorating depression within participants.

[c]Higher scores represent improving life satisfaction within participants.

[d]Not applicable.

XSL•FO
RenderX

Figure 1. Predicted life satisfaction score by wave of study and frequency of internet usage from fixed effects model. Model adjusted for time-varying frequency of internet use, age, age square, working status, couple status, and health status.

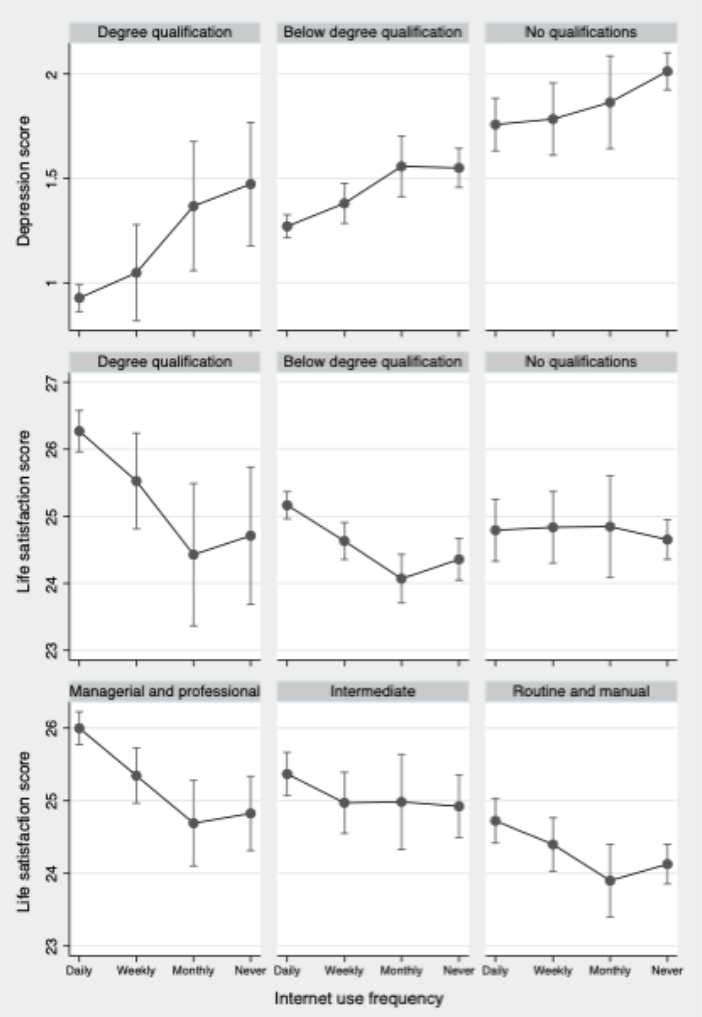


Socioeconomic Position as a Moderator of the Association Between Internet Use and Mental Health (Random Effects Models)

Figure 2 graphically illustrates the predictions from the random effects modeling, which showed that education and occupational class significantly moderated the longitudinal association between internet use frequency and mental health (likelihood ratio test $P<.10$ for interaction terms). The model estimates from the full models are presented in Multimedia Appendix 2. The differences in the predicted scores of depression ($P=.09$) and life satisfaction ($P=.02$) by internet use were steeper for those with a degree. For example, the predicted life satisfaction score for those who had no formal educational qualifications was almost 25 irrespective of internet use frequency, whereas the predicted score for those with a degree was over 1 unit higher for daily users compared with monthly or never users. Similarly, the models suggested that occupational class moderated the association between internet use frequency and life satisfaction ($P=.05$) but not depression. There was no evidence to suggest that wealth moderated the associations.

XSL•FO
RenderX

Figure 2. Predicted mental health score by socioeconomic position and frequency of internet usage from random effects model. Model adjusted for time-varying frequency of internet use, age, age square, working status, couple status, and health status and for time-invariant education, occupational class, wealth, and sex, including interaction terms for education and frequency of internet use and occupational class and frequency of internet use.



Association Between Purpose of Internet Use and Mental Health (Random Effects Models)

Table 3 presents the estimates from the random effects model for mutually adjusted purposes of internet usage on mental health (full estimates are presented in Multimedia Appendix 3). Compared with those not using the internet for this stated purpose, using the internet for communication purposes was associated with a lower depression score between waves 6 and 8 (β=−0.24; *P*=.002) and a higher life satisfaction score (β=.97; *P*<.001) when controlling for time-varying and time-invariant variables. In contrast, using the internet for information access was associated with a worse life satisfaction score (β=−0.86; *P*=.001).

XSL•FO
RenderX

Table 3. Random effects model coefficients for the purpose of internet use on mental health outcomes.

Purpose of internet use[a,b]	Depression (95% CI)[c]		Life satisfaction (95% CI)[d]	
	Coefficient (95% CI)	*P* value	Log odds (95% CI)	*P* value
Entertainment	0.047 (−0.027 to 0.121)	.21	0.093 (−0.023 to 0.412)	.56
Communication	−0.242 (−0.391 to −0.092)	.002	0.968 (0.509-1.428)[e]	<.001
Information access	−0.034 (−1.344 to 0.105)	.62	−0.858 (−1.344 to −0.372)[e]	.001
Electronic commerce	−0.098 (−0.199 to 0.004)	.06	0.322 (−0.035 to 0.678)	.08
Finance	−0.073 (−0.153 to 0.007)	.07	0.210 (−0.110 to 0.531)	.19

[a]Models adjust for time-varying age, age squared, working status, couple status, and health status and for time-invariant socioeconomic position and sex.

[b]Reference category was not using the internet for each purpose.

[c]Higher scores represent deteriorating depression.

[d]Higher scores represent improving life satisfaction.

Analysis of more specific purposes (Multimedia Appendix 4) showed that using the internet for email was the only communication purpose associated with better mental health longitudinally and that using the internet for job searching was the only information access purpose associated with deteriorating mental health. Adjustment for more detailed working status categories did not change our findings, suggesting that using the internet for job searching was associated with lower life satisfaction longitudinally over and above being unemployed or out of work.

Discussion

Principal Findings

This study partially confirmed our first hypothesis (H1) that more frequent internet use improves mental health. After adjustment for time-varying confounding factors, there was a positive longitudinal effect of using the internet daily compared with monthly or less on life satisfaction. Internet use frequency was not associated longitudinally with depression. The second hypothesis (H2) that SEP moderates the association between internet usage frequency and mental health was confirmed. This association was stronger for those with an educational degree. There was also a stronger association in those in managerial occupations and professional occupations in life satisfaction. We confirmed our third hypothesis (H3) that using the internet for communication purposes (specifically use of email) was associated with better mental health and that using the internet for information access (specifically for job searching) was associated with worse mental health.

Existing studies have found mixed results regarding the association between internet use and mental health. Notwithstanding methodological differences, our findings are consistent with studies that found that frequent internet use can have benefits for older adults' depression and life satisfaction [13,15,21,34]. The discrepancy between the findings could be because of methodological issues, such as measurement of mental health [18], and varying sample characteristics, sizes, and context [37]. Very few previous studies have been able to analyze internet use frequency and its relationship with mental health longitudinally. Our findings are consistent with the limited literature available [15,16].

This is the first study to our knowledge that explores SEP as a potential moderator of longitudinal associations between internet use and mental health among older adults. Several studies draw conclusions from sample sizes of fewer than 300 respondents and are not generalizable to middle- and older-aged adults. In contrast, the results of this study come from a large and nationally representative sample of persons aged 50 years and older. Furthermore, the range of questions on internet use allowed us to extend knowledge on the impact of different purposes of internet use on mental health.

Evidence from this study suggests that markers of SEP moderate the association between internet use and mental health; frequent usage in the highest educational and occupational groups had a larger positive impact on mental health than frequent use among those in the lowest groups. This might suggest that low usage in higher SEP groups through digital exclusion, for example, may be particularly problematic for mental health. Our findings, however, need to be viewed in light of our mental health outcomes. The measure of mental health used in our study is determined solely on the affective (depression) and cognitive (life satisfaction) aspects of hedonic well-being, which may not be fully indicative of overall mental health, which also includes eudemonic aspects such as a sense of purpose or control. We cannot rule out the possibility that internet use fulfills eudemonic needs among older adults in disadvantaged SEP groups. Future studies should use a broader measure of mental health to assess whether the longitudinal association between internet use and mental health varies significantly across SEP groups. Further work could also explore whether the relationship between internet use frequency and mental health is moderated by SEP in younger samples in whom almost all use the internet. One might posit that our finding is a cohort effect because later-born generations are expected to enter older age with a much higher prevalence of internet usage.

Our findings highlight heterogeneity among different purposes of internet use. Using the internet for communication through email was associated with better mental health. This provides empirical support for the importance of communication for

XSL•FO
RenderX

positive mental health, as it allows individuals to maintain strong social ties and contacts, thereby reducing the negative impacts of social isolation and loneliness on mental health [7,15,38,39]. In contrast, using the internet for job searching was associated with higher depression and worse life satisfaction over time. This is comparable with previous analyses of ELSA data. This finding may indicate dissatisfaction with one's current job or economic activity status rather than a direct effect of using the internet for this purpose [21]. Our results emphasize the importance of disaggregating different internet usage purposes, as internet use among older adults is not homogenous [40] and impacts different aspects of mental health in different ways.

Policy Recommendations

Our findings point to clear policy implications, the most pressing of which is the need for interventions to encourage older adults to use the internet, so as to build and maintain strong social ties, thereby avoiding the damaging impacts of social isolation and loneliness on mental health. Strategies to promote and increase internet usage among older adults could include setting up public Wi-Fi in places frequently visited by older people and where usage might be low because of barriers to access and digital exclusion. Design of policy interventions to increase internet usage should consider the fact that nonusers of the internet in older age are not homogenous entities but are heterogeneous with respect to gender, age, SEP, household composition, and attitudes toward the internet [41]. It is also important to recognize that for many older people, intrinsic training in how to use the internet did not take place during their formal education or working life, and therefore, the barriers to access can be much more than about availability. Therefore, modifying any psychological barriers to active engagement with the internet is also important. More specifically, to promote internet use for communication, public policy makers should encourage private providers of social networking tools to create age-friendly interfaces. Hardware also needs to be appropriate for novice users and users with impairments such as poor eyesight. This requires manufacturers to be sensitive to these needs when designing products and making the accessibility settings easy to navigate to change the text size, for example.

Limitations

Several limitations of this study need to be acknowledged. The 8-item CES-D administered in ELSA does not measure depression, rather symptomology, and its association with internet usage may be confounded by omitted variables not included or imperfectly measured in this study. These might include the diagnosis of mental health problems and access to mental health services. Further research should theorize the possible pathways between internet usage and mental health to minimize potential confounding. The measurement of our main exposure is also imperfect, as we were unable to disaggregate daily use of the internet (almost half of the sample) by the number of hours spent on the internet on a typical day. Evidence from other studies suggests that spending an excessive amount of time on the internet correlates with negative mental health in younger age samples [42]. We are not aware of any evidence of the relationship between very frequent internet usage and mental health in older age. A further limitation is the limited 4 years of follow-up. More data points are necessary to explore in detail how any change in the frequency and purposes for internet usage impact mental health trajectories.

A final limitation of our study is the potential bias in our estimates because of the selective nature of participation in aging cohort studies, such as ELSA. Our study sample included participants from the refreshment samples added at waves 3, 4, and 6 to ensure a nationally representative sample of persons aged 50 years and older living in private households at the time of wave 6 (2012-2013). Although our analyses used the appropriate weights available for use with the ELSA study and used multiple imputed data to reduce (wave and item) nonresponse bias, the healthier nature of surviving participants could have inevitably led to an underestimation of the internet use and mental health associations.

Conclusions

In summary, internet use can serve as a channel for improving the mental health of older adults. With population aging and increases in the prevalence of poor mental health, promotion of (affordable) internet use and minimization of psychological barriers to digital engagement could potentially help to promote healthy living in older age.

Acknowledgments

This work was supported by the UCL Q-Step Centre funded by Nuffield Foundation, Economic and Social Research Council, and Higher Education Funding Council for England Q-Step programme.

Authors' Contributions

All authors contributed to the development, analysis, and writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full fixed effects models results.
[DOCX File , 19 KB - jmir_v22i7e15683_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Full random effects models results for education by internet usage results.
[DOCX File , 18 KB - jmir_v22i7e15683_app2.docx]

Multimedia Appendix 3
Full random effects models results for broad purpose of use.
[DOCX File , 16 KB - jmir_v22i7e15683_app3.docx]

Multimedia Appendix 4
Random effects model fixed effect coefficients for detailed purpose of internet use on mental health.
[DOCX File , 14 KB - jmir_v22i7e15683_app4.docx]

References

1. Rodda J, Walker Z, Carter J. Depression in older adults. Br Med J 2011 Sep 28;343:d5219. [doi: 10.1136/bmj.d5219] [Medline: 21957206]
2. Blazer DG. Depression in late life: review and commentary. J Gerontol A Biol Sci Med Sci 2003 Mar;58(3):249-265. [doi: 10.1093/gerona/58.3.m249] [Medline: 12634292]
3. Cotten SR, Goldner M, Hale TM, Drentea P. The importance of type, amount, and timing of internet use for understanding psychological distress. Soc Sci Q 2011;92(1):119-139. [doi: 10.1111/j.1540-6237.2011.00760.x] [Medline: 21534270]
4. Mellor D, Firth L, Moore K. Can the internet improve the well-being of the elderly? Ageing Int 2008 Feb 9;32(1):25-42. [doi: 10.1007/s12126-008-9006-3]
5. Sum S, Mathews MR, Pourghasem M, Hughes I. Internet technology and social capital:how the internet affects seniors' social capital and wellbeing. J Comput Mediat Commun Indexing 2008;14(1):202-220. [doi: 10.1111/j.1083-6101.2008.01437.x]
6. Tsai H, Tsai Y, Wang H, Chang Y, Chu HH. Videoconference program enhances social support, loneliness, and depressive status of elderly nursing home residents. Aging Ment Health 2010 Nov;14(8):947-954. [doi: 10.1080/13607863.2010.501057] [Medline: 21069600]
7. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
8. Banks J, Nazroo J, Steptoe A. The Dynamics of ageing: Evidence from the English Longitudinal Study of Ageing 2002-2016 (Wave 8). Institute For Fiscal Studies - IFS. 2002. URL: https://www.ifs.org.uk/publications/13510 [accessed 2020-06-10]
9. Internet Users, UK: 2019. Office for National Statistics. 2019. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2019 [accessed 2020-05-22]
10. Forsman AK, Nordmyr J. Psychosocial links between internet use and mental health in later life: a systematic review of quantitative and qualitative evidence. J Appl Gerontol 2017 Dec;36(12):1471-1518. [doi: 10.1177/0733464815595509] [Medline: 26245208]
11. Dickinson A, Gregor P. Computer use has no demonstrated impact on the well-being of older adults. Int J Human-Comput Stud 2006 Aug;64(8):744-753. [doi: 10.1016/j.ijhcs.2006.03.001]
12. Shapira N, Barak A, Gal I. Promoting older adults' well-being through internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
13. White H, McConnell E, Clipp E, Branch LG, Sloane R, Pieper C, et al. A randomized controlled trial of the psychosocial impact of providing internet training and access to older adults. Aging Ment Health 2002 Aug;6(3):213-221. [doi: 10.1080/13607860220142422] [Medline: 12217089]
14. Slegers K, van Boxtel MP, Jolles J. Effects of computer training and Internet usage on the well-being and quality of life of older adults: a randomized, controlled study. J Gerontol B Psychol Sci Soc Sci 2008 May;63(3):P176-P184. [doi: 10.1093/geronb/63.3.p176] [Medline: 18559683]
15. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
16. Lissitsa S, Chachashvili-Bolotin S. Life satisfaction in the internet age – changes in the past decade. Comput Hum Behav 2016 Jan;54:197-206. [doi: 10.1016/j.chb.2015.08.001]
17. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. Lancet 2015 Feb 14;385(9968):640-648 [FREE Full text] [doi: 10.1016/S0140-6736(13)61489-0] [Medline: 25468152]
18. Vanhoutte B. The multidimensional structure of subjective well-being in later life. J Popul Ageing 2014;7(1):1-20 [FREE Full text] [doi: 10.1007/s12062-014-9092-9] [Medline: 25089162]
19. Jivraj S, Nazroo J, Vanhoutte B, Chandola T. Aging and subjective well-being in later life. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69(6):930-941 [FREE Full text] [doi: 10.1093/geronb/gbu006] [Medline: 24569002]

XSL•FO
RenderX

20. Ryff CD, Singer B. The contours of positive human health. Psychol Inq 2009 Nov 19;9(1):1-28. [doi: 10.1207/s15327965pli0901_1]
21. Matthews K, Nazroo J. Understanding Digital Engagement in Later Life. Developer Documentation: Government of UK. 2015. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/495105/FullReport-UnderstandingDigitalEngagementinLaterLife.docx.pdf [accessed 2020-05-22]
22. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
23. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010 Feb 28;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
24. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
25. Sum S, Mathews RM, Hughes I, Campbell A. Internet use and loneliness in older adults. Cyberpsychol Behav 2008 Apr;11(2):208-211. [doi: 10.1089/cpb.2007.0010] [Medline: 18422415]
26. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]
27. Khan AS, Rahman AU, Qazi LT. The relationship between internet usage, socioeconomic status, subjective health and social status. Bus Econ Rev 2016 Oct 1;8(Special Edition):67-82. [doi: 10.22547/ber/8.se.5]
28. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May 16;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
29. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. Int J Epidemiol 2013 Dec;42(6):1640-1648 [FREE Full text] [doi: 10.1093/ije/dys168] [Medline: 23143611]
30. Mindell J, Biddulph JP, Hirani V, Stamatakis E, Craig R, Nunn S, et al. Cohort profile: the health survey for England. Int J Epidemiol 2012 Dec;41(6):1585-1593. [doi: 10.1093/ije/dyr199] [Medline: 22253315]
31. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for epidemiologic studies depression scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychol Aging 1997 Jun;12(2):277-287. [doi: 10.1037//0882-7974.12.2.277] [Medline: 9189988]
32. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
33. Pavot W, Diener E. The satisfaction with life scale and the emerging construct of life satisfaction. J Posit Psychol 2008 Apr;3(2):137-152. [doi: 10.1080/17439760701756946]
34. Forsman AK, Nordmyr J, Wahlbeck K. Psychosocial interventions for the promotion of mental health and the prevention of depression among older adults. Health Promot Int 2011 Dec;26(Suppl 1):i85-107. [doi: 10.1093/heapro/dar074] [Medline: 22079938]
35. Banks J, Karlsen S, Oldfield Z. Socio-economic position. In: Marmot M, Banks J, Blundell R, Lessof C, Nazroo J, editors. Health, Wealth and Lifestyles of the Older Population in England: The 2002 English Longitudinal Study of Ageing. London, UK: Institute for Fiscal Studies; 2002.
36. Pollack CE, Chideya S, Cubbin C, Williams B, Dekker M, Braveman P. Should health studies measure wealth? A systematic review. Am J Prev Med 2007 Sep;33(3):250-264. [doi: 10.1016/j.amepre.2007.04.033] [Medline: 17826585]
37. Morgan C, Cotten SR. The relationship between internet activities and depressive symptoms in a sample of college freshmen. Cyberpsychol Behav 2003 Apr;6(2):133-142. [doi: 10.1089/109493103321640329] [Medline: 12804025]
38. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T, Scherlis W. Internet paradox. A social technology that reduces social involvement and psychological well-being? Am Psychol 1998 Sep;53(9):1017-1031. [doi: 10.1037//0003-066x.53.9.1017] [Medline: 9841579]
39. Moody EJ. Internet use and its relationship to lonelines. Cyberpsychol Behav 2001 Jun;4(3):393-401. [doi: 10.1089/109493101300210303] [Medline: 11710265]
40. van Boekel LC, Peek ST, Luijkx KG. Diversity in older adults' use of the internet: identifying subgroups through latent class analysis. J Med Internet Res 2017 May 24;19(5):e180 [FREE Full text] [doi: 10.2196/jmir.6853] [Medline: 28539302]
41. van Deursen AJ, Helsper EJ. A nuanced understanding of internet use and non-use among the elderly. Eur J Commun 2015 Apr 9;30(2):171-187. [doi: 10.1177/0267323115578059]
42. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among US young adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]

Abbreviations

CES-D: Center for Epidemiologic Studies Depression Scale
ELSA: English Longitudinal Study of Ageing
HRS: Health and Retirement Study

XSL•FO
RenderX

SEP: socioeconomic position

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by P Schulz, P Romero-Sanchiz; comments to author 03.11.19; revised version received 20.11.19; accepted 26.01.20; published 28.07.20.

Please cite as:
Lam SSM, Jivraj S, Scholes S
Exploring the Relationship Between Internet Use and Mental Health Among Older Adults in England: Longitudinal Observational Study
J Med Internet Res 2020;22(7):e15683
URL: https://www.jmir.org/2020/7/e15683
doi:10.2196/15683
PMID:32718913

©Sabrina Sze Man Lam, Stephen Jivraj, Shaun Scholes. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding the Dimensions of Medical Crowdfunding: A Visual Analytics Approach

Jie Ren[1*], PhD; Viju Raghupathi[2*], PhD; Wullianallur Raghupathi[1*], PhD

[1]Gabelli School of Business, Fordham University, New York, NY, United States

[2]Koppelman School of Business, Brooklyn College of the City University of New York, Brooklyn, NY, United States

[*]all authors contributed equally

Corresponding Author:
Wullianallur Raghupathi, PhD
Gabelli School of Business
Fordham University
140 W 62nd Street
New York, NY, 10023
United States
Phone: 1 2126367230
Email: raghupathi@fordham.edu

Abstract

Background: Medical crowdfunding has emerged as a growing field for fundraising opportunities. Some environmental trends have driven the emergence of campaigns to raise funds for medical care. These trends include lack of medical insurance, economic backlash following the 2008 financial collapse, and shortcomings of health care regulations.

Objective: Research regarding crowdfunding campaign use, reasons, and effects on the provision of medical care and individual relationships in health systems is limited. This study aimed to explore the nature and dimensions of the phenomenon of medical crowdfunding using a visual analytics approach and data crawled from the GoFundMe crowdfunding platform in 2019. We aimed to explore and identify the factors that contribute to a successful campaign.

Methods: This data-driven study used a visual analytics approach. It focused on descriptive analytics to obtain a panoramic insight into medical projects funded through the GoFundMe crowdfunding platform.

Results: This study highlighted the relevance of positioning the campaign for fundraising. In terms of motivating donors, it appears that people are typically more generous in contributing to campaigns for children rather than those for adults. The results emphasized the differing dynamics that a picture posted in the campaign brings to the potential for medical crowdfunding. In terms of donor's motivation, the results show that a picture depicting the pediatric patient by himself or herself is the most effective. In addition, a picture depicting the current medical condition of the patient as severe is more effective than one depicting relative normalcy in the condition. This study also drew attention to the optimum length of the title. Finally, an interesting trend in the trajectory of donations is that the average amount of a donation decreases with an increase in the number of donors. This indicates that the first donors tend to be the most generous.

Conclusions: This study examines the relationship between social media, the characteristics of a campaign, and the potential for fundraising. Its analysis of medical crowdfunding campaigns across the states offers a window into the status of the country's health care affordability. This study shows the nurturing role that social media can play in the domain of medical crowdfunding. In addition, it discusses the drivers of a successful fundraising campaign with respect to the GoFundMe platform.

(J Med Internet Res 2020;22(7):e18813) doi:10.2196/18813

KEYWORDS

crowdfunding; medical crowdfunding; GoFundMe; fundraising; health care; health care affordability; patient; Facebook; fundraiser

Introduction

The crowdfunding phenomenon extends the traditional idea of fundraising into a contemporary internet platform–based funding vehicle [1-3]. In this exploratory study, we applied visual analytics to study medical crowdfunding, in which individuals raise funds for their medical treatment or for research projects via digital platforms such as GoFundMe [4-8].

XSL•FO
RenderX

Crowdfunding is an old phenomenon. Its earliest form, designed by Joseph Pulitzer, channeled Pulitzer's newspaper to raise funds for the pedestal of the Statue of Liberty [1,9]. Similarly, this avenue has been used by artists, composers, and inventors to procure funding from backers willing to invest in their projects [10].

Crowdfunding has evolved into a powerful alternative to other traditional investment forms, such as venture capital, initial public offerings, and angel investments [1]. This trend has been facilitated by the ubiquity of digital media, including the internet, social media, e-commerce, and web-based advertising [11,12]. These platforms bestow comfort and ease in web-based transactions [1].

In the United States, the entrepreneurial culture has increased public awareness and the desire to support entrepreneurs [1,13]. As an alternative form of investment, crowdfunding offers opportunities for entrepreneurs to access various forms of funding [14]. As an example, Formlabs, a 3D printing startup, raised US $2 million in seed funding. The business used the Kickstarter crowdfunding platform to raise an additional US $2.9 million in the first year and US $19 million in the following year [1]. Kickstarter was initially used to raise small amounts of capital. However, its funding potential has progressively increased [15]. It is projected that the platform will generate more than US $300 million in funding by 2025 [16].

The process of crowdfunding begins with the entrepreneur's pitch (or narrative) on details such as their background, funding-reward structure, and overview of the project or product for which funding is needed. This stage is intended for prospective backers [1]. The narrative is used throughout the funding period such that potential funders can evaluate the venture or the entrepreneur can make a funding decision. Owing to its potential, the phenomenon of crowdfunding is deployed in various domains, including law and medicine [6,17-20].

This study aimed to explore the nature and dimensions of the phenomenon of medical crowdfunding. It used a visual analytics approach and data crawled from the popular GoFundMe platform. This type of exploration sheds light on the characteristics and magnitude of the phenomenon, offering ethical, health, and social policy implications and recommendations [4-8]. This exploratory study addressed the following questions:

1. What are the dimensions of medical crowdfunding?
2. What factors create a successful medical crowdfunding campaign?

The next section provides the background information on crowdfunding, particularly medical crowdfunding. Following this is the outline of the methodology. Next, the results of visual analytics are presented, along with the scope and limitations of the study. Finally, the implications and future research are covered, along with the conclusions.

Background

The success of crowdfunding in raising financial resources for new ventures has attracted the attention of entrepreneurs who explore this avenue for sourcing new ideas and projects [12,21-24]. In particular, crowdfunding facilitates entrepreneurs who are in the formative stages of gathering financial resources to turn raw ideas into real business [21,23,25,26]. As this field grows, its conceptualization also evolves [12,15,21,23,27-30].

Research has identified key factors and definitions related to this phenomenon, including crowd, project creator (funding requester or fundraiser), and platform [21,27,28,30-35]. Some researchers suggest that a comprehensive definition of crowdfunding should include these elements as well as how each relates to or impacts others [21,34].

Microfinancing and crowdsourcing definitions have adopted a general conceptualization of crowdfunding [21,26,29,32,33,36-39]. The scope extends to both commercial and noncommercial forms of crowdfunding [14,21,27,35,40]. For example, Paschen [26] defined crowdfunding as "the outsourcing of an organizational function, through information technology, to a strategically defined network of actors (ie, the crowd) in the form of an open call specifically, requesting monetary contributions toward a commercial or social business goal." The objective of crowdfunding is that a crowd provides the "financial resources either in form of donation or in exchange for some form of reward and/or voting rights to support initiatives for specific purposes" [41]. In the context of new ventures and entrepreneurial financing, crowdfunding is defined as "the efforts by entrepreneurial individuals and groups—cultural, social, and for-profit—to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the internet, without standard financial intermediaries" [29]. Other crowdfunding definitions, which focus on backers, are based on the contributions of an interested crowd of people in a specific phenomenon [42].

Using the context-specific approach, scholars have classified crowdfunding into several types [27,29,43-45]. Typically, 4 models can be broadly identified.

Donation

In this type of crowdfunding, funders support charity projects [46-49]. JustGiving, based in the United Kingdom, focuses on donation crowdfunding. Donors can be grouped based on the expectation of receiving rewards [26]. The first group (pure donors) financially supports crowdfunding projects without either monetary or nonmonetary returns [26]. The second group does not receive monetary rewards. Instead, it receives nontangible rewards such as recognition or tokens [21,50].

Investment (Equity)

This crowdfunding option allows funders to invest in a project or business and acquire equity in return. SoMoLend in the United States and Crowdcube in the United Kingdom are examples of this model [46].

Lending (Debt)

This type of crowdfunding is when funders lend money to a project or business and expect repayment with or without interest. Kiva is a well-known example of lending.

XSL•FO
RenderX

Rewards

In this system, funders receive tangible or intangible items as rewards [46]. Fundraisers can select individual reward schemes for their contributions. This is a popular form of crowdfunding [14,27,35,40,44,46,51-55]. Examples include GoFundMe, Indiegogo, and Kickstarter.

Among platforms, GoFundMe, Indiegogo, and Kickstarter were the top crowdfunding websites in 2019 in terms of both fundraising and volume [56]. Other crowdfunding platforms include Fundly, JustGiving, Crowdrise, Indiegogo, Teespring, Patreon, YouCaring [50], Chuffed, ArtistShare, MightyCause, InKind, Crowdfunder, Kiva, and GiveWP. Fundraisers can find support for novel projects through crowdfunding, which would have been difficult through traditional funding sources. In addition, fundraisers have a choice in selecting platforms in which they do not have to pay commissions unless the effort is a success. For example, Kickstarter collects a 5% commission only if a project reaches its funding goal [50].

The number of crowdfunding platforms has grown in the past few years. Platforms such as Kickstarter, GoFundMe, and Indiegogo have shown increasing trends in fundraising [22]. According to AP News, the global crowdfunding market is expected to exceed US $28.77 billion by 2025 [57]. The potential for crowdfunding has urged regulatory efforts in the domain, such as the Jumpstart Our Business Startups Act [22]. In addition, many nonprofit businesses and governments are considering crowdfunding as a source for financing community programs aimed at serving the public [58]. In this regard, crowdfunding analytic tools can offer insights into the design of a successful fundraising campaign [22].

Next, this study focuses on the object of its research, that is, medical crowdfunding.

Medical Crowdfunding

In general, crowdfunding is being used for the development of community or social projects [6,29,58,59]. Health care can be considered both a community and social project because it incorporates programs such as poverty reduction and child education [6,29,59,60]. In this context, medical crowdfunding has emerged as a growing area of opportunity for crowdfunding [5,20,61-65].

What is the niche for medical crowdfunding? In recent years, crowdfunding has been adopted for entrepreneurial finance and litigation. More recently, individuals have deployed it for health issues and costs [6]. The increased need for funding health initiatives—the development of vaccines for public health or improving research protocols and systems—makes it a viable channel for deployment in the medical domain [5,6,66].

There are some renowned cases of medical crowdfunding [5]. For example, a campaign on RocketHub developed an affordable cost genome drug analysis test for patients [6,67]. This successful fundraising campaign spurred other campaigns like a rare genomic project for helping children with rare genetic diseases [68]. A campaign on Indiegogo analyzed microorganisms found on common public surfaces. The research aimed to prevent the spread of contagious diseases through mobile alerts [6,8]. These, and other similar projects, motivated contributions by creating donor awareness about the potential for medical research [6,69,70].

Some US-based websites are exclusively devoted to helping patients. These include GiveForward, GoFundMe, and YouCaring [6,20,61,62,66]. People who are ill or with disabilities resort to crowdfunding as a means to raise funds for meeting health care costs [4,20,61,66]. Environmental trends drive patients and caregivers to these platforms to raise funds for medical care. Factors include lack of medical insurance; aftereffects of the financial downturn in 2008; and lack of health care regulations, such as the Affordable Care Act [4,6-8,68]. Although crowdfunding can be used for many forms of medical needs, campaigns for acute and exceptional medical needs are more likely to be funded than those for chronic needs [63]. Medical crowdfunding campaigns can include raising funds for a variety of reasons, such as treatment of diseases in adults or children, hospital expenses, postoperative care, homecare needs, and general support in terms of drugs and postdiagnosis protocols, among others.

There are some dedicated medical crowdfunding websites. YouCaring had 15,880 active medical campaigns in 2016. FundRazr had 5326 campaigns [8]. GoFundMe reported an increase from 8000 campaigns (raising US $1.6 million) in 2011 to over 600,000 (raising nearly US $150 million) in 2014 [5]. GiveForward reported that medical-related needs are its most popular form of crowdfunding campaign [63]. These websites are expected to grow annually at a rate of 25% [71].

Despite this trend, little attention has been given to medical crowdfunding campaigns in terms of exploring reasons for use, impact on the provision of medical care, and individual relationships to health systems [4,8,17,66]. This exploratory study was conducted to address the paucity and obtain a panoramic view of the current state of medical crowdfunding. It explored the dimensions of medical crowdfunding as well as the factors of a successful campaign.

Methods

Visual Analytics Approach

This data-driven study used a visual analytics approach with primarily descriptive analytics [72,73] to obtain a panoramic insight into medical projects funded through GoFundMe. Visual analytics provides researchers and policy makers with effective ways to comprehend and analyze large datasets while acting on the findings in real time [74]. By integrating the computer's capabilities with that of humans, visual analytics allows for the discovery of unexpected patterns and insights, which can lead to innovative and novel solutions [75,76]. It can address the challenge of information overload by translating information into viable opportunities, allowing researchers to examine results and the processes leading to those results [74,77,78]. The goal is to tell a compelling story through information visualization and the pillars of visualization, statistics, and data mining [75,77].

Descriptive analytics is based on the idea of describing data *as is* (without a preconceived assumption). Descriptive analytics

XSL•FO
RenderX

is more data driven than other models; it allows for the understanding of past and current patterns and data trends, using the insight for informed decision making [72,79]. Through categorization, characterization, and aggregation or classification of data, information is presented visually in the form of meaningful charts and reports for analyzing business decisions [72,79].

In the current context of medical crowdfunding, we deployed visualization with descriptive analytics to address insightful questions such as the following. Drawing from the literature on medical crowdfunding and from general insights in the domain of crowdfunding, particularly in the context of web-based platforms, we explored the following in our study:

1. What factors lead to the success of a medical crowdfunding campaign?
2. Is there an optimum length for a fundraising campaign to be successful?
3. Do gender and age of fundraisers play a role in the amount of funds that can be raised?
4. Do pictures posted by fundraisers impact the success of the campaign?
5. What role does social media play in crowdfunding, particularly medical crowdfunding?
6. How does an analysis of medical crowdfunding campaigns across the states offer a window into the status of health care affordability in the country?

The Platform: GoFundMe

The data source for this study is the GoFundMe crowdfunding platform, founded in 2010. It is the world's largest crowdfunding platform in terms of the total funding amount raised and the total number of active campaigns [80]. GoFundMe campaigns include fundraising for several categories such as medical, memorials, emergency situations, and charitable causes.

In this study, we extracted data between June 19 and June 25, 2019, from the medical category of GoFundMe. We used the Selenium and Beautiful Soup packages in Python to crawl data from the medical category. Although it is possible that medical campaigns may exist in other categories, the medical category exclusively holds the bulk of these campaigns [7]. Therefore, we consider this to be an appropriate repository for medical crowdfunding campaigns. In addition, to ensure that only medical campaigns were included from this category in our data, we performed a manual perusal of the project descriptions upon extraction. Textbox 1 outlines the methodology used.

Table 1 describes the entities in the study and the variables for each entity. The entities in the crowdfunding study include the *fundraiser* (the person who is running the campaign and raising medical funds for himself or herself or another), the *campaign*, and the *donor*. We also show the source of each variable in terms of whether the code was directly downloaded from the data source or had to be coded after a manual perusal of the data. The variables are described in further detail in the context of the data analyses and results.

Using the web crawler, we collected data on 1000 medical crowdfunding campaigns. Three campaigns were eliminated for lack of data, resulting in a sample size of 997 fundraising campaigns. The raw data were extracted in a comma-separated value format and loaded into Tableau (Tableau Software Inc), a visual analytics tool. The following section discusses the results of the analyses.

Textbox 1. Methodology.

Data collection

- GoFundMe

Data preparation

- Data crawled from GoFundMe, extracted in a comma-separated values format, and prepared for an analytic tool

Platform selection and implementation

- Tableau 2019.2

XSL•FO
RenderX

Table 1. Variables in the research.

Entity and variable	Description	Source of variable code
Patient (for whom funds are being raised in the campaign)		
Gender	Gender refers to the gender of the patient for whom the campaign is developed (M: male, F: female)	Directly from the data source
Age	Age refers to the age of the patient for whom the campaign is developed. People below the age of 18 years are considered children, while others are considered adults (C: children, A: adult)	Directly from the data source
Campaign		
Group picture	Represents a picture that is posted in the campaign. If there is a single person in the picture, it is deemed to be a single picture. All other pictures are categorized as a group picture (G: group picture; S: single picture)	Manual review of data
Status	Status provides details regarding the picture of the patient (P: picture in which the patient appears healthy and not in a medical setting; N: picture in which the patient appears ill and in a medical setting)	Manual review of data
Position	Websites have multiple pages. Therefore, campaigns on the homepage are marked as *position 0*. Those on successive pages are marked with successive numbers (0-333)	Manual review of data
Location	The geographical location of the fundraiser (city)	Directly from the data source
Goal	The dollar amount that is sought through the campaign (US$)	Directly from the data source
Length of fundraising	The time period for which the campaign has been active (months)	Directly from the data source
Amount raised	Amount raised denotes the total dollar amount that the campaign generated (US$)	Directly from the data source
Facebook shares	Number of people who shared the campaign link on their Facebook page (#)	Directly from the data source
Favorite hearts	The number of people who liked the campaign on GoFundMe. This shows support toward the campaign (#)	Directly from the data source
Donor		
Number of donors	Number of people who donated to a campaign (#)	Directly from the data source

Results

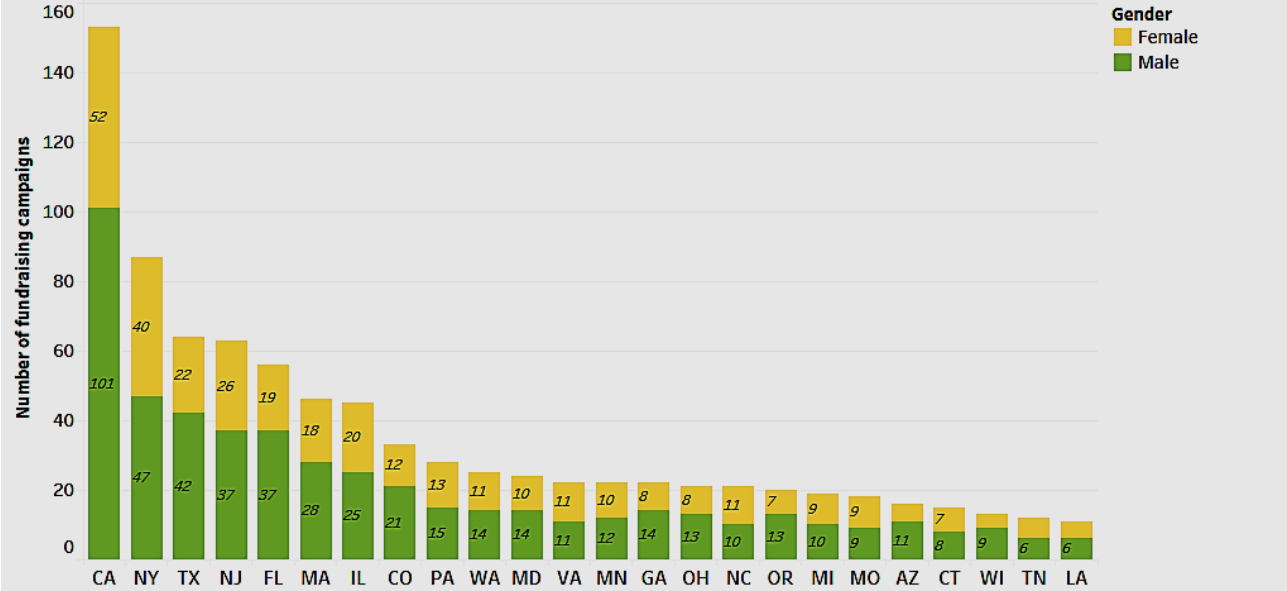
Distribution of Campaigns

We first explored the geographic distribution of the number of medical fundraising campaigns across various states in the United States (Figure 1). In addition to the geographical distribution, the figure depicts the distribution by gender.

Figure 1 shows that California has the highest number of fundraising campaigns (n=163), whereas the average number of campaigns for most states is approximately 20. This is followed by New York with 90, Texas with 65, New Jersey with 64, and Florida with 58 campaigns. We can only speculate on the distribution of fundraising amounts by state. For one, according to the 2019 US Census, these are also the states with a population of over 10 million [81]. It is also suggested that platforms such as GoFundMe, Fundly, and JustGiving are gaining popularity as people struggle to meet rising health care costs [82]. On the basis of this information, our results suggest the need for further empirical exploration on whether the high number of campaigns in these states is, in some way, a reflection of the high population that naturally entails a higher usage of the platform, or more people using the crowdfunding platform to cope with rising health care costs [81,82]. Figure 1 also reveals the analysis for gender. It shows that there are more campaigns by male patients 59.2% (59.2/100) compared with female patients 39.4% (39.4/100). In terms of age, our data show that most patients are primarily adults 80.9% (80.9/100).

XSL•FO
RenderX

Figure 1. Distribution of fundraising campaigns by state.



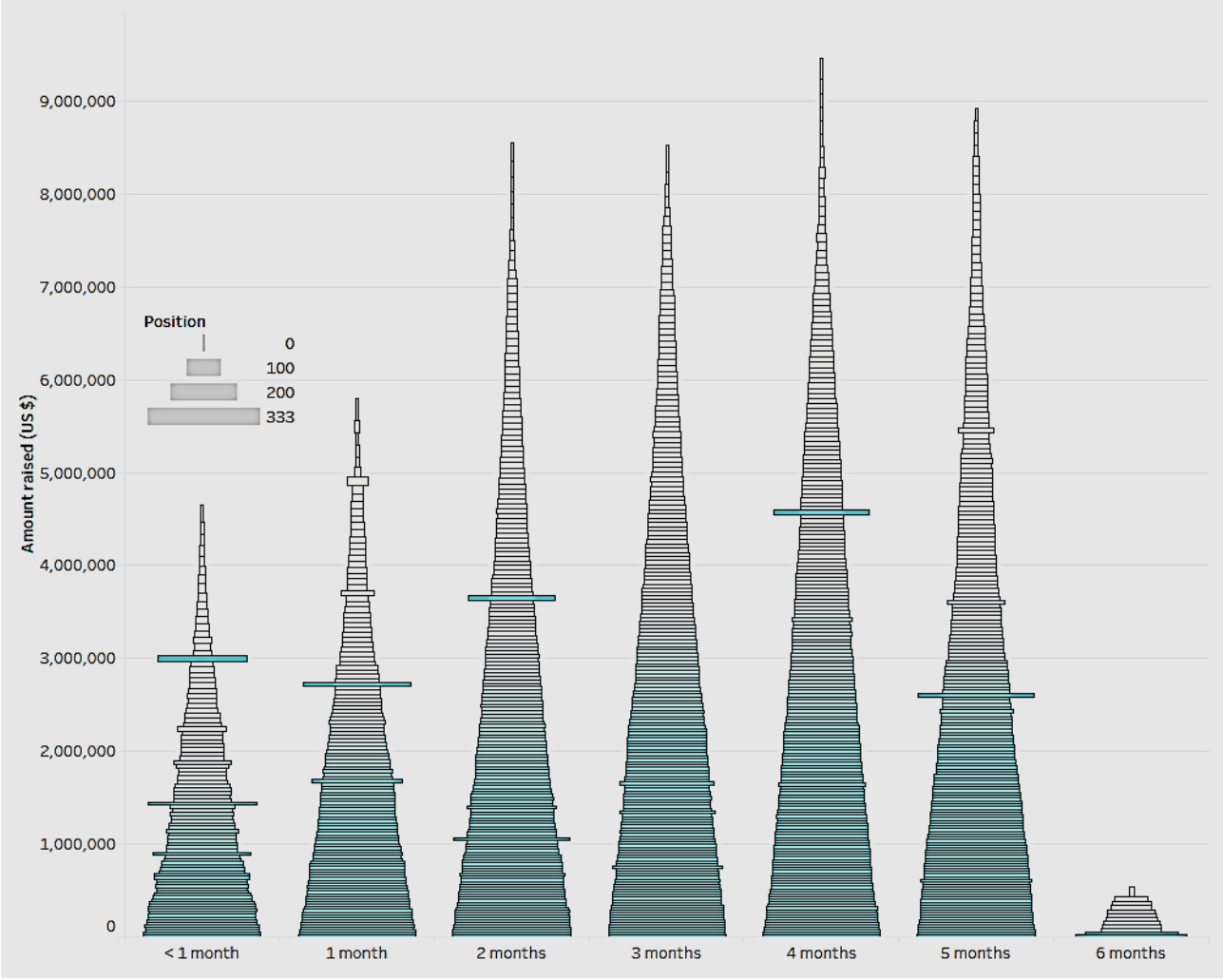
Factors Driving the Success of Campaigns in Terms of Amount Raised

We then explored factors that can drive the success of a campaign. For one, if the position of the campaign posting on the website and the length of time that the fundraising campaign was active had any influence on the amount of funds raised (Figure 2). In Figure 2, the x-axis shows the duration of the campaign in months, and the y-axis shows the total amount raised. The position is represented by the size of the band for each campaign within each bar. The position of the campaign is the page on the website on which the campaign appears. The smallest bar represents a campaign positioned on the first page of the website (0). The size of the bar increases with the change in position on successive pages. The positioning of the campaign can play a major role in raising funds. Per the rationale, campaigns on the front pages receive higher visibility with donors. Thus, these campaigns have a higher chance of funding.

The results in Figure 2 show a positive correlation between the position of the campaign and the amount raised. Campaigns posted in the first few pages of the website do have a higher amount raised. With respect to the length of the campaigns, we found that campaigns that ran for 2 to 5 months received better funding and had a higher chance of meeting their goals. This shows that there is an optimal period for a campaign to remain active, beyond which the potential for increasing returns is questionable. Although this is an interesting finding, there is no prior research or literature to substantiate the reasons for the lack of an incremental effect on the amount raised after 5 months. We can only speculate on a few. For one, because of the increasing number of campaigns, after a period of time, donor attention may be diverted to other newer campaigns on the platform. Second, in medical crowdfunding, since the cause for funding is medical, this may have an effect on the perception of the sense of urgency of the need. A campaign that is open for a longer period may not convey an imminent sense of urgency as another that is open for a shorter term. However, further studies are needed to explore this in greater detail, incorporating more sophisticated techniques.

XSL•FO
RenderX

Figure 2. Impact of positioning and length of fundraising.



Factors in Closing the Gap Between the Goal and the Actual Amount Raised

In the context of fundraising, it is important to identify which features of the campaign have a favorable influence in terms of closing the gap between the goal amount and the amount raised. The aim is to ensure that the difference is reduced and that the goal is reached.

In this regard, we studied the role of the campaign length in influencing the gap between the goal and the amount raised. Figure 3 shows a combination line chart and a bar-in-bar chart. The blue bar shows the average goal of the fundraising campaign, the orange bar shows the average amount raised, and the yellow line chart shows the difference between the goal and the actual amount raised. These are depicted in the context of the length of fundraising (the x-axis).

According to Figure 3, as the length of the fundraising period (in months) increases, the difference between the goal and the actual amount raised decreases. The decrease lasts up to a period of 5 months. After that, the difference increases. This trend indicates that the optimal fundraising period to meet a goal is between 2 and 5 months.

Next, we explored the role of other variables related to the patient (such as gender and age) and the picture (status and group picture) posted in the campaign (Figure 4). Gender refers to whether the patient is male (M) or female (F); age refers to whether the patient is a child (C) or an adult (A); status refers to the depiction of the status of the patient as reflected in the campaign picture—whether the patient appears healthy and not in a medical setting (P) or the patient appears ill and in a medical setting (N); and a group picture denotes whether it is a single picture (S) portraying only the patient or a group picture (G). Figure 4 uses a combination chart to show the average percentage of the goal reached and the average amount raised by each of the 4 variables.

In Figure 4, analysis by the status of the picture shows that campaigns with the picture of a patient appearing ill and in a medical setting (N) had a higher average amount raised and a higher average percentage of goal compared with campaigns with pictures where the patient appears healthy and not in a medical setting (P). On analyzing by gender, campaigns for male patients raised more funds (approximately 8%) on average than those for female patients; when comparing the average percentage of goal reached, campaigns for male groups lagged (by about 4%). When considering both the criteria of the average amount raised and the percentage of goal reached, campaigns for children raised more funds (about 2.5%) than those for adults.

XSL•FO
RenderX

Figure 3. Role of the length of fundraising in closing the gap between the goal and the amount raised.

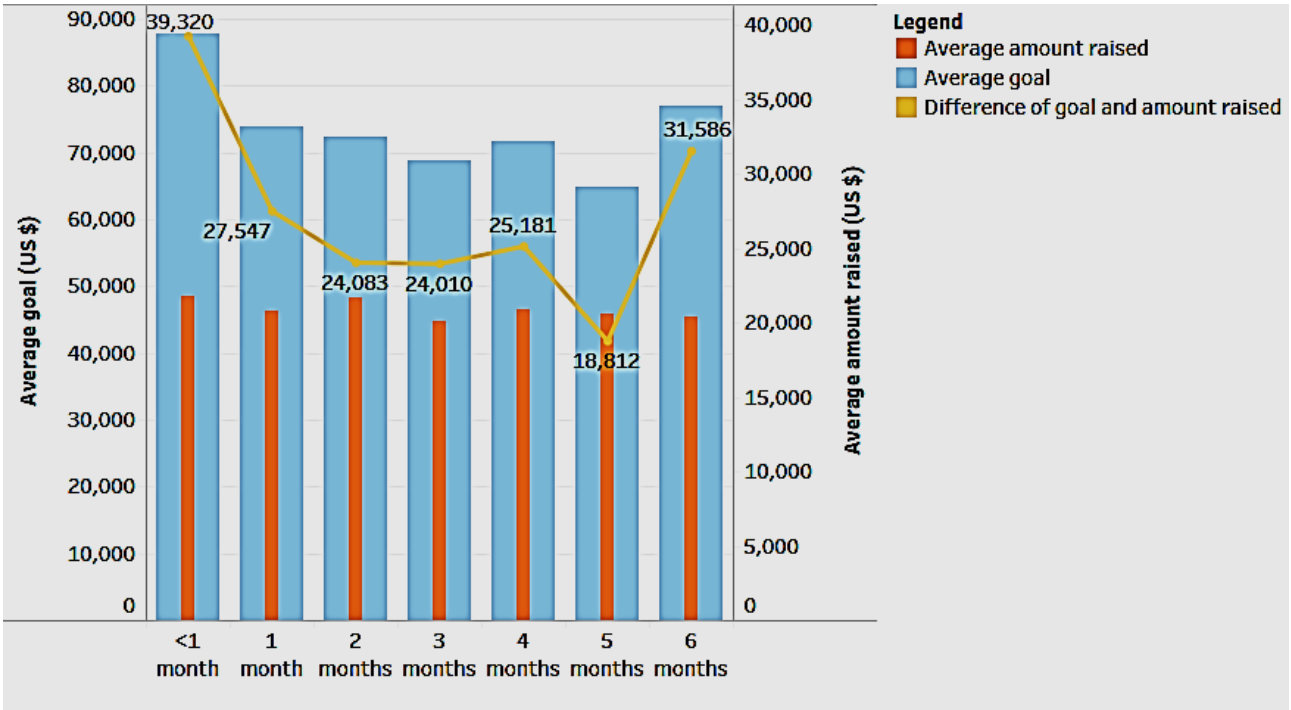
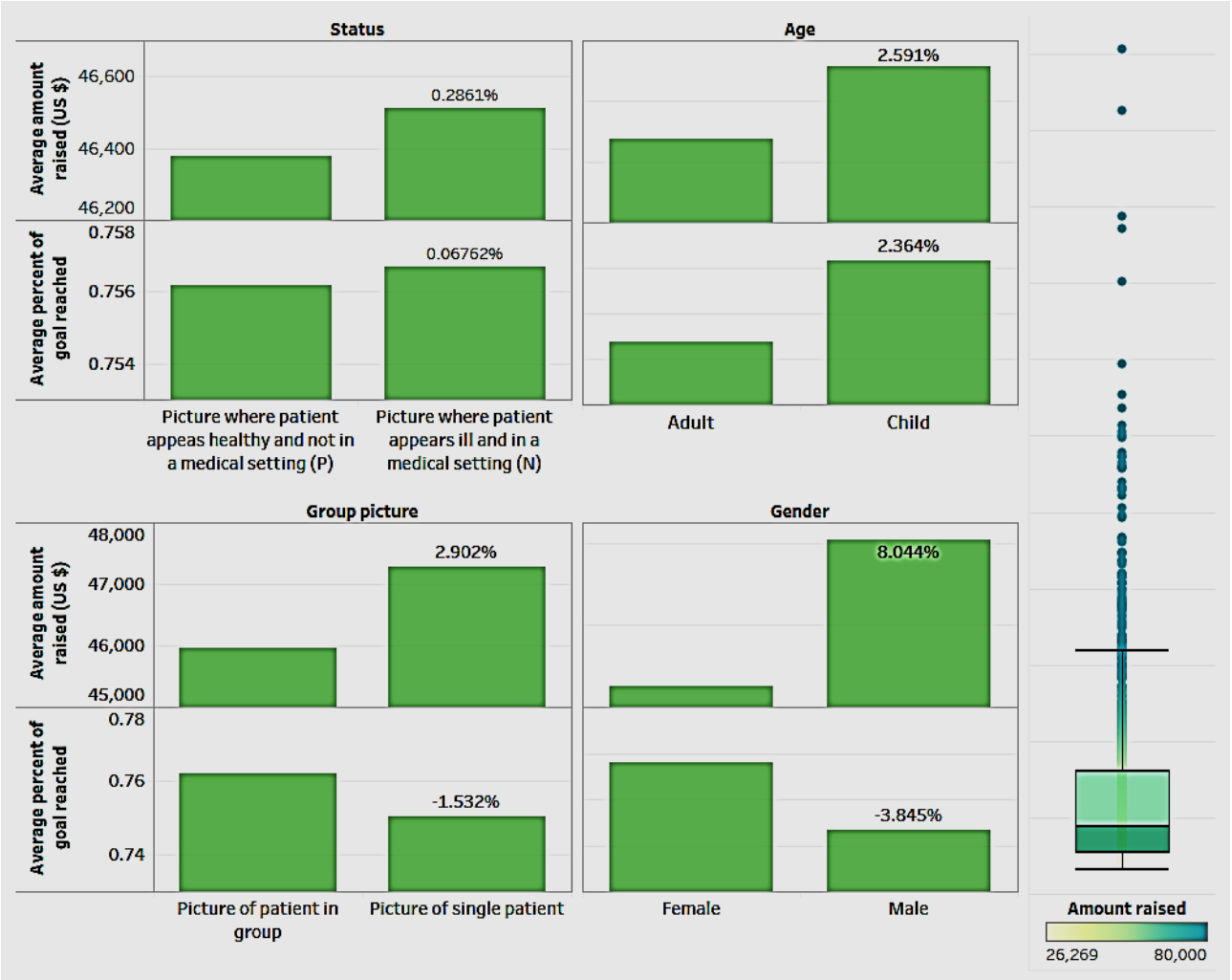


Figure 4. The average amount raised and percentage of goal reached by the different variables.

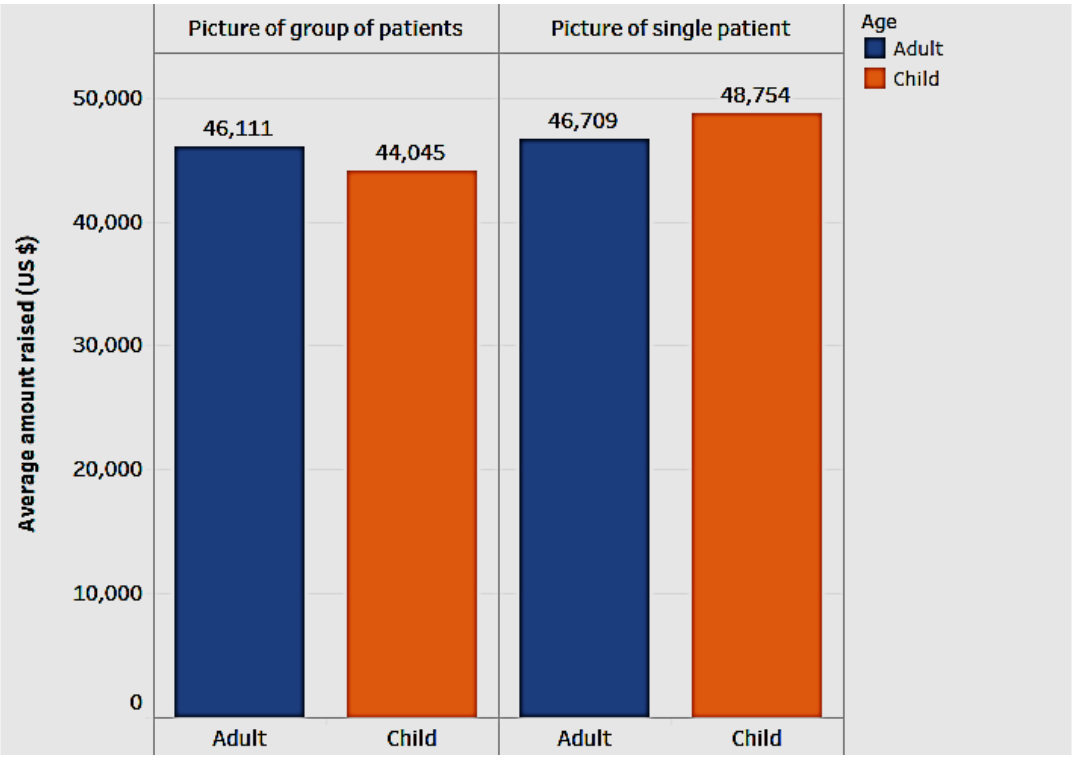


XSL•FO
RenderX

While the above analysis revealed the separate influence of age and picture on the average amount raised, we now analyzed their combined influence on the average amount raised. Figure 5 depicts the amount raised for adult and pediatric patients, differentiated by the type of picture in the campaign (single: S or group: G).

Figure 5 depicts the amount raised for adult and pediatric patients differentiated by the type of picture in the campaign. Some campaigns use a picture showing a single patient while others show the patient in a group. Our findings from the figure show that for adult patients, the average amount raised with the picture of a single patient is slightly lower than that raised using a group picture. However, in the case of children, the trend is reversed, with a higher amount being raised with the picture of a single patient. Therefore, for adults, the difference is in the opposite direction and much less visible than for children. This holds interesting prospects for future research to further analyze the factors behind the differential effect being more pronounced for children. In addition, campaign developers can design campaigns with this differential in mind.

Figure 5. Influence of age and type of picture on the average amount raised.



Association of Popularity Measures of the Campaign With Amount Raised

Campaigns have several measures that denote their popularity. In our study sample from GoFundMe, these include favorite hearts and Facebook shares. Favorite hearts denote the number of people who liked the campaign on GoFundMe. Facebook shares denote the number of people who shared the campaign on their social media platform, Facebook. In addition to these 2 measures, we also analyzed the influence of the goal and number of donors on the amount raised. We further showed the percentage of goals reached in the scatterplots (Figure 6). The dashboard in Figure 6 shows the associations for each of the 4 variables.

Figure 6 shows that favorite hearts have a significant and positive correlation (P<.001; R^2=0.26) with the amount raised. Facebook shares did not have a significant correlation with the amount raised (P=.06). There was a significant positive correlation between funding goals and the amount raised for a campaign (P<.001; R^2=0.30). Therefore, the higher the goal, the higher the amount raised. Finally, the number of donors had a significant positive correlation with the amount raised (P<.001; R^2=0.47), which indicates that the higher the number of donors the higher is the amount raised. Keeping in mind that the average donation amount tends to be the same, the higher the number of donors correlates with the higher probability of raising a larger amount. From the analyses, we can see that in the association of goal with the amount raised, the colored bubbles show that the higher the goal amount, the higher the possibility of the percentage of the goal being achieved.

XSL•FO
RenderX

Figure 6. Association of amount raised with campaign popularity measures.

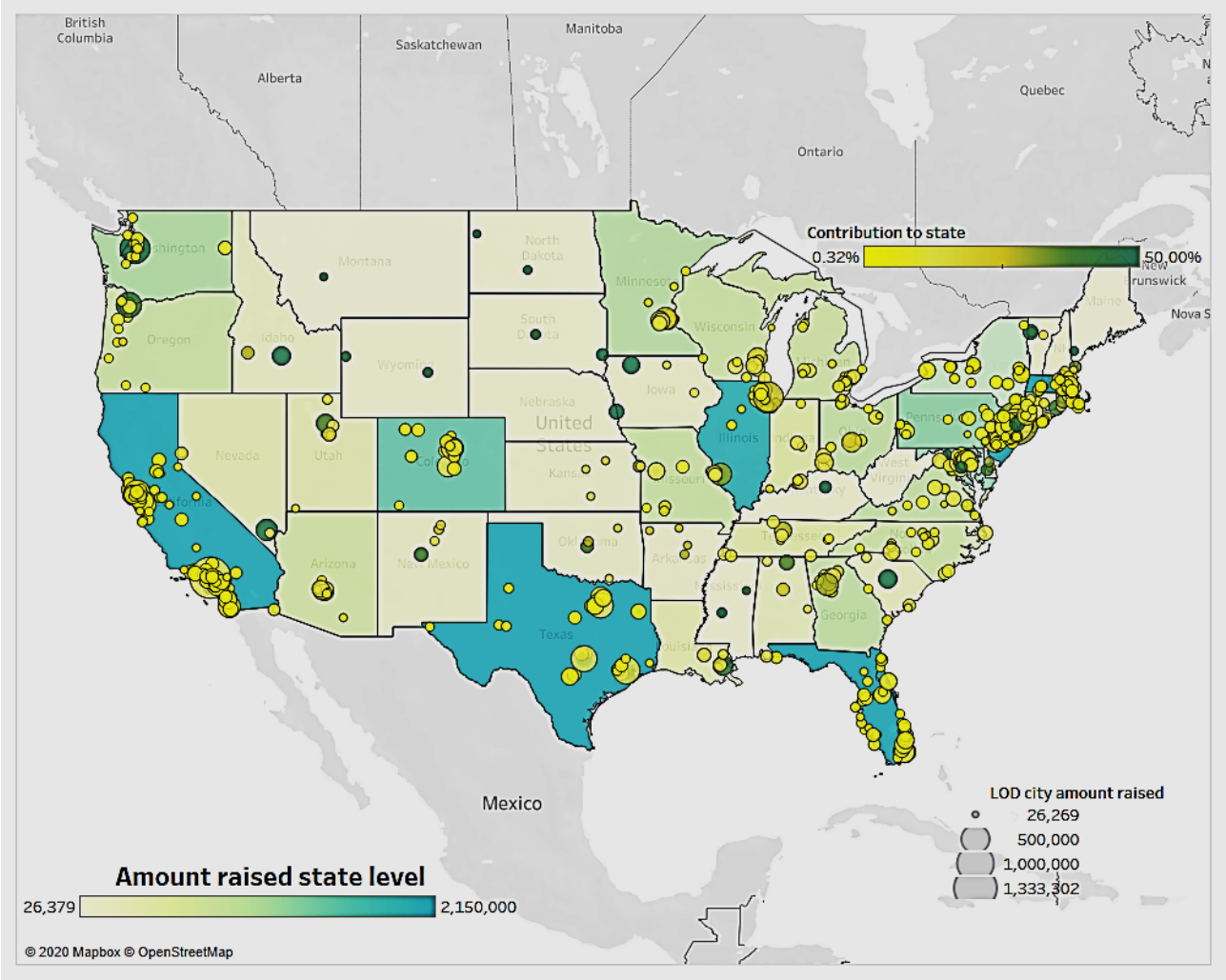


Figure 7 illustrates the state- and city-level distribution of the amounts raised. This includes 3 characteristics: the intensity of color inside the states denotes the amount raised by the state, such that the darker colors show larger amounts raised; the bubbles inside each state represent the amount raised at the city level; and the bubble color indicates the contribution of the city-level amount raised to the state's overall campaign amounts raised.

Figure 7 shows that California, New York, Texas, and Illinois have the highest amount raised compared with the other states. It also depicts clusters in the above states. These represent the locations of most of the crowdfunding campaigns. For example, areas around New York City and Los Angeles are the most concentrated in terms of campaigns created as well as amounts raised.

XSL•FO
RenderX

Figure 7. Amount raised by state and city.

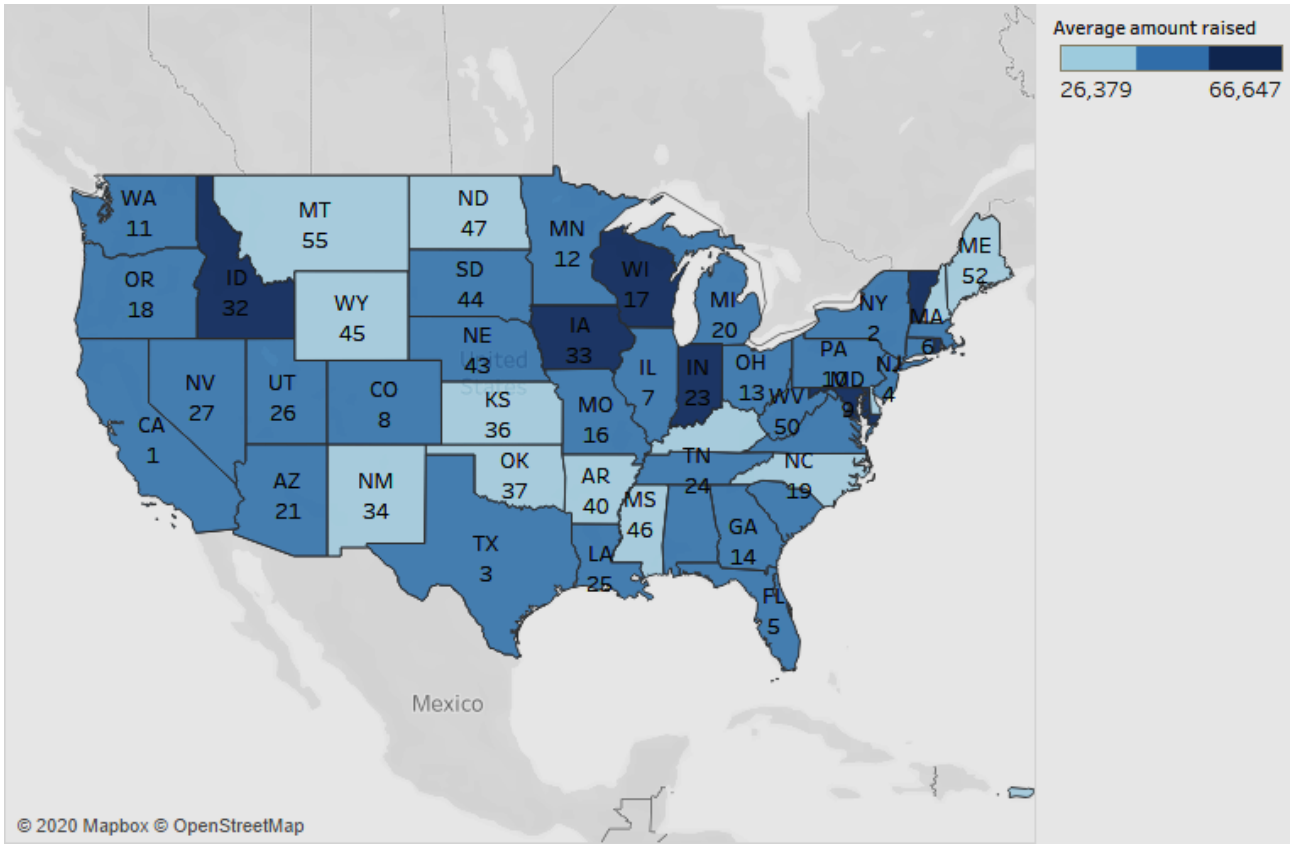


Along the lines of state-level distribution, we performed a state-level comparison of the average amount raised and the total amount raised (Figure 8).

In Figure 8, in addition to the average amount raised in each state (denoted by the color), the ranking in terms of the total amount raised is indicated as a label. The figure shows Iowa, Indiana, Wisconsin, and Idaho as the states with the highest average amount raised. However, in terms of the total amount raised, these states are ranked as 33rd, 23rd, 17th, and 32nd, respectively. In terms of the total amount raised, California ranks first, followed by New York, Texas, New Jersey, and Florida. These results highlight the potential for platforms such as GoFundMe to customize their campaign strategies based on the funding performance of the different states.

XSL•FO
RenderX

Figure 8. Average amount raised by state and ranking by total amount raised.

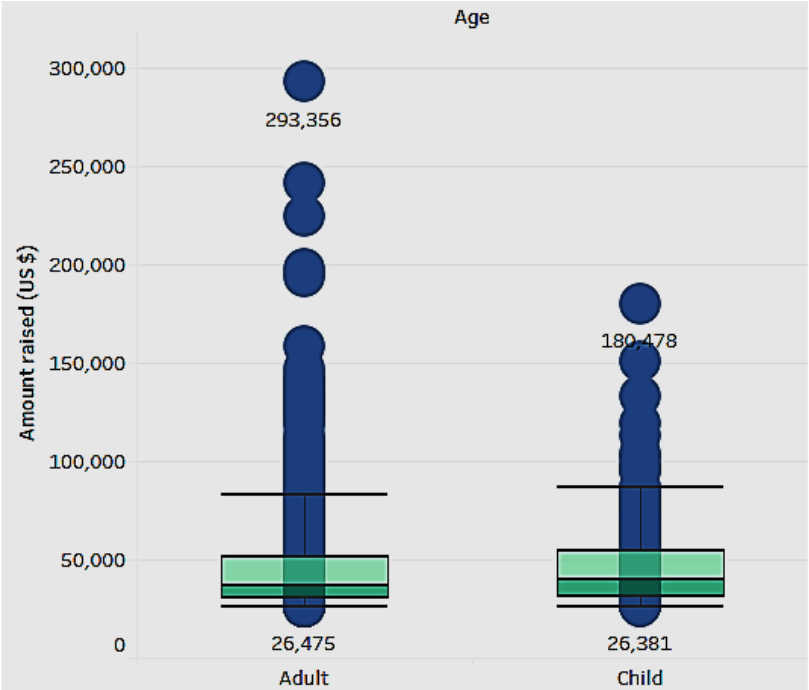


Association of the Amount Raised With Relevant Campaign and Fundraiser Variables

The box plot in Figure 9 shows the association of the amount raised with the age of the patients (child or adult) for whom the fundraising campaigns were developed.

Figure 9 shows that campaigns for adult and child patients have a similar median (approximately US $30,000 to US $40,000). However, the adult group has a higher biased distribution with more outliers. We propose that crowdfunding platforms pay special attention to these outliers because they may represent unique circumstances requiring specialized assistance and strategies for funding.

Figure 9. The median amount raised by age.

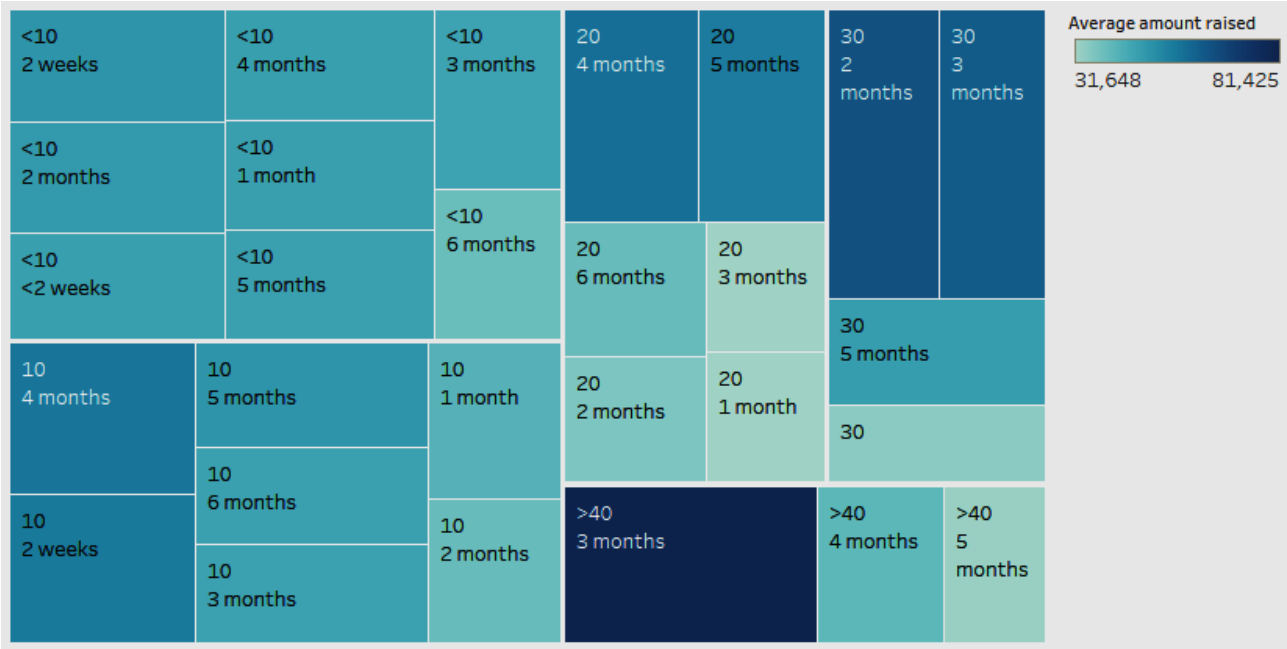


XSL•FO
RenderX

We compared the amount raised with the number of updates and the length of the fundraising campaign (Figure 10). In the figure, the intensity of the color denotes the average amount raised, with the darker color representing higher amounts. The 2 labels inside the boxes represent the number of updates (top label) and the length of the fundraising campaign in months (bottom label). The updates represent those posted to the campaign regarding health and treatment.

Figure 10 shows that campaigns with an average of 30 to 40 updates can raise much higher amounts in a 2- to 3-month period. The highest average amount raised is for a fundraiser with more than 40 updates and a period of 3 months. It should be noted that these are average values. In calculating these averages, we did not consider the position or the goal of the campaign. They were assumed to be the same across campaigns. These results indicate that we can focus on the length and the number of updates as relevant criteria to achieve a fundraising goal.

Figure 10. The average amount raised, number of updates, and length of campaigns.



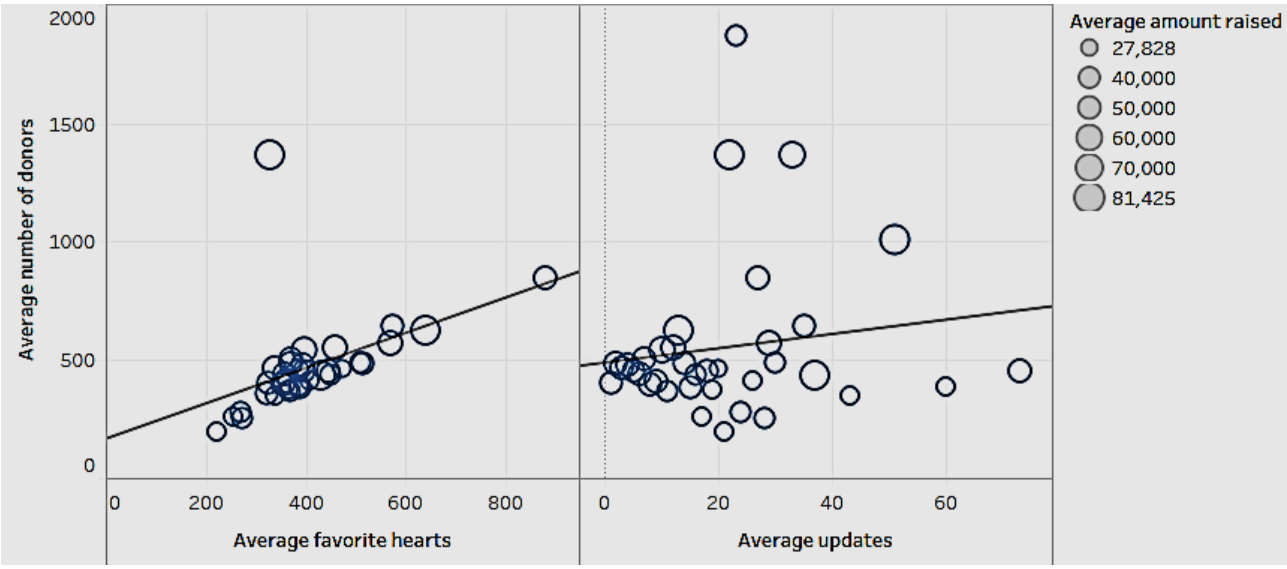
Trajectory of Donations and Influence of Favorite Hearts and Updates

Next, we looked for patterns in donations over time based on the number of updates or the number of favorite hearts in a campaign (Figure 11).

Figure 11 shows the relationship between the average number of donors with favorite hearts and the average number of updates in a campaign. The size of the bubbles in the scatterplots indicates the average amount raised. In Figure 11, we can see that the number of favorite hearts shows a significant positive association (P=.04) with the number of donors. Favorite hearts represent the number of people who liked the campaign on the GoFundMe website. It appears that people are more motivated to contribute if they see others endorsing a campaign in some manner. However, considering the lack of significance (P=.06) about the number of updates, it appears that donors focus on the initial story without following successive updates.

XSL•FO
RenderX

Figure 11. The average number of donors, favorite hearts, and updates.



Role of the Title and Description of the Campaign on Donations

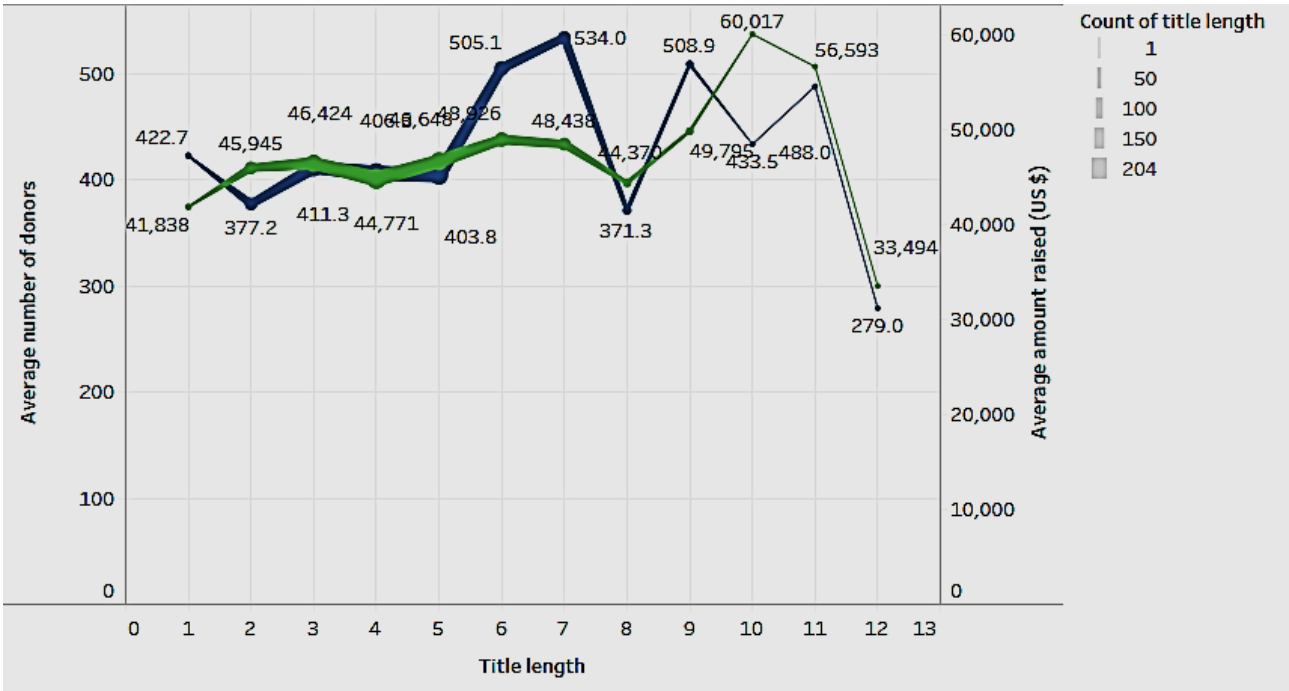
The title of the campaign was important. Therefore, we wanted to explore whether it influenced the average number of donors and the average amount raised (Figure 12). Figure 12 shows the two-line charts. The green line indicates the average amount raised, and the blue line indicates the number of donors.

Figure 12 shows that campaigns with titles having a length of 6, 7, and 9 words had a higher average number of donors. Titles with 10 or 11 words had a higher average amount raised. It can be inferred that the title plays an important role. An interesting title with 6 to 11 words can have a strong association with raising funds. The figure shows that most campaigns use 2- to 7-word titles. Future studies can explore the content of titles in addition to the length of in-depth associations.

Multimedia Appendix 1 depicts a word cloud for keywords that appear more than 500 times in the textual descriptions of campaigns.

A word cloud is a visual representation of the frequency and value of a word. It is used to highlight popular or trending terms based on the frequency of use and prominence in a corpus of text. The more times a keyword is present in a dataset, the bigger and bolder the keyword appears. Multimedia Appendix 1 shows that frequently occurring words include help, family, cancer, support, medical, treatment, time, and hospital. These represent the most discussed topics as well as the key factors in medical crowdfunding campaigns. The word cloud analysis shows that a large number of campaigns were designed to help with costs related to cancer treatment.

Figure 12. Influence of title length.



XSL•FO
RenderX

Discussion

Principal Findings

The visual analysis provides key insights that can help in developing successful medical crowdsourcing campaigns. States such as California and New York show the highest number of campaigns in the country. This indicates the need for further empirical investigation into a few possibilities: whether the high population in these states [81] naturally suggests higher platform usage or being that the objective of medical crowdfunding is primarily to meet medical costs [82], whether high platform usage actually reflects people reaching out for health care needs [81,82]. This study showed that some factors influence the amount raised in a medical crowdfunding campaign. For example, the positioning of a campaign is crucial. Posting a campaign within the first 10 pages of the website enhances its visibility to donors.

In terms of donor motivation, people are typically more generous in contributing to campaigns for children. It is also interesting to note the dynamics that a picture brings to the fundraising potential of a medical campaign. This study showed that a single patient picture is more effective for children. A group picture is more effective when fundraising for adults. A picture depicting the current medical condition of the patient (as severe) is more effective in motivating donations than one that depicts normalcy in the patient's condition.

We draw attention to the importance and effectiveness of the title in a medical crowdfunding campaign. In this regard, an optimum length contains 6 to 11 words. Regarding the role of gender and age of patients for whom the campaigns are developed, it appears that campaigns for female children are more successful, raising up to 78% of the goal amount.

Finally, an interesting trend in the trajectory of donations is that the average donation decreases with an increase in the number of donors. This indicates that the first donors tend to be the most generous. These initial donations are typically substantive, suggesting that they may be from the family or friends of the fundraiser. The public at large seems to donate marginally less than the initial donors.

Scope and Limitations

This study has some limitations. We crawled the data during a specific time period, that is, 2019. Therefore, it provides a snapshot of crowdfunding activity. Future studies should span a longitudinal timeframe and a more expansive set of variables. In addition, although we used a visual analytics method, alternative techniques such as data mining (ie, clustering, association) can be deployed. Although visual analytics, as a methodology, offers descriptive analysis, further empirical investigation is needed to draw quantitative conclusions. Furthermore, although the research focuses on the amount raised, goal, and position of the campaign, information on the fundraiser's background is not included. Access to fundraiser information would help ascertain whether certain profiles are more successful in attracting donations. Including specific information about the donors may also enable researchers to conclude whether the donor knew the patient or how many people donated to an unknown cause. Future studies can analyze differences in the type of patients for whom funds are raised (eg, medical expenses for a patient or pet). In addition, future studies can incorporate details on the insurance coverage of fundraisers. This might help ascertain their actual medical requirements. It will also be interesting to determine the current situation of fundraisers with regard to whether the treatment is completed or ongoing.

Implications

Despite these limitations, this research presents numerous contributions to the literature on medical crowdfunding and health care. This research highlights factors that are key to the success of a medical crowdfunding campaign. It also demonstrates the critical role of social media in the domain of health care. For example, research shows that campaigns with photos, frequent updates, and descriptions get more hearts (likes) and Facebook shares.

Through this research, we show how data-driven analytics can help donors make educated, fact-based contribution decisions in medical crowdfunding. The visual analytics methodology offers a holistic perspective on the phenomenon, including insight for policymaking in the arena of medical crowdfunding. Furthermore, this study offers insight into the geographic distribution of crowdfunding campaigns. This highlights the need for advanced analytics to empirically explore the contributing factors for differentials in platform use and fundraising success by state.

Conclusions and Future Research

This study focused on the dimensions and factors of a medical crowdfunding campaign using the most popular platform, GoFundMe. We examined the relationship between the use of social media, the characteristics of a campaign, and the potential for fundraising. The analysis of medical crowdfunding campaigns across the states offers a window into the status of health care affordability in the United States. The research shows the nurturing role that social media can play in the domain of medical crowdfunding. We also add to the drivers of a successful fundraising campaign with respect to the GoFundMe platform.

This research is significant because the topic of medical crowdfunding has been gaining public attention. Ethical concerns have been raised in relation to these kinds of campaigns. Most of these concerns focus on issues relating to both fundraisers and donors, such as exposure to fraudulent campaigns, loss of privacy, and fairness in fund distribution [6,8,65]. There is also a debate on whether campaigns can be designed to raise funds for medical treatment or conditions related to pets.

In the meantime, systematic efforts to comprehend the scope of these ethical concerns are lacking. For example, a search on Google Scholar using the terms *medical* and *crowdfunding* returned only 3 results focusing on crowdfunding specifically for covering medical expenses rather than for research [6,8,68]. Researchers and academicians can add insight into the phenomenon of medical crowdfunding given the interest in the social dimensions of health care combined with the ability to deploy a variety of methods from different disciplines [6,8,65].

XSL•FO
RenderX

For instance, health geographers can offer insight into the spatial dimensions of campaigns, such as the extent of the donor network from an ethics of care approach [6,8,68]. Health economists can explore the regulatory aspects related to the possible competing interests of parties in medical crowdfunding. Health informatics experts can incorporate leading-edge technology such as text analytics or sentiment analysis to analyze whether the severity of the medical condition influences the funds raised for medical crowdfunding campaigns.

Drawing from our research and the related literature, we propose ethical and social considerations for future medical crowdfunding research. First, it is important to explore how medical crowdfunding impacts the national phenomenon of health care, specifically affordability and cost of care. The exploration of whether the age group of patients has a differential impact on fundraising potential is also important. Second, crowdfunding platforms are all technology based. Therefore, research needs to explore fundraiser information privacy and security [6,8,65]. Third, the associated aspect of regulating misuse of platforms for fraudulent campaigns, including misrepresenting the severity of illness for securing funds or misappropriating funds, needs to be investigated [8,62]. Fourth, the results of the study emphasize the role of social media endorsements for crowdfunding campaigns, such as Facebook shares. It also questions whether social media companies should regulate the authenticity of endorsements as well as their role in influencing potential donors. At this stage, more social, political, and economic issues will continue to be explored as the prevalence and popularity of crowdfunding technology platforms rise.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word cloud of frequently used words in textual descriptions.
[PNG File , 493 KB - jmir_v22i7e18813_app1.png]

References

1. Allison T, Davis B, Webb J, Short JC. Persuasion in crowdfunding: an elaboration likelihood model of crowdfunding performance. J Bus Ventur 2017 Nov;32(6):707-725. [doi: 10.1016/j.jbusvent.2017.09.002]
2. Anglin AH, Short JC, Drover W, Stevenson RM, McKenny AF, Allison TH. The power of positivity? The influence of positive psychological capital language on crowdfunding performance. J Bus Ventur 2018 Jul;33(4):470-492. [doi: 10.1016/j.jbusvent.2018.03.003]
3. Borst I, Moser C, Ferguson J. From friendfunding to crowdfunding: relevance of relationships, social media, and platform activities to crowdfunding performance. New Media Soc 2018 Apr;20(4):1396-1414 [FREE Full text] [doi: 10.1177/1461444817694599] [Medline: 30581357]
4. Berliner LS, Kenworthy NJ. Producing a worthy illness: personal crowdfunding amidst financial crisis. Soc Sci Med 2017 Aug;187:233-242. [doi: 10.1016/j.socscimed.2017.02.008] [Medline: 28274601]
5. Cameron P, Corne DW, Mason CE, Rosenfeld J. Crowdfunding genomics and bioinformatics. Genome Biol 2013;14(9):134 [FREE Full text] [doi: 10.1186/gb-2013-14-9-134] [Medline: 24079746]
6. -. Crowdfunding. A new option for funding health projects. Arch Argent Pediat 2015 Apr 1;113(2):154-157. [doi: 10.5546/aap.2015.eng.154]
7. Sisler J. Crowdfunding for medical expenses. Can Med Assoc J 2012 Feb 7;184(2):E123-E124 [FREE Full text] [doi: 10.1503/cmaj.109-4084] [Medline: 22231688]
8. Snyder J, Mathers A, Crooks VA. Fund my treatment!: a call for ethics-focused social science research into the use of crowdfunding for medical care. Soc Sci Med 2016 Nov;169:27-30. [doi: 10.1016/j.socscimed.2016.09.024] [Medline: 27665200]
9. The Statue of Liberty and America's Crowdfunding Pioneer. BBC News. 2013. URL: https://www.bbc.com/news/magazine-21932675 [accessed 2020-06-12]
10. Steinberg D. The Kickstarter Handbook: Real-Life Crowdfunding Success Stories. Philadelphia, PA: Quirk Books; 2012.
11. Goldfarb A, Tucker CE. Privacy regulation and online advertising. Manag Sci 2011 Jan;57(1):57-71. [doi: 10.1287/mnsc.1100.1246]
12. Li S, Srinivasan K, Sun B. Internet auction features as quality signals. J Mark 2009 Jan;73(1):75-92. [doi: 10.1509/jmkg.73.1.075]
13. Spinelli S, Adams R. McGraw-Hill. New York, USA: McGraw-Hill; 2012.
14. Belleflamme P, Lambert T, Schwienbacher A. Individual crowdfunding practices. Venture Cap 2013 Oct;15(4):313-333. [doi: 10.1080/13691066.2013.785151]
15. Allison TH, Davis BC, Short JC, Webb JW. Crowdfunding in a prosocial microlending environment: examining the role of intrinsic versus extrinsic cues. Enterp Theory Pract 2014 Apr 21;39(1):53-73. [doi: 10.1111/etap.12108]
16. About EquityNet. EquityNet: Leading Business Funding Platform. 2014. URL: https://www.equitynet.com/about.aspx [accessed 2020-06-12]

XSL•FO
RenderX

17. Burtch G, Chang J. Reducing Medical Bankruptcy Through Crowdfunding: Evidence From Giveforward. In: 35th International Conference on Information Systems: Building a Better World Through Information Systems. 2014 Presented at: ICIS'14; December 14-17, 2014; Auckland, New Zealand. [doi: 10.17705/4icis]
18. Gómez M. Crowdfunded justice: on the potential benefits and challenges of crowdfunding as a litigation financing tool. USFL Rev 2015;49:307 [FREE Full text]
19. Perry R. Crowdfunding civil justice. SSRN J 2018;59:1357-1395. [doi: 10.2139/ssrn.3041129]
20. Snyder J. GoFundMe and Other Crowdfunding Medical Campaigns Pose Ethical Dilemma. CBC. 2015. URL: https://www.cbc.ca/news/canada/british-columbia/gofundme-and- [accessed 2020-06-12]
21. Bagheri A, Chitsazan H, Ebrahimi A. Crowdfunding motivations: a focus on donors' perspectives. Technol Forecast Soc Change 2019 Sep;146:218-232. [doi: 10.1016/j.techfore.2019.05.002]
22. Fan-Osuala O, Zantedeschi D, Jank W. Using past contribution patterns to forecast fundraising outcomes in crowdfunding. Int J Forecast 2018 Jan;34(1):30-44. [doi: 10.1016/j.ijforecast.2017.07.003]
23. Kuppuswamy V, Bayus BL. Crowdfunding creative ideas: the dynamics of project backers in kickstarter. SSRN J 2014:-. [doi: 10.2139/ssrn.2234765]
24. Pollack J, Bosse DA. When do investors forgive entrepreneurs for lying? J Bus Ventur 2014 Nov;29(6):741-754 [FREE Full text] [doi: 10.1016/j.jbusvent.2013.08.005]
25. Clauss T, Breitenecker RJ, Kraus S, Brem A, Richter C. Directing the wisdom of the crowd: the importance of social interaction among founders and the crowd during crowdfunding campaigns. Econ Innov New Technol 2017 Nov 9;27(8):709-729 [FREE Full text] [doi: 10.1080/10438599.2018.1396660]
26. Paschen J. Choose wisely: crowdfunding through the stages of the startup life cycle. Bus Horiz 2017 Mar;60(2):179-188 [FREE Full text] [doi: 10.1016/j.bushor.2016.11.003]
27. Belleflamme P, Lambert T, Schwienbacher A. Crowdfunding: tapping the right crowd. J Bus Ventur 2014 Sep;29(5):585-609. [doi: 10.1016/j.jbusvent.2013.07.003]
28. Mariani A, Annunziata A, Aprile MC, Nacchia F. Crowdfunding and wine business: some insights from Fundovino experience. Wine Econ Policy 2017 Jun;6(1):60-70 [FREE Full text] [doi: 10.1016/j.wep.2017.02.001]
29. Mollick E. The dynamics of crowdfunding: an exploratory study. J Bus Ventur 2014 Jan;29(1):1-16 [FREE Full text] [doi: 10.1016/j.jbusvent.2013.06.005]
30. Ordanini A, Miceli L, Pizzetti M, Parasuraman A. Crowd‐funding: transforming customers into investors through innovative service platforms. J Serv Manag 2011 Aug 9;22(4):443-470 [FREE Full text] [doi: 10.1108/09564231111155079]
31. Colombo MG, Franzoni C, Rossi-Lamastra C. Internal social capital and the attraction of early contributions in crowdfunding. Entrep Theory Pract 2014 Aug 13;39(1):75-100 [FREE Full text] [doi: 10.1111/etap.12118]
32. Kraus S, Richter C, Brem A, Cheng C, Chang M. Strategies for reward-based crowdfunding campaigns. J Innov Knowl 2016 Jan;1(1):13-23 [FREE Full text] [doi: 10.1016/j.jik.2016.01.010]
33. Tomczak A, Brem A. A conceptualized investment model of crowdfunding. Ventur Capit 2013 Oct;15(4):335-359. [doi: 10.1080/13691066.2013.847614]
34. Valančienė L, Jegelevičiūtė S. Valuation of crowdfunding: benefits and drawbacks. Econ Manag 2013 Jun 17;18(1):- [FREE Full text] [doi: 10.5755/j01.em.18.1.3713]
35. Zhang H, Chen W. Backer motivation in crowdfunding new product ideas: is it about you or is it about me? J Prod Innov Manag 2018 Oct 27;36(2):241-262 [FREE Full text] [doi: 10.1111/jpim.12477]
36. Brem A, Bilgram V, Marchuk A. How crowdfunding platforms change the nature of user innovation – from problem solving to entrepreneurship. Technol Forecast Soc Chang 2019 Jul;144:348-360 [FREE Full text] [doi: 10.1016/j.techfore.2017.11.020]
37. Forbes H, Schaefer D. Guidelines for successful crowdfunding. Procedia 2017;60:398-403 [FREE Full text] [doi: 10.1016/j.procir.2017.02.021]
38. Gerber E, Hui J. Crowdfunding: motivations and deterrents for participation. ACM Trans Comput-Hum Interact 2013 Dec;20(6):1-32 [FREE Full text] [doi: 10.1145/2530540]
39. Harrison R. Crowdfunding and the revitalisation of the early stage risk capital market: catalyst or chimera? Ventur Capit 2013 Oct;15(4):283-287 [FREE Full text] [doi: 10.1080/13691066.2013.852331]
40. Ahlers G, Cumming D, Günther C, Schweizer D. Signaling in equity crowdfunding. Entrep Theory Pract 2015 Mar 24;39(4):955-980 [FREE Full text] [doi: 10.1111/etap.12157]
41. Kim J, Newberry PW, Qiu C. An empirical analysis of a crowdfunding platform. SSRN J 2010:1. [doi: 10.2139/ssrn.2672075]
42. Short J, Ketchen DJ, McKenny AF, Allison RH, Ireland RD. Research on crowdfunding: reviewing the (very recent) past and celebrating the present. Entrep Theory Pract 2017 Feb 7;41(2):149-160 [FREE Full text] [doi: 10.1111/etap.12270]
43. Beaulieu T, Sarker S, Sarker S. A conceptual framework for understanding crowdfunding. Commun Assoc Inf Syst 2015;37:- [FREE Full text] [doi: 10.17705/1cais.03701]
44. Bretschneider U, Leimeister JM. Not just an ego-trip: exploring backers' motivation for funding in incentive-based crowdfunding. J Strateg Inf Syst 2017 Dec;26(4):246-260 [FREE Full text] [doi: 10.1016/j.jsis.2017.02.002]
45. Davidson R, Poor N. Factors for success in repeat crowdfunding: why sugar daddies are only good for Bar-Mitzvahs. Inf Commun Soc 2015 Oct 19;19(1):127-139 [FREE Full text] [doi: 10.1080/1369118x.2015.1093533]

XSL•FO
RenderX

46. Kromidha E, Robson P. Social identity and signalling success factors in online crowdfunding. Entrep Regional Dev 2016 Jun 29;28(9-10):605-629. [doi: 10.1080/08985626.2016.1198425]
47. Wang S, Jank W, Shmueli G, Smith P. Modeling price dynamics in ebay auctions using differential equations. J Am Stat Assoc 2012 Jan;103(483):1100-1118. [doi: 10.1198/016214508000000670]
48. Wei Shi S. Crowdfunding: designing an effective reward structure. Int J Mark Res 2018 Jan 12;60(3):288-303. [doi: 10.1177/1470785317744113]
49. Xiao S, Yue Q. Investors' inertia behavior and their repeated decision-making in online reward-based crowdfunding market. Decis Support Syst 2018 Jul;111:101-112. [doi: 10.1016/j.dss.2018.05.005]
50. Chen S, Thomas S, Kohli C. What really makes a promotional campaign succeed on a crowdfunding platform? J Adv Res 2016 Feb 12;56(1):81-94. [doi: 10.2501/jar-2016-002]
51. Choy K, Schlagwein D. Crowdsourcing for a better world. Inf Technol People 2016 Mar 7;29(1):221-247. [doi: 10.1108/itp-09-2014-0215]
52. da Cruz JV. Beyond financing: crowdfunding as an informational mechanism. J Bus Ventur 2018 May;33(3):371-393. [doi: 10.1016/j.jbusvent.2018.02.001]
53. Giudici G, Guerini M, Rossi-Lamastra C. Reward-based crowdfunding of entrepreneurial projects: the effect of local altruism and localized social capital on proponents' success. Small Bus Econ 2017 Jan 7;50(2):307-324. [doi: 10.1007/s11187-016-9830-x]
54. Gorczyca M, Hartman RL. The new face of philanthropy: the role of intrinsic motivation in millennials' attitudes and intent to donate to charitable organizations. J Nonprofit Public Sect Mark 2017 Sep 6;29(4):415-433. [doi: 10.1080/10495142.2017.1326349]
55. Jian L, Shin J. Motivations behind donors' contributions to crowdfunded journalism. Mass Commun Soc 2015 Feb 10;18(2):165-185. [doi: 10.1080/15205436.2014.911328]
56. Comparing The Top Online Fundraising And Crowdfunding Platforms. Crowdfunding: Best Fundraising Sites Comparison. 2019. URL: https://www.crowdfunding.com/ [accessed 2020-06-12]
57. Global Crowdfunding Market to Exceed USD 28.77 Billion by 2025. Associated Press News. 2019. URL: https://apnews.com/Newswire/278cd583cea2d5d4582d07495eb9208c [accessed 2020-06-12]
58. Lindsay D. Local Governments and Nonprofits Test Crowdfunding for Civic Projects. The Chronicle of Philanthropy. 2015. URL: https://www.philanthropy.com/article/LocalGovernments-and/152005 [accessed 2020-06-12]
59. Davies R. Three provocations for civic crowdfunding. Inf Commun Soc 2014 Dec 23;18(3):342-355. [doi: 10.1080/1369118x.2014.989878]
60. Lee CH, Zhao JL, Hassna G. Government-incentivized crowdfunding for one-belt, one-road enterprises: design and research issues. Financ Innov 2016 Jan 26;2(1):2. [doi: 10.1186/s40854-016-0022-0]
61. Barclay E. The Sick Turn To Crowd Funding To Pay Medical Bills. National Public Radio. 2012. URL: https://www.npr.org/sections/health-shots/2012/10/23/163489063/the-sick-turn-to- [accessed 2020-06-12]
62. Bourque A. Is Crowdfunding Your Medical Expenses Risky or Reasonable? HuffPost. 2017. URL: https://www.huffpost.com/entry/is-crowdfunding-your-medi_b_7088486 [accessed 2020-06-12]
63. Gustke C. Managing Health Costs With Crowdfunding. The New York Times. 2015. URL: https://www.nytimes.com/2015/01/31/business/managing-health-costs-with-crowdfunding.html [accessed 2020-06-12]
64. Park A. Crowdfunding a cure: the sick are getting strangers to pay their medical bills. Time 2012 Dec 3;180(23):22. [Medline: 23326967]
65. Romm C. Is It Fair to Ask the Internet to Pay Your Hospital Bill? The Atlantic. 2015. URL: https://www.theatlantic.com/health/archive/2015/03/is-it-fair-to-ask-the-internet-to-pay-your-hospital-bill/387577/ [accessed 2020-06-12]
66. Bassani G, Marinelli N, Vismara S. Crowdfunding in healthcare. J Technol Transf 2018 Apr 20;44(4):1290-1310. [doi: 10.1007/s10961-018-9663-7]
67. del Savio L. Mass communication and societythe place of crowdfunding in the discovery of scientific and social value of medical research. Bioethics 2017 Jun;31(5):384-392. [doi: 10.1111/bioe.12339] [Medline: 28182288]
68. Renwick MJ, Mossialos E. Crowdfunding our health: economic risks and benefits. Soc Sci Med 2017 Oct;191:48-56 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.035] [Medline: 28889030]
69. Dragojlovic N, Lynd LD. Crowdfunding drug development: the state of play in oncology and rare diseases. Drug Discov Today 2014 Nov;19(11):1775-1780. [doi: 10.1016/j.drudis.2014.06.019] [Medline: 24973645]
70. Vance A. Crowdfunded Searches for Medical Miracles. Bloomberg. 2012. URL: https://www.bloomberg.com/news/articles/2012-08-02/crowdfunded-searches-for-medical-miracles [accessed 2020-06-12]
71. Simon R. Crowdfunding Sites Like GoFundMe and YouCaring Raise Money—and Concerns. The Wall Street Journal. 2016. URL: https://www.wsj.com/articles/crowdfunding-sites-like-gofundme-and-youcaring-raise-moneyand-concerns-1456775949 [accessed 2020-06-12]
72. Raghupathi WR. An overview of health analytics. J Health Med Informat 2013;4(3):-. [doi: 10.4172/2157-7420.1000132]
73. Sun G, Wu Y, Liang R, Liu S. A survey of visual analytics techniques and applications: state-of-the-art research and future challenges. J Comput Sci Technol 2013 Sep 17;28(5):852-867. [doi: 10.1007/s11390-013-1383-8]

XSL•FO
RenderX

74. Börner K, Bueckle A, Ginda M. Data visualization literacy: definitions, conceptual frameworks, exercises, and assessments. Proc Natl Acad Sci U S A 2019 Feb 5;116(6):1857-1864 [FREE Full text] [doi: 10.1073/pnas.1807180116] [Medline: 30718386]
75. Kohlhammer J, Keim D, Pohl M, Santucci G, Andrienko G. Solving problems with visual analytics. Procedia Comput Sci 2011;7:117-120 [FREE Full text] [doi: 10.1016/j.procs.2011.12.035]
76. Thomas J, Cook K. Illuminating the Path: The Research and Development Agenda for Visual Analytics. Washington, DC: United States Department of Homeland Security; 2005.
77. Keim DA. Visual exploration of large data sets. Commun ACM 2001 Aug;44(8):38-44. [doi: 10.1145/381641.381656]
78. Wong PC, Thomas J. Visual analytics. IEEE Comput Graph Appl 2004;24(5):20-21. [doi: 10.1109/mcg.2004.39] [Medline: 15628096]
79. Raghupathi W, Raghupathi V. An empirical study of chronic diseases in the United States: a visual analytics approach. Int J Environ Res Public Health 2018 Mar 1;15(3):431 [FREE Full text] [doi: 10.3390/ijerph15030431] [Medline: 29494555]
80. Mac R. How Anti-Kickstarter GoFundMe Became The Crowdfunding King With Causes Not Projects. Forbes. 2019. URL: https://www.forbes.com/sites/ryanmac/2015/09/24/gofundme-largest-crowdfunding-platform-1-billion-donations/#2a91b67c6629 [accessed 2020-06-23]
81. US Population Estimates Continue to Show the Nation?s Growth Is Slowing. United States Census Bureau. 2019. URL: https://www.census.gov/newsroom/press-releases/2019/popest-nation.html# [accessed 2019-12-30]
82. Coleman E. Millions of Americans Rely on Crowdfunding for Medical Expenses. Route Fifty - State and Local News and Analysis. -: - URL: https://www.routefifty.com/health-human-services/2020/02/crowdfunding-medical-expenses/163233/ [accessed 2020-02-20]

Edited by G Eysenbach; submitted 19.03.20; peer-reviewed by A Kulanthaivel, D Juzwishin, M Janodia, R Lee; comments to author 08.04.20; revised version received 07.05.20; accepted 03.06.20; published 03.07.20.

Please cite as:
Ren J, Raghupathi V, Raghupathi W
Understanding the Dimensions of Medical Crowdfunding: A Visual Analytics Approach
J Med Internet Res 2020;22(7):e18813
URL: https://www.jmir.org/2020/7/e18813
doi:10.2196/18813
PMID:32618573

©Jie Ren, Viju Raghupathi, Wullianallur Raghupathi. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study

Edmund WJ Lee[1,2,3], BA, MA, PhD; Mesfin Awoke Bekalu[1,2], PhD; Rachel McCloud[1,2], MPH, PhD; Donna Vallone[4,5,6], MPH, PhD; Monisha Arya[7,8], MPH, MD; Nathaniel Osgood[9], PhD; Xiaoyan Li[9], MSc; Sara Minsky[1], MPH; Kasisomayajula Viswanath[1,2], PhD

[1]Dana-Farber Cancer Institute, Boston, MA, United States

[2]Harvard TH Chan School of Public Health, Boston, MA, United States

[3]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[4]Schroeder Institute, Truth Initiative, Washington, DC, United States

[5]College of Global Public Health, New York University, New York, NY, United States

[6]Department of Health, Behavior and Society, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

[7]Baylor College of Medicine, Houston, TX, United States

[8]Center for Innovation in Quality, Effectiveness and Safety, Michael E DeBakey VA Medical Center, Houston, TX, United States

[9]Department of Computer Science, College of Arts and Science, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:
Edmund WJ Lee, BA, MA, PhD
Dana-Farber Cancer Institute
375 Longwood Avenue
Boston, MA, 02215
United States
Phone: 1 6178587988
Email: Edmund_Lee@dfci.harvard.edu

Abstract

Background: People from underserved communities such as those from lower socioeconomic positions or racial and ethnic minority groups are often disproportionately targeted by the tobacco industry, through the relatively high levels of tobacco retail outlets (TROs) located in their neighborhood or protobacco marketing and promotional strategies. It is difficult to capture the smoking behaviors of individuals in actual locations as well as the extent of exposure to tobacco promotional efforts. With the high ownership of smartphones in the United States—when used alongside data sources on TRO locations—apps could potentially improve tobacco control efforts. Health apps could be used to assess individual-level exposure to tobacco marketing, particularly in relation to the locations of TROs as well as locations where they were most likely to smoke. To date, it remains unclear how health apps could be used practically by health promotion organizations to better reach underserved communities in their tobacco control efforts.

Objective: This study aimed to demonstrate how smartphone apps could augment existing data on locations of TROs within underserved communities in Massachusetts and Texas to help inform tobacco control efforts.

Methods: Data for this study were collected from 2 sources: (1) geolocations of TROs from the North American Industry Classification System 2016 and (2) 95 participants (aged 18 to 34 years) from underserved communities who resided in Massachusetts and Texas and took part in an 8-week study using location tracking on their smartphones. We analyzed the data using spatial autocorrelation, optimized hot spot analysis, and fitted power-law distribution to identify the TROs that attracted the most human traffic using mobility data.

Results: Participants reported encountering protobacco messages mostly from store signs and displays and antitobacco messages predominantly through television. In Massachusetts, clusters of TROs (Dorchester Center and Jamaica Plain) and reported smoking behaviors (Dorchester Center, Roxbury Crossing, Lawrence) were found in economically disadvantaged neighborhoods. Despite the widespread distribution of TROs throughout the communities, participants overwhelmingly visited a relatively small number of TROs in Roxbury and Methuen. In Texas, clusters of TROs (Spring, Jersey Village, Bunker Hill Village, Sugar Land, and

XSL•FO
RenderX

Missouri City) were found primarily in Houston, whereas clusters of reported smoking behaviors were concentrated in West University Place, Aldine, Jersey Village, Spring, and Baytown.

Conclusions: Smartphone apps could be used to pair geolocation data with self-reported smoking behavior in order to gain a better understanding of how tobacco product marketing and promotion influence smoking behavior within vulnerable communities. Public health officials could take advantage of smartphone data collection capabilities to implement targeted tobacco control efforts in these strategic locations to reach underserved communities in their built environment.

(J Med Internet Res 2020;22(7):e17451) doi:10.2196/17451

KEYWORDS

mobile health; mobile phone; tobacco use; big data; spatial analysis; data science

Introduction

Background

Tobacco use is a major risk factor for lung cancer and premature morbidity [1] and also one of the leading preventable causes of death in the United States [2,3]. Approximately 480,000 deaths (1 in 5 deaths) in the United States annually could be attributed to smoking or tobacco consumption [4]. Although tobacco use is a prevalent public health problem in the United States, research has shown that the health and economic burden of tobacco use disproportionately affect underserved communities and that these communities have not benefitted from tobacco control efforts as much as others [5]. Research has also shown that the density of tobacco retail outlets (TROs) is higher in low-income neighborhoods [6] as well as in communities with higher percentages of ethnic minorities [7,8]. In addition, people from lower socioeconomic positions (SEPs) are often the target of the tobacco industry's advertising, including the places of their residence [9]. The presence of TROs, together with a disproportionate exposure to protobacco messages, is associated with smoking behaviors and may attenuate the attempts of smokers to quit by allowing easy access to tobacco products as well as encouraging impulse purchases by providing environmental cues for smoking [10].

Although many studies have examined how factors such as the proximity of a residence to TROs and exposure to protobacco messages are related to smoking attitudes and initiation, smoking prevalence, and even hospital admissions [3,11-14], few studies have investigated the exposure and awareness of antitobacco messaging within underserved communities, particularly in relation to TROs and across media platforms. For instance, what are the media platforms where underserved communities are exposed to anti- or protobacco messages? Are people strategically exposed to antitobacco messages near locations such as TROs or places they are most likely to smoke or other places? To date, research on the extent of such antitobacco and protobacco message exposure is limited by its reliance on aggregated cross-sectional self-reported survey data.

The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts

With the ubiquity of smartphone ownership, particularly within vulnerable communities, data from smartphone apps provide a potential data collection mechanism for informing health policy makers where to focus antitobacco messaging efforts [15,16], especially when they are used to complement traditional data sources such as the location of TROs. After all, smartphone penetration in the United States is high, with about 81% of the population owning a smartphone; the smartphone ownership figures are also high for underserved communities, such as people living in rural areas (71%), those making less than US $30,000 annually (71%), and in minority communities (approximately 79% to 80%) [17]. Health promotion organizations and policy makers could take advantage of the contextual information provided by smartphones to identify strategic areas to help ensure adequate exposure to antitobacco messages [18].

Smartphones enable researchers to *passively* engage in data collection at-scale within people's naturalistic environments. For instance, by enabling geolocation tracking with the explicit consent of users, smartphone apps can collect temporally ordered information on the precise locations they visited or the paths they have taken, without being intrusive. The ability to track the behaviors of those in underserved communities in situ is a huge advantage over traditional survey methods that rely on self-reported recall [19]. Methods based on recall are limited in that participants may not remember all locations they have visited accurately, or they may omit details for the sake of social desirability (eg, underreporting of places they may perceive as undesirable). Geolocation tracking can provide valuable information for health organizations in helping to strategically target antitobacco messaging efforts by accurately identifying where protobacco messaging is being encountered. For example, the mobility or path data that outline how individuals move through their communities would be useful for identifying the most popular TROs and other place-based platforms where messages are being aired.

This type of smartphone data collection might also provide opportunities for health organizations to partner with underserved communities for *participatory science* efforts [20]. Although collecting data from the underserved communities can be extremely difficult [21], the use of smartphones could circumvent this problem as it adds a minimum logistical burden given the role of smartphones in their lives. The data collection could employ ecological momentary assessment (EMA) techniques, which assess particular events in the lives of subjects at periodic intervals, such as smoking behavior or exposure to specific types of messages, which are automatically prompted. These data, together with geolocation of smoking-related behaviors, could be used to map smoking *hot spots* [22]—defined as locations where there are non-random observed patterns of clustering—which are areas where there is a

XSL•FO
RenderX

statistically significant clustering of respondents who report smoking in the same area.

These types of data collection efforts can facilitate the proactive reporting of exposure to tobacco messages that have high temporal specificity and are capable of capturing details of even ephemeral exposures (eg, a photo of an advertisement on a rotating billboard or as part of a video at a gas station and radio advertisements). Insights into where and how anti- and protobacco messages are reaching those in underserved communities can assist tobacco control practitioners and policy makers in helping reduce the disproportionate burden of tobacco use within vulnerable communities.

Objectives of the Study

This study aimed to examine how smartphone app data collection could complement existing data sources to help inform tobacco control efforts for underserved communities. There are three specific objectives of this paper. First, we sought to identify if there were concentrations of TROs in Massachusetts and Texas. Second, using both passive (ie, geolocations) and active (ie, self-reports) data, we aimed to identify (1) the most popular TROs, denoted by a small number of TROs that attracted the most human traffic; (2) the areas in which participants were most likely to smoke; and (3) the locations where the participants reported exposure to tobacco messages and where the concentrations of pro- and antitobacco messages were. Third, we drew suggestions for tobacco control based on our results.

Methods

Study Design and Recruitment

To address the objectives of our study, we conducted a small-scale feasibility test using a smartphone app in underserved communities in Massachusetts and Texas. We have chosen to conduct the study in these 2 states, given the diversity in tobacco control policy implementation, where Massachusetts had stricter tobacco laws as compared with Texas [23,24]. Ethics approval was obtained from the respective institutional review boards (IRBs) of Harvard University, Baylor College of Medicine, and the University of Saskatchewan after extensive review, which ensured that adequate layers of protection were in place for our participants. Upon receiving the IRB approvals, we recruited 95 participants (smokers and nonsmokers) aged 18 to 34 years who resided in different cities within Massachusetts and Texas to participate in our 8-week smartphone tobacco tracking study.

Participants were required to meet the following criteria: (1) existing Android smartphone users with a data plan (although we covered the costs of their plan for the duration of the study) or would be willing to change their primary phone to a study-compatible phone, (2) consented to download a location-tracking smartphone app called *Ethica* and to keep their location-tracking feature switched on for the duration of the study, and (3) were willing to complete a pretest at the start of the study and a posttest at the end of the study as well as respond to EMAs that would be pushed to them. Ethica is a smartphone app designed to collect sensor-based data (eg, geolocations, accelerometry, and electrodermal activity) as well as contextual self-reports (eg, EMAs).

Once the participants downloaded Ethica (assisted by study staff) and registered using their email and a password, the study staff helped ensure that the location-tracking feature on their phones was enabled. The participants were then asked to complete the pretest via the app on their smartphone. This pretest contained questions pertaining to demographics, smoking status, number of cigarettes smoked in the past 30 days, and other related smoking attitudinal and behavioral questions. The location-tracking app began collecting geolocation data once registration was complete. At the end of the study, the participants completed a similar posttest.

Profile of Participants

Among the 95 participants (49 females, 42 males, 4 nonresponse), 51 were from Massachusetts and 43 were from Texas, with 1 nonresponse. Of all the 95 participants, when asked if they were of Hispanic, Latino, or Spanish origin, 54 (57%) reported they were “not of Hispanic, Latino, or Spanish origin”; 7 (7%) were “Mexican, Mexican American, Chicano”; 10 (11%) were “Puerto Rican”; 12 (13%) were “Dominican”; 1 (1%) was “Cuban”; 6 (6%) were “another Hispanic, Latino, or Spanish origin (eg, Guatemalan, Salvadoran, Honduran, Nicaraguan, Panamanian, Colombian, Venezuelan, Peruvian)”; and 5 (5%) with no response. In terms of race, 37/95 (39%) of our participants identified as “Black or African American”, 33/95 (35%) as “White”, and the rest identified themselves as a combination of different ethnic groups (eg, “American Indian or Alaska Native”, “Asian”, “Native Hawaiian”, or “other Pacific Islander”).

The median total combined household income was between US $20,000 and US $29,000 (from 1 [US $10,000] to 9 [≥US $75,000]; median 3.00 [US $20,000 and US $29,000]; SD 2.11), and the median education status was having *some college* (1 [completed grade school or less] to 8 [completed graduate or professional school after college]; median 5.00 [some college]; SD 1.32). In total, 53 participants self-identified as smokers, whereas 41 were nonsmokers, with 1 nonresponse. The participants were also asked to report the total number of cigarettes they smoked in the past 30 days (mean 11.3, SD 13.4).

Data Management and Processing

Geolocations of Tobacco Retail Outlets

The geolocations of TROs were obtained from the North American Industry Classification System (NAICS). Developed by the Office of Management and Budget, NAICS is the federal standard business classification system based on the primary activities of businesses. We identified and extracted records for 252 TROs in Massachusetts and 1422 in Texas, based on the NAICS classification of *Tobacco Stores*. For this study, we chose to focus our analysis on TROs that were solely cigars, cigarettes, and tobacco dealers and retailers or smoke shops (NAICS8 code: Tobacco Stores) and excluded retailers whose primary descriptions were not in the area of tobacco sales (eg, beer, wine, and liquor stores; convenience stores; and gasoline stations with convenience stores), even though they might sell tobacco products. The reason for excluding these stores was

XSL•FO
RenderX

that people might pass by or linger at these places because of reasons (eg, when shopping for groceries) other than tobacco purchase and consumption.

Participants' Geolocations

Ethica collected approximately 31 million time-stamped geolocations from all the participants recorded in millisecond intervals. To increase the reliability of the data (eg, as there were multiple geolocations of individuals recorded when they were stationary), we collapsed participants' geolocations into multiple 10-min time intervals and extracted the most accurate and representative longitude and latitude locations for each interval. In total, there were 279,840 geolocations of participants from Massachusetts and 227,991 geolocations for participants in Texas.

Geolocations of Smoking Behaviors

Questions regarding smoking behaviors were randomly administered 4 times a day to smokers via Ethica (between 8 AM and 9 PM on weekdays and between 10 AM and 9 PM on weekends). The geolocations of participants when they were smoking were captured from their responses to the question "Have you smoked in the past hour," in which they were asked to select from the following responses: (1) I smoked a cigarette in the past hour, (2) I smoked an electronic cigarette (e-cigarette) in the past hour, (3) I used another tobacco product in the past hour, (4) I am smoking a cigarette right now, (5) I am smoking an e-cigarette right now, (6) I am using another tobacco product right now, and (7) I have not smoked. To obtain the geolocations of the participants when they were smoking, we extracted the longitude and latitude of smokers at the time if they indicated that they were smoking a cigarette or an e-cigarette or using another tobacco product *right now.* A total of 10,393 smoking geolocations in Massachusetts and 10,187 in Texas were recorded.

Geolocations of Tobacco Message Exposure

Through Ethica, participants were able to take or upload photos of tobacco messages and advertisements they came across in their communities (eg, billboards, TROs) or on the internet. After this, they were prompted to answer an EMA survey where the participants were given the options to identify the messages as either antitobacco or protobacco and to report when they saw the message (where 1=I see it right now; 2=in the past hour; 3=in the past 1-5 hours; 4=more than 5 hours ago). The latitude and longitude of the photos and EMA surveys were logged using Ethica.

Statistical Analysis

Data were imported into ArcMap 10.6.1 for mapping and statistical analyses, where we conducted *spatial autocorrelation* and *optimized hot spot analysis* as well as *power-law analysis* in R studio (version 1.8383) to address all the study objectives. The global spatial autocorrelation was used to test for the presence of spatial variation in a given dataset [25], specifically in examining the correlation among data points that are close to one another and to determine if there is a nonrandom spatial clustering among data points that were in close proximity [26]. The global Moran index is a statistic that indicates the presence of statistically significant spatial clustering, which produces a number between –1 and +1. A negative value indicates the presence of negative spatial autocorrelation, which is the tendency for dissimilar values to be located together. On the other hand, positive values indicate the presence of positive spatial autocorrelation, where data points with similar values are clustered together [25]. If the presence of spatial clustering was detected, we then conducted the optimized hot spot analysis to examine the locations of the *hot spots* and *cold spots* of TROs, smoking, and antitobacco messages.

The power-law analysis aimed to test if there was an observable power-law distribution in the data. A power-law distribution is also known as a heavy tail distribution, where smaller values on the *x*-axis correspond to large values in the *y*-axis. In other words, in the context of this study, if a power-law distribution is observable, a small number of TROs would attract the most human traffic. This may suggest that certain TROs are more popular, or centrally located, such that people are more likely to pass by as compared with TROs located in obscure locations.

Results

Our first objective was to identify if there were concentrations of TROs (ie, TRO hot spots) in Massachusetts and Texas (see Figures 1-3). To do so, we conducted spatial autocorrelation on the geolocations of TROs and determined if there was a statistically significant spatial clustering of TROs. The results suggest that there was evidence of clustering of TROs in both Massachusetts (global Moran index=0.79; z=8.04; P<.001) and Texas (global Moran index=0.69; z=5.85; P<.001). Next, optimized hot spot analysis in ArcMap showed that there was a statistically significant clustering of TROs in the city of Boston, with the most significant clusters found in Dorchester Center, Jamaica Plain, and Hyde Park (z≥3.50; P<.001). The TRO hot spots in Texas were found in Houston, and they were in places such as Spring, Jersey Village, Bunker Hill Village, Sugar Land, and Missouri City.

XSL•FO
RenderX

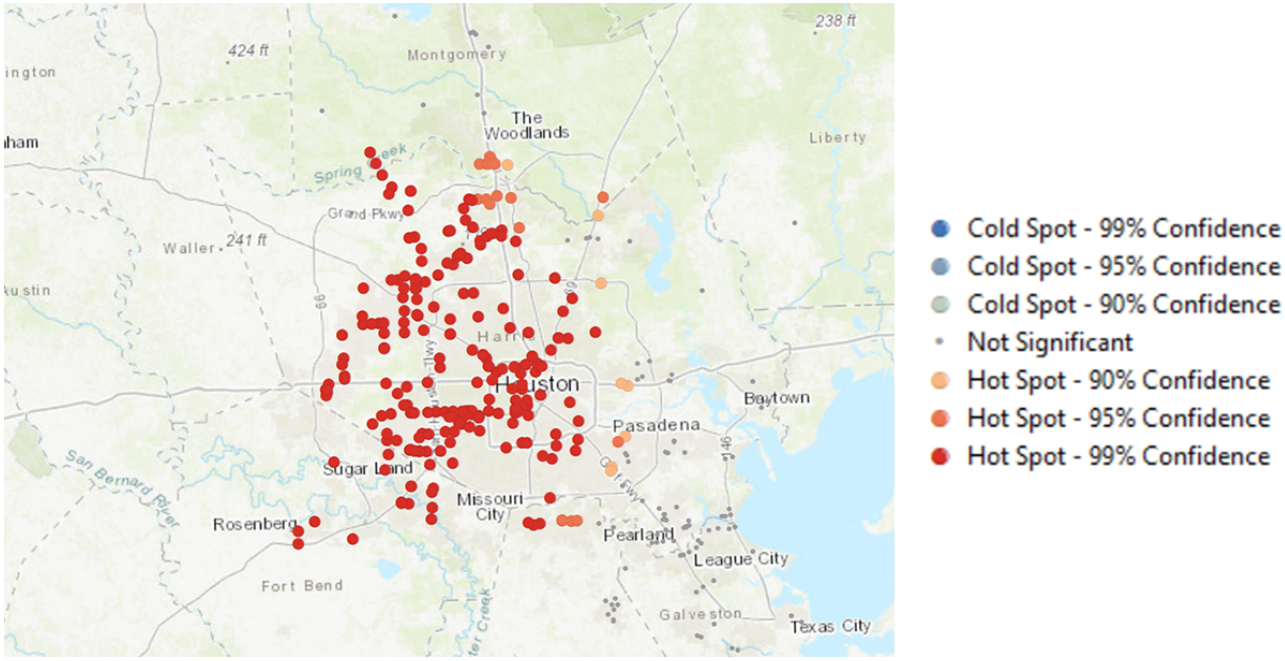
Figure 1. Hot spots of tobacco retail outlets in the state of Massachusetts.



Figure 2. Zoomed-in view of the most significant hot spots in the city of Boston.



Figure 3. Hot spots of tobacco retail outlets in the state of Texas.



XSL•FO
RenderX

Our second objective was to draw upon both passive (ie, geolocations) and active (ie, self-reports) data to identify (1) the most popular TROs, (2) the areas in which participants were most likely to smoke, and (3) the locations where participants reported being exposed to tobacco messages and where these concentrations of protobacco and antitobacco messages were.

To examine which were the most popular TROs—if a small number of TROs attracted the most traffic—we tested if a power-law distribution was observable by analyzing geolocations of participants near the TROs. To do so, we created a 100-m buffer around all the TROs in our dataset and performed a spatial join with all the geolocations that intersected within the buffer. The selection of a 100-m buffer was consistent with previous research [27]. We then exported the data to R studio to fit a power-law distribution in accordance with the steps recommended by Clauset et al [28]: (1) construct a discrete power-law distribution object, (2) estimate the x_{min} and exponent α of the power law and assign them to the power-law object, and (3) bootstrap to obtain the *P* value for the hypothesis test of if the data followed a power-law distribution. In the Kolmogorov-Smirnov test, the null hypothesis is that observations will follow a specific distribution, whereas the alternative hypothesis specifies that a set of distribution does not follow a specific distribution. As such, to claim that observations follow a power law distribution, the *P* value would have to be equal or more than .05 for the null hypothesis to be accepted, thereby indicating the presence of a power-law distribution. The analysis found marginal support for the power-law distribution (D=0.12; P=.05) for TROs in Massachusetts (Figure 4) but not for Texas. The top TROs that attracted the most human traffic from our sample were in the neighborhood of Roxbury in the city of Boston and Methuen, a city close to Boston.

Next, to identify the areas where participants were most likely to smoke (ie, smoking hot spots), we conducted spatial autocorrelation on geolocations where the participants reported their smoking behavior through the EMAs and determined if there was a statistically significant spatial clustering of smokers who reported smoking (Figure 5). The results suggest that there was significant clustering in both Massachusetts (global Moran index=0.29; z=34; P<.001) and Texas (global Moran index=0.25; z=63.5; P<.001). Next, we conducted optimized hot spot analysis, and the results showed that in Massachusetts, the heaviest smokers (based on the number of cigarettes smoked in the past 30 days) tended to report that they smoked in Dorchester Center, Roxbury Crossing, Lawrence, and Peabody (z≥2.84; P<.001). In Texas, the heaviest smokers tended to report smoking in West University Place, Aldine, Jersey Village, Spring, and Baytown (z≥2.67; P<.001).

To identify the locations where the participants reported being exposed to antitobacco and protobacco messages, we examined the photos taken by the participants through the app where they rated if the messages were either antitobacco or protobacco. In Massachusetts, there were 41 antitobacco and 48 protobacco messages reported (see Multimedia Appendix 1). The top 3 most frequent platforms for exposure to antitobacco messages in Massachusetts were on (1) television and others (19.5% each, 8/41 for television and 8/41 for others), (2) store sign or display (7/41, 17.1%), and (3) billboard/bus/train stop advertisements (6/41, 14.6%). The top 3 highest exposures to protobacco messages were on (1) store sign or display (27/48, 56.3%), (2) newspaper or magazine (7/48, 14.6%), and (3) website (5/48, 10.4%).

In Texas, there were 63 antitobacco and 43 protobacco messages (Multimedia Appendix 1). The top three highest exposures to antitobacco messages in Texas were (1) others (21/63, 33.3%), (2) television (14/63, 22.2%), and (3) store sign or display (7/63, 11.1%). The top 3 highest exposures to protobacco messages were on (1) store sign or display (25/43, 58.1%), (2) others (7/43, 16.3%), and (3) television (4/43, 9.3%).

Finally, we aimed to examine if there were spatial clustering of tobacco messages and if such clusters were located near TROs or smoking hot spots. We analyzed the data using spatial autocorrelation, and the results suggested that there was evidence of antitobacco message clustering in Massachusetts (global Moran index=0.28; z=1.89; P=.06) but not in Texas (global Moran index=–0.12; z=0.73; P=.07). We then conducted an optimized hot spot analysis for antitobacco messages in Massachusetts, and the results showed that the clustering of antitobacco messages (z=3.85; P<.001) only occurred in Lawrence in Massachusetts (Figure 6). There was no evidence of protobacco message clusters.

XSL•FO
RenderX

Figure 4. Cumulative mean of power-law analysis for traffic of tobacco retail outlets in Massachusetts.

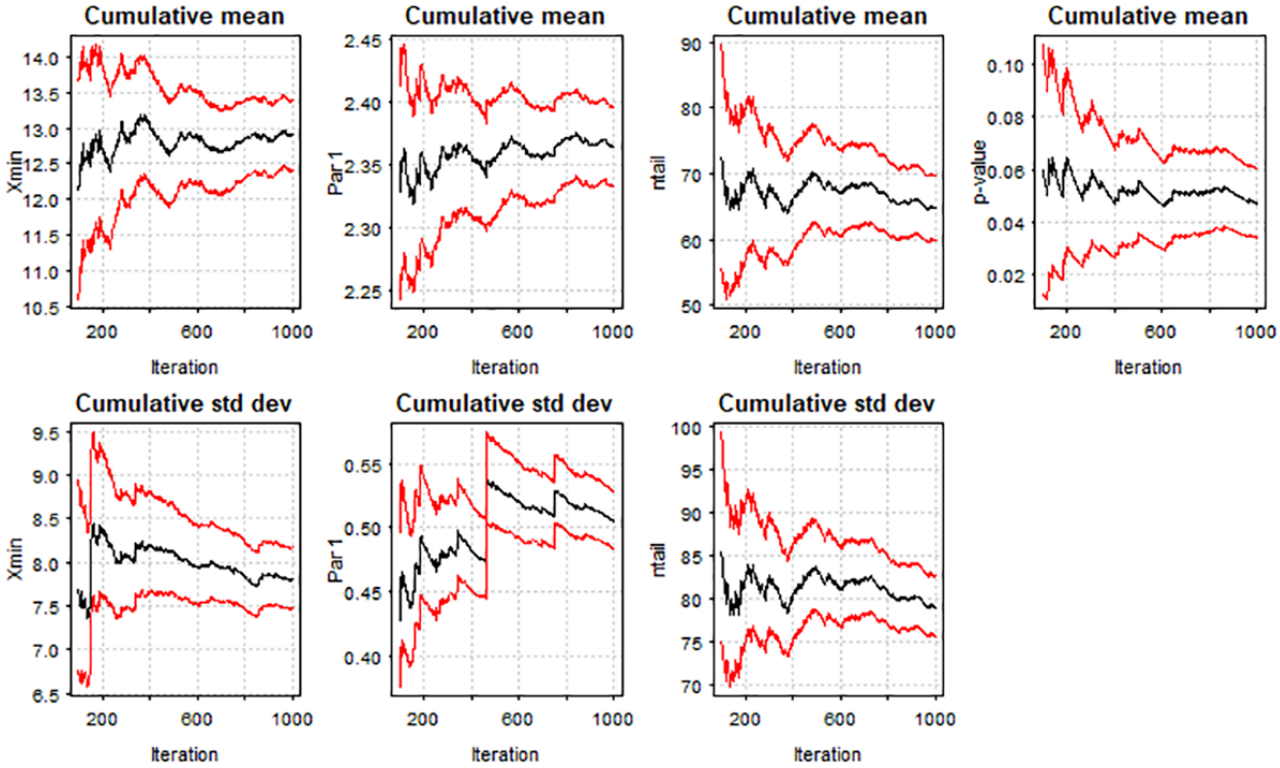
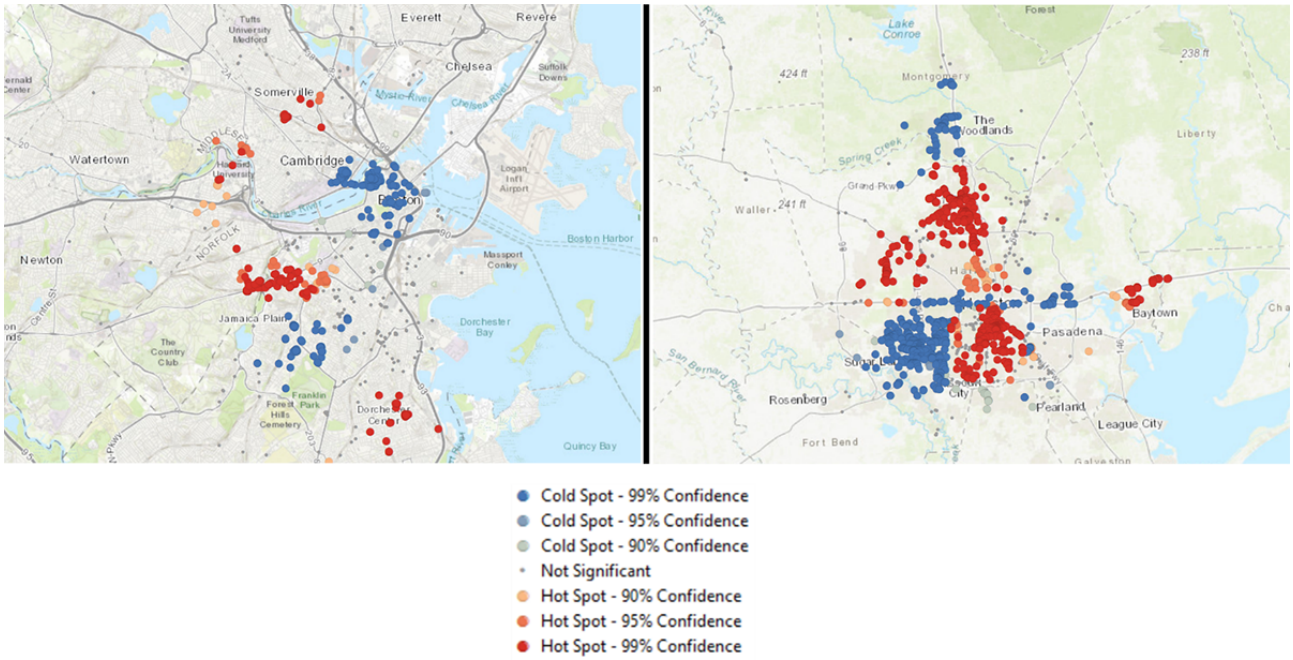
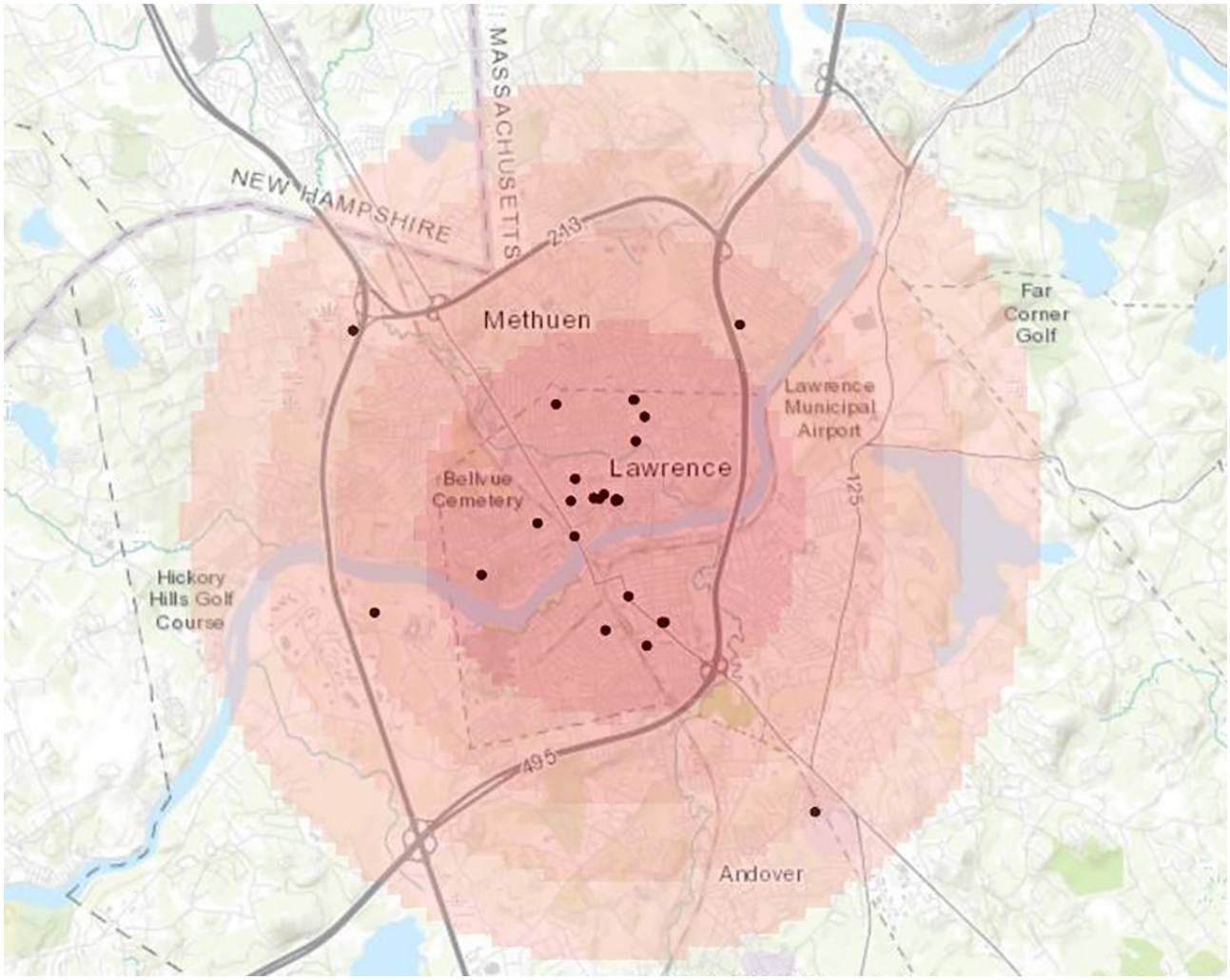


Figure 5. Clusters of smoking hot spots in Massachusetts (left) and Texas (right).



XSL•FO
RenderX

Figure 6. Clusters of reported antitobacco messages in Lawrence, Massachusetts.



Discussion

This study showcases how data from smartphone apps could significantly inform tobacco control communication efforts when used to complement existing data sources, such as the geolocations of TROs in our study obtained from NAICS [29]. Using these methods, we were able to identify the locations where our participants were exposed to tobacco messages (anti- and pro-), the specific TROs that attract the highest level of patrons, as well as areas where individuals were most likely to smoke. There are several notable findings from our results. The data showed that physical locations still matter more than online tobacco messages when considering where people were most likely to encounter external cues for tobacco use, such as locations of TROs and areas where protobacco messaging were reported. From NAICS data, there was a concentration of TROs in economically disadvantaged areas within Boston, such as the Dorchester Center and Jamaica Plain [30]. In Texas, our data showed that there was a high concentration of TROs across Houston.

The data from smartphones complement traditional tobacco surveillance data, such as population health surveys [31], in that smartphone data provide context to how access to tobacco products and exposure to marketing and promotional efforts may influence tobacco use behavior within underserved communities. For example, we found that in both Massachusetts and Texas, participants reported that they predominantly encountered protobacco messages at store signs or displays as compared with web-based sources. This is somewhat surprising considering the increasing concern about the influence of social media posts in the promotion of tobacco, either through user-generated content on social media or through targeted industry web-based advertising efforts [32,33]. In contrast, although our participants reported encountering antitobacco messages through web-based and offline sources, they were most likely to come across antitobacco messages on television. In addition to airing messages on mainstream media, public health officials should consider boosting efforts in placing antitobacco messages around TROs.

Another finding is that in both Massachusetts and Texas, participants reported encountering fewer antitobacco messages in newspapers or magazines as compared with protobacco messages. This is consistent with the findings from a recent study [34], which aimed to determine the extent of exposure to federal court-ordered antismoking advertisements—where tobacco companies were required to pay for these advertisements to correct smoking misinformation [35]—among a nationally representative sample of the adult population in the United States in 2018. The study found that the overall estimated exposure to antismoking advertisements was generally low

XSL•FO
RenderX

(40.6%), with the lowest exposure rates found among people aged 18 to 34 years (37.4%), those who had high school education or less (34.5%), those who earned less than US $35,000 annually (37.5%), and Hispanic smokers (42.2%). Although it was difficult to definitively pinpoint why our participants reported low exposure to antitobacco advertisements in newspapers, one plausible reason was that young people such as those in our sample may not be using print newspapers and magazines as much as the internet and social media [36,37], and thus, they would be less likely to come across antitobacco messages across traditional media platforms. In addition, research has documented that people from underserved communities were less likely to use newspapers as their primary news sources as compared with individuals from higher SEPs [38].

Second, this type of smartphone data collection allows one to target strategic areas for antitobacco message placement. For instance, in the state of Massachusetts, there was evidence of antitobacco messages only in Lawrence, which traditionally has a higher percentage of adult smokers and TROs per 1000 adults as compared with other parts of Massachusetts [39]. Although this was a positive step, there was a need for broader dissemination of antitobacco messages to reach other areas where popular TRO hot spots were found (Dorchester Center, Jamaica Plain, and Hyde Park), specific TROs (Roxbury and Methuen) with highest human traffic, as well as areas where smoking was concentrated (Dorchester Center, Roxbury Crossing, and Peabody).

Third, it is evident that the use of smartphone data to inform antitobacco messaging efforts for underserved communities is not a magic pill solution, as it would need concurrent *supply side* tobacco control regulations to be most effective. In Houston, the widespread prevalence of TROs remained problematic for targeted antitobacco messaging to be efficacious. In other words, effective and targeted antitobacco messaging in Texas would need to be accompanied by concurrent supply side solutions, such as restricting the number of TROs or increasing tobacco taxes.

Despite the study's significant strengths, there are limitations. First, we relied on a small sample of individuals from underserved communities, and the results would not be generalizable to the overall population. For example, the locations of popular smoking areas could be heavily influenced by the characteristics of our sample. Second, as in all studies that employ smartphone apps, the geolocations were only captured when the smartphones were operational. Third, this methodology does not guarantee that exposure to all antitobacco messages is captured. Participants might not be able to snap a picture of the antitobacco message on a billboard in time if they were driving or traveling in a car. Finally, we recognized that, similar to many smartphone tracking studies, there are issues pertaining to privacy because of the amount of data collected that may not relate directly to the study's objectives. Considering that we were working with underserved communities that were arguably more vulnerable than the general population, we prioritized the privacy protection of participants from the beginning of this study and took significant steps in communicating with our participants the privacy protection measures we have implemented.

At the policy and system architecture level, Ethica was built to be compliant with the General Data Protection Regulation requirements, which extended data protection for different types of health data collected from individuals [18,40]. In other words, our participants had the right to access and delete their own data. If the participants did not have the technical skills to do so, Ethica would provide technical support as needed. In addition, Ethica allowed the participants to request a copy of their data, and the support staff would provide them with a machine-readable file containing all the data collected about them. On a practical level, Ethica was designed in such a way that participants could snooze their study participation for some time. For instance, there was an incognito function where participants could pause data collection (eg, tracking of their geolocations) at any time they wanted.

Despite these limitations, this study presents a novel way of integrating passive and active data from smartphones with traditional tobacco surveillance information to help inform tobacco control efforts within underserved communities. We recommend that public health researchers continue to explore how to capitalize on big data from smartphones for tobacco control. For instance, future studies could extend our study by recruiting a larger sample of participants from different states and examining how fluctuations in emotions (captured by the EMA) could play a role in influencing tobacco use. Future research could also design smartphone-based interventions examining the optimal locations and time to administer antitobacco messages to people from underserved communities. In conclusion, smartphone data can inform tobacco control efforts in a powerful way, and health organizations and public health researchers should take advantage of this data revolution to strengthen tobacco control efforts to benefit the health of underserved communities [20].

Acknowledgments

This work was funded by the *Truth Initiative* (KV, principal investigator). The authors would like to thank the study coordinators from both Massachusetts (Carmenza Bruff) and Texas (Amanda Zhao and Anna Kimutis) for their efforts in recruitment and data collection as well as Mohammad Hashemian from *Ethica* for app-related technical assistance. In addition, the authors would like to acknowledge Dr Jill Kelly and Dr Matthew Wilson for providing technical guidance on ArcMap through the annual Geographical Information Systems Summer Institute 2019 organized by the Center for Geographic Analysis at Harvard University.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tobacco messages reported in Massachusetts (Table 1) and Texas (Table 2).
[DOCX File , 16 KB - jmir_v22i7e17451_app1.docx]

References

1. Fu SH, Jha P, Gupta PC, Kumar R, Dikshit R, Sinha D. Geospatial analysis on the distributions of tobacco smoking and alcohol drinking in India. PLoS One 2014;9(7):e102416 [FREE Full text] [doi: 10.1371/journal.pone.0102416] [Medline: 25025379]
2. Centers for Disease Control and Prevention. Health Effects of Cigarette Smoking URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm [accessed 2020-05-12]
3. Maloney EK, Cappella JN. Does vaping in e-cigarette advertisements affect tobacco smoking urge, intentions, and perceptions in daily, intermittent, and former smokers? Health Commun 2016;31(1):129-138. [doi: 10.1080/10410236.2014.993496] [Medline: 25758192]
4. Centers for Disease Control and Prevention. 2019. Current Cigarette Smoking Among Adults in the United States URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm#states [accessed 2020-04-23]
5. Sakuma KK, Felicitas-Perkins JQ, Blanco L, Fagan P, Pérez-Stable EJ, Pulvers K, et al. Tobacco use disparities by racial/ethnic groups: California compared to the United States. Prev Med 2016 Oct;91:224-232 [FREE Full text] [doi: 10.1016/j.ypmed.2016.08.035] [Medline: 27575315]
6. Tucker-Seeley RD, Bezold CP, James P, Miller M, Wallington SF. Retail pharmacy policy to end the sale of tobacco products: what is the impact on disparity in neighborhood density of tobacco outlets? Cancer Epidemiol Biomarkers Prev 2016 Sep;25(9):1305-1310 [FREE Full text] [doi: 10.1158/1055-9965.EPI-15-1234] [Medline: 27302724]
7. Lee J, Sun D, Schleicher N, Ribisl K, Luke D, Henriksen L. Inequalities in tobacco outlet density by race, ethnicity and socioeconomic status, 2012, USA: results from the ASPiRE study. J Epidemiol Community Health 2017 May;71(5):487-492 [FREE Full text] [doi: 10.1136/jech-2016-208475] [Medline: 28249990]
8. Fakunle D, Milam A, Furr-Holden C, Butler J, Thorpe R, LaVeist T. The inequitable distribution of tobacco outlet density: the role of income in two black mid-Atlantic geopolitical areas. Public Health 2016 Jul;136:35-40 [FREE Full text] [doi: 10.1016/j.puhe.2016.02.032] [Medline: 27076440]
9. Bekalu MA, Minsky S, Viswanath K. Beliefs about smoking-related lung cancer risk among low socioeconomic individuals: the role of smoking experience and interpersonal communication. Glob Health Promot 2019 Sep;26(3):88-93. [doi: 10.1177/1757975917732758] [Medline: 29110569]
10. Cantrell J, Anesetti-Rothermel A, Pearson JL, Xiao H, Vallone D, Kirchner TR. The impact of the tobacco retail outlet environment on adult cessation and differences by neighborhood poverty. Addiction 2015 Jan;110(1):152-161 [FREE Full text] [doi: 10.1111/add.12718] [Medline: 25171184]
11. Barnes R, Foster SA, Pereira G, Villanueva K, Wood L. Is neighbourhood access to tobacco outlets related to smoking behaviour and tobacco-related health outcomes and hospital admissions? Prev Med 2016 Jul;88:218-223. [doi: 10.1016/j.ypmed.2016.05.003] [Medline: 27154350]
12. Shortt NK, Tisch C, Pearce J, Richardson EA, Mitchell R. The density of tobacco retailers in home and school environments and relationship with adolescent smoking behaviours in Scotland. Tob Control 2016 Jan;25(1):75-82 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051473] [Medline: 25370699]
13. Cantrell J, Pearson JL, Anesetti-Rothermel A, Xiao H, Kirchner TR, Vallone D. Tobacco retail outlet density and young adult tobacco initiation. Nicotine Tob Res 2016 Feb;18(2):130-137 [FREE Full text] [doi: 10.1093/ntr/ntv036] [Medline: 25666816]
14. Loomis BR, Kim AE, Busey AH, Farrelly MC, Willett JG, Juster HR. The density of tobacco retailers and its association with attitudes toward smoking, exposure to point-of-sale tobacco advertising, cigarette purchasing, and smoking among New York youth. Prev Med 2012 Nov;55(5):468-474. [doi: 10.1016/j.ypmed.2012.08.014] [Medline: 22960255]
15. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
16. Baskerville NB, Struik LL, Dash D. Crush the crave: development and formative evaluation of a smartphone app for smoking cessation. JMIR Mhealth Uhealth 2018 Mar 2;6(3):e52 [FREE Full text] [doi: 10.2196/mhealth.9011] [Medline: 29500157]
17. Pew Research Center. 2019. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2020-04-23]
18. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. Front Commun 2020 Feb 27;5(11):1-10. [doi: 10.3389/fcomm.2020.00011]

XSL•FO
RenderX

19. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
20. Lee EW, Viswanath K. Big data in context: addressing the twin perils of data absenteeism and chauvinism in the context of health disparities research. J Med Internet Res 2020 Jan 7;22(1):e16377 [FREE Full text] [doi: 10.2196/16377] [Medline: 31909724]
21. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
22. Harris NL, Goldman E, Gabris C, Nordling J, Minnemeyer S, Ansari S, et al. Using spatial statistics to identify emerging hot spots of forest loss. Environ Res Lett 2017 Feb 7;12(2):024012. [doi: 10.1088/1748-9326/aa5a2f]
23. Truth Initiative. 2019. Tobacco Use in Texas 2019 URL: https://truthinitiative.org/research-resources/smoking-region/tobacco-use-texas-2019 [accessed 2020-04-23]
24. Truth Initiative. 2019. Tobacco Use in Massachusetts 2019 URL: https://truthinitiative.org/research-resources/smoking-region/tobacco-use-massachusetts-2019 [accessed 2020-04-23]
25. Haining R. Spatial autocorrelation. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Toronto, Canada: Elsevier; 2015:105-110.
26. Tsai P, Lin M, Chu C, Perng C. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. BMC Public Health 2009 Dec 14;9(1):464 [FREE Full text] [doi: 10.1186/1471-2458-9-464] [Medline: 20003460]
27. Lipperman-Kreda S, Morrison C, Grube JW, Gaidus A. Youth activity spaces and daily exposure to tobacco outlets. Health & Place 2015 Jul;34:30-33. [doi: 10.1016/j.healthplace.2015.03.013] [Medline: 25879915]
28. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM Rev 2009 Nov 4;51(4):661-703. [doi: 10.1137/070710111]
29. Watkins KL, Regan SD, Nguyen N, Businelle MS, Kendzor DE, Lam C, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. Nicotine Tob Res 2014 May;16(Suppl 2):S93-101 [FREE Full text] [doi: 10.1093/ntr/ntt135] [Medline: 24057995]
30. Boston Planning & Development Agency. 2014. Poverty in Boston URL: http://www.bostonplans.org/getattachment/f1ecaf8a-d529-40b6-a9bc-8b4419587b86 [accessed 2020-04-23]
31. Vereen RN, Westmaas JL, Bontemps-Jones J, Jackson K, Alcaraz KI. Trust of information about tobacco and e-cigarettes from health professionals versus tobacco or electronic cigarette companies: differences by subgroups and implications for tobacco messaging. Health Commun 2020 Jan;35(1):89-95. [doi: 10.1080/10410236.2018.1544875] [Medline: 30422690]
32. Allem J, Dharmapuri L, Leventhal A, Unger J, Cruz TB. Hookah-related posts to Twitter from 2017 to 2018: thematic analysis. J Med Internet Res 2018 Nov 19;20(11):e11669 [FREE Full text] [doi: 10.2196/11669] [Medline: 30455162]
33. Chu K, Colditz JB, Primack BA, Shensa A, Allem J, Miller E, et al. JUUL: spreading online and offline. J Adolesc Health 2018 Nov;63(5):582-586 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.08.002] [Medline: 30348280]
34. Chido-Amajuoyi OG, Yu RK, Agaku I, Shete S. Exposure to court-ordered tobacco industry antismoking advertisements among US adults. JAMA Netw Open 2019 Jul 3;2(7):e196935 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6935] [Medline: 31298713]
35. Mathias T. Reuters. 2019. Tobacco Industry Anti-smoking Ads Reached Less Than Half of US Adults URL: https://www.reuters.com/article/us-health-tobacco-industry/tobacco-industry-anti-smoking-ads-reached-less-than-half-of-u-s-adults-idUSKCN1V91V2 [accessed 2020-04-23]
36. Lee EW, Ho SS, Lwin MO. Explicating problematic social network sites use: a review of concepts, theoretical frameworks, and future directions for communication theorizing. New Media Soc 2016 Oct 6;19(2):308-326. [doi: 10.1177/1461444816671891]
37. Lee EW, Ho SS, Lwin MO. Extending the social cognitive model—examining the external and personal antecedents of social network sites use among Singaporean adolescents. Comput Human Behav 2017 Feb;67:240-251. [doi: 10.1016/j.chb.2016.10.030]
38. Lee EW, Ho SS. The perceived familiarity gap hypothesis: examining how media atteThe perceived familiarity gap hypothesis: examining how media attention and reflective integration relate to perceived familiarity with nanotechnology in Singaporention and reflective integration relate to perceived familiarity with nanotechnology in Singapore. J Nanopart Res 2015 May 22;17(5):1-15. [doi: 10.1007/s11051-015-3036-z]
39. Department of Public Health Massachusetts. Tobacco Community Fact Sheet. Make Smoking History. 2019. URL: https://makesmokinghistory.org/wp-content/themes/makesmokinghistory/tafi/pdf/TAFI-Lawrence.pdf
40. Ethica. 2020. Is Ethica GDPR Compliant? URL: https://learn.ethicadata.com/faq/is-ethica-gdpr-compliant/ [accessed 2020-04-23]

Abbreviations

e-cigarette: electronic cigarette
GED: General Education Diploma
IRB: institutional review board

XSL•FO
RenderX

NAICS: North American Industry Classification System
SEP: socioeconomic position
TRO: tobacco retail outlet

Edited by G Eysenbach; submitted 13.12.19; peer-reviewed by E Mwashuma, I Contreras; comments to author 11.03.20; revised version received 20.03.20; accepted 21.03.20; published 07.07.20.

Please cite as:
Lee EWJ, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, Li X, Minsky S, Viswanath K
The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study
J Med Internet Res 2020;22(7):e17451
URL: https://www.jmir.org/2020/7/e17451
doi:10.2196/17451
PMID:32673252

©Edmund WJ Lee, Mesfin Awoke Bekalu, Rachel McCloud, Donna Vallone, Monisha Arya, Nathaniel Osgood, Xiaoyan Li, Sara Minsky, Kasisomayajula Viswanath. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Diagnosing Parkinson Disease Through Facial Expression Recognition: Video Analysis

Bo Jin[1], PhD; Yue Qu[1], MS; Liang Zhang[2], PhD; Zhan Gao[3], PhD

[1]Dalian University of Technology, Dalian, China

[2]Dongbei University of Finance and Economics, Dalian, China

[3]Beijing Haoyisheng Cloud Hospital Management Technology Ltd, Beijing, China

Corresponding Author:
Liang Zhang, PhD
Dongbei University of Finance and Economics
217 Jianshan St, Shahekou District
Dalian
China
Phone: 86 15524709655
Email: liang.zhang@dufe.edu.cn

Abstract

Background: The number of patients with neurological diseases is currently increasing annually, which presents tremendous challenges for both patients and doctors. With the advent of advanced information technology, digital medical care is gradually changing the medical ecology. Numerous people are exploring new ways to receive a consultation, track their diseases, and receive rehabilitation training in more convenient and efficient ways. In this paper, we explore the use of facial expression recognition via artificial intelligence to diagnose a typical neurological system disease, Parkinson disease (PD).

Objective: This study proposes methods to diagnose PD through facial expression recognition.

Methods: We collected videos of facial expressions of people with PD and matched controls. We used relative coordinates and positional jitter to extract facial expression features (facial expression amplitude and shaking of small facial muscle groups) from the key points returned by Face++. Algorithms from traditional machine learning and advanced deep learning were utilized to diagnose PD.

Results: The experimental results showed our models can achieve outstanding facial expression recognition ability for PD diagnosis. Applying a long short-term model neural network to the positions of the key features, precision and F1 values of 86% and 75%, respectively, can be reached. Further, utilizing a support vector machine algorithm for the facial expression amplitude features and shaking of the small facial muscle groups, an F1 value of 99% can be achieved.

Conclusions: This study contributes to the digital diagnosis of PD based on facial expression recognition. The disease diagnosis model was validated through our experiment. The results can help doctors understand the real-time dynamics of the disease and even conduct remote diagnosis.

(J Med Internet Res 2020;22(7):e18697) doi:10.2196/18697

KEYWORDS

Parkinson disease; face landmarks; machine learning; artificial intelligence

Introduction

The population overall is currently aging. While an aging population represents the triumph of medical and social advances over disease, it also presents daunting challenges. Age is a crucial parameter for the occurrence, development, and diagnosis of diseases. As age increases, the central nervous system's morphology, metabolism, and function undergo different degrees of decline, which results in certain neurological diseases [1]. One typical example is Parkinson disease (PD), which is caused by a decrease in dopamine secretion. PD, also known as tremor palsy, is a common neurodegenerative disease; the manifestations are mainly bradykinesia, myotonia, resting tremors, and unstable posture [2]. Clinical symptoms may also include nonmotor symptoms such as olfactory function decline, constipation, and depression. As a progressive disease, the various symptoms, both motor and nonmotor, become more serious as the disease course develops, and various

XSL•FO
RenderX

complications, such as the "on-off" phenomenon, dyskinesia, and decline in drug efficacy, often occur in the later stages. Patients with serious illness may be plagued by balance disorders, frozen gait, falls, and speech disorders, resulting in an inability to take care of themselves and a decrease in quality of life [3]. Famous people like Xiaoping Deng, Jingrun Chen, Jin Ba, and Muhammad Ali have all been deeply affected by PD. Hence, an accurate diagnosis and medication that works immediately are important.

In recent years, with the development of computer vision technology, facial image recognition has been used for disease diagnosis. In 2017, the National Human Genome Research Institute developed facial diagnosis software to identify whether a child has DiGeorge syndrome [4]. It is a very rare disease, with a pediatric incidence rate ranging from 1/6000 to 1/3000 worldwide. With such a wide range of morbidity, disease diagnosis can be difficult. However, all children with DiGeorge syndrome have clinically obvious facial features, which were utilized by the National Human Genome Research Institute for disease diagnosis. In 2016, FDNA Inc developed the Face2Gene system, which can help doctors diagnose genetic diseases via facial recognition [5]. Some relatively rare genetic diseases can even be discovered by comparing photos of patients with healthy people. This helps patients detect their disease and obtain treatment in a relatively short time.

For PD, the disease-specific facial expressions of patients have attracted researchers' attention. In the 1860s, Charcot first described the characteristics of "masked face" in patients with PD [6]. Based on years of research, "mask face" is considered one of the common symptoms of PD. Its symptoms involve a faceless and binocular gaze [7]. Facial expressions in humans are expressed by the superficial muscles of the face. These muscles are called the "facial muscles" or "expression muscles." Smiles and other facial expressions of patients with PD often appear to be unresponsive and have significantly smaller amplitude. The time it takes to form an expression is also extended. To many patients' families, the "mask face" symptom causes issues because it is an expression of sullenness. Therefore, exploring a diagnostic method using facial feature point recognition is promising. Note that detection of a speech disorder in patients with PD, which is also a noninvasive diagnostic method, has been shown to be effective [8,9]. Hence, facial expression recognition can be combined with speech disorder recognition to obtain a more comprehensive, multidirectional, noninvasive, remote diagnosis.

Several researchers have utilized machine learning and computer vision technology to explore the inner relationship in the "mask face." Bandini et al [10] proposed an automatic method in 2017 to analyze videos of facial expressions in patients with PD. They extracted the average distances between facial key points using face tracing. Then, they found that patients with PD have much smaller distances of facial movement. In contrast to video-based research, Rajnoha et al [11] designed an automatic detection method based on static facial images using convolutional neural network models. However, the best achieved accuracy was only 67.33%. In 2019, Langevin et al [12] designed the Parkinson's Analysis with Remote Kinetic-tasks framework to analyze PD characteristics, finding that facial features and motion features in the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) could be extracted from videos.

In order to explore facial expressions in humans and test whether they are accompanied by tremors, we need to extract the key points of the face. In this paper, we investigated the ability to diagnose PD by recognizing changes in key points of the face in a short video. At present, to the best of our knowledge, we are the first to utilize sequential changes in key points of the face to diagnose PD.

Methods

Data Collection

People with PD usually suffer from loss of facial expression on both sides of the face. Moreover, their eye movement is reduced, like wearing a mask. This symptom is usually called the "mask face" [13]. Motion retardation is one of the main symptoms of PD. When facial symptoms appear, movement in facial expressions is significantly reduced, accompanied by local tremor symptoms of the small tendon group. When patients with PD try to smile or form other expressions, the facial muscles move slowly and often show excessive expressions. Based on the main facial expression symptoms of patients with PD, we recorded the smiles of patients with PD. Each video was 5 seconds long, and we recorded 2-3 videos per patient. The expression of a smile can be used to distinguish a patient with PD from someone without PD by comparing the magnitude of the expression and trembling of the small muscles of the face.

Patients with PD at the First Affiliated Hospital of Dalian Medical University were recruited and provided video data. Patients provided written informed consent before collecting data. Data for the control group, or people without PD, were randomly collected from senior people who liked to exercise frequently. Finally, we collected data from 64 senior citizens, including 33 people with PD. Each person recorded a smile expression 3 times. The length of every video was 5 seconds. After filtering out the videos that could not be used, we finally collected 176 records. The label is the subject status (ie, whether he or she had PD), as shown in Figure 1.

XSL•FO
RenderX

Figure 1. Smiles of a patient with Parkinson disease (left) and person without Parkinson disease (right).



Acquisition Equipment

In order to make the captured video clear enough and not affect the experimental results due to human jitter, the videos were captured using a Canon 700D camera placed on a standard tripod.

Collection Plan

In the process of recording facial expressions, we showed emoticons, which were printed on photos, to the patients with PD and asked them to imitate the emoticons. This data collection process aimed to explore the vibration of the small muscle groups of the face to distinguish patients with PD from people without PD. The underlying reason for asking the participants to imitate an emoticon in a photo was to prevent the subjective emotions of the patients with PD from affecting the data collection. At the same time, it guaranteed consistency of data collection and ensured that each patient's understanding of a smile or anger was identical. The data collection process was as follows. First, we recorded the participant's gender, age, and history of other neurological diseases. Then, the subjects filled out the video data collection registration form, with full awareness of the PD detection research project. Table 1 displays an example of the information collected via the registration form. Third, the emoticon photos were given to the subjects, and they were asked to imitate them. We collected 3 videos of smiling facial expressions, each of which lasted for 5 seconds. After all the recordings were completed, the videos were classified into 2 classes: patients with PD and people without PD. Different folders were sorted according to the patient ID.

Table 1. Example data collected using the registration form to collect data via video of patients with Parkinson disease.

Patient Number	Age (years)	Gender	Confirmed	Other neurological disease	Length of disease (month)	Date of collection
1	60	Male	Yes	No	10	11/13/2017
2	55	Male	Yes	No	24	11/13/2017
3	60	Male	Yes	No	10	11/13/2017
4	63	Female	Yes	No	14	11/13/2017

Converting Video to Images

To obtain the participant's facial information, we split the recorded video into individual frames that were extracted directly from the video every 0.1 seconds. In this experiment, we used ffmpeg to implement this function. ffmpeg is a set of open source computer programs that convert digital audio and video into streams [14]. In practice, processing multiple videos in multiple folders involves the files' operation commands. We leveraged a subprocess module in the python programming environment that encapsulates the running terminal commands, which perfectly met our needs. The subprocess package mainly executes external commands and programs and uses the function *subprocess.call*() to call external commands. The video was converted into several images according to a preset time interval and stored in the specified folder for analysis.

Data Preprocessing

Face++ is a well-known service for face recognition and facial landmark detection. There are many interesting applications for using Face++, such as inferring the demographics of social media users from profile pictures [15]. We chose Face++ because of its outstanding ability to localize facial landmarks. Based on previous literature and market research on key points of faces, we found that the Face++ interface can provide 106 coordinate points to create very accurate faces at this stage.

Users simply log in to the Face++ official website and apply for an API key. By calling the "Detect API" interface while using the "POST" method and setting the *return_landmark* parameter value to 2, the system will return 106 key points of the human face, as shown in Figure 2. These 106 points cover most of the key points of a human's face, including the facial contour, eyebrows, eyes, nose, and mouth that can describe almost all facial expressions. The returned coordinates are numbers using pixels as the unit. The top left corner of the image is used as the origin point.

XSL•FO
RenderX

Figure 2. Face key points (n=106) returned by the Face++ interface.



The facial muscles of patients with PD tend to be more rigid than those of people without PD, which causes a smaller facial expression range. We obtained the magnitude of the patient's expression by calculating the coordinate positions' range. However, during the video recording process, we found that although the video equipment is stable, it is difficult to ensure the participant, especially a patient with PD, does not move. To capture a relatively accurate magnitude of the facial expression, we converted the absolute coordinates into relative coordinates. Through the variation of the relative coordinates, the magnitude of the change in the facial expression of the patient can be reflected, and the error caused by postural changes of the body can be avoided.

To transform absolute coordinates into relative coordinates, we used the midpoint between the inside corners of the eyes as the origin of the coordinate (0,0). Then, we set the line connecting the inside corners of the eye as the x axis and the line connecting the nose and the origin as the y axis. A non-Cartesian coordinate system could then be created. As for the units, the coordinates of the inside corners of both eyes were quantized as (–1,0) and (1,0), and the coordinates of the nose were quantized as (0,–1), as shown in Figure 3. The black coordinate system was used to record the position of the pixels in the image, so we called this the absolute coordinate system. The blue coordinate system represents the relative coordinate system.

Assuming that the unit vector along the x axis in the relative coordinate system is vector a (a_1, a_2) and the unit vector along the y axis direction is vector b (b_1, b_2), we only need to compute (x,y), as shown in Figure 4. In Figure 4, the absolute coordinates (m_1, n_1), (m_2, n_2), (a_1, a_2), (b_1, b_2) can be converted to the relative coordinates (0, 0), (x, y), (1, 0), (0, 1), respectively.

XSL•FO
RenderX

Figure 3. Relative face coordinate system, where the Cartesian, or absolute, coordinate system is represented by the black coordinate system, which was used to record the position of pixels in the image, and the non-Cartesian, or relative, coordinate system is represented by the blue coordinate system, which was used to record the relative position of key points on the face.

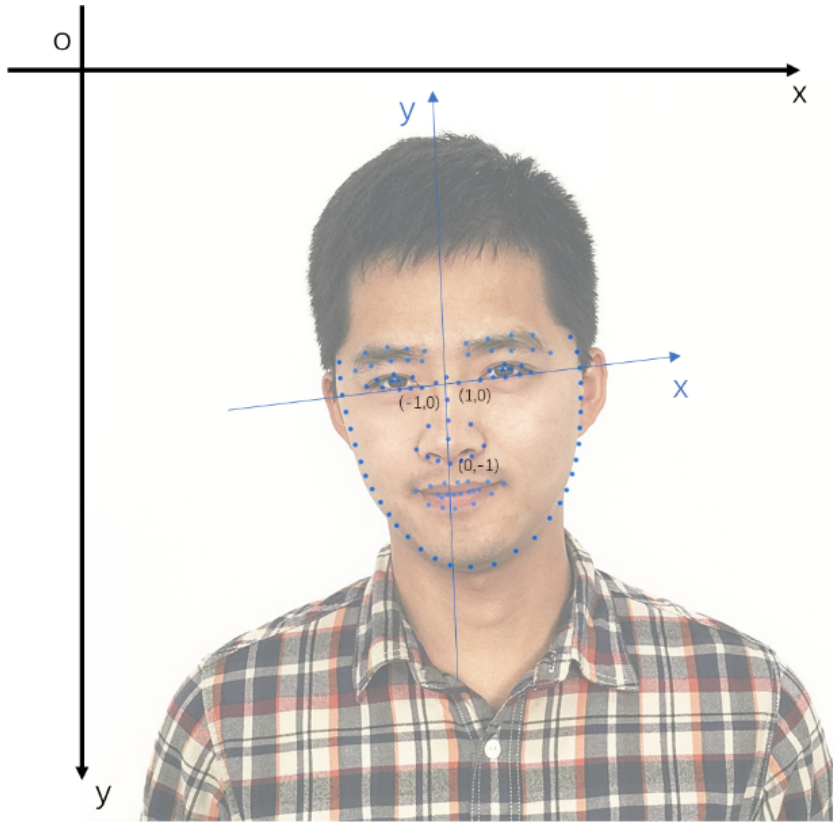
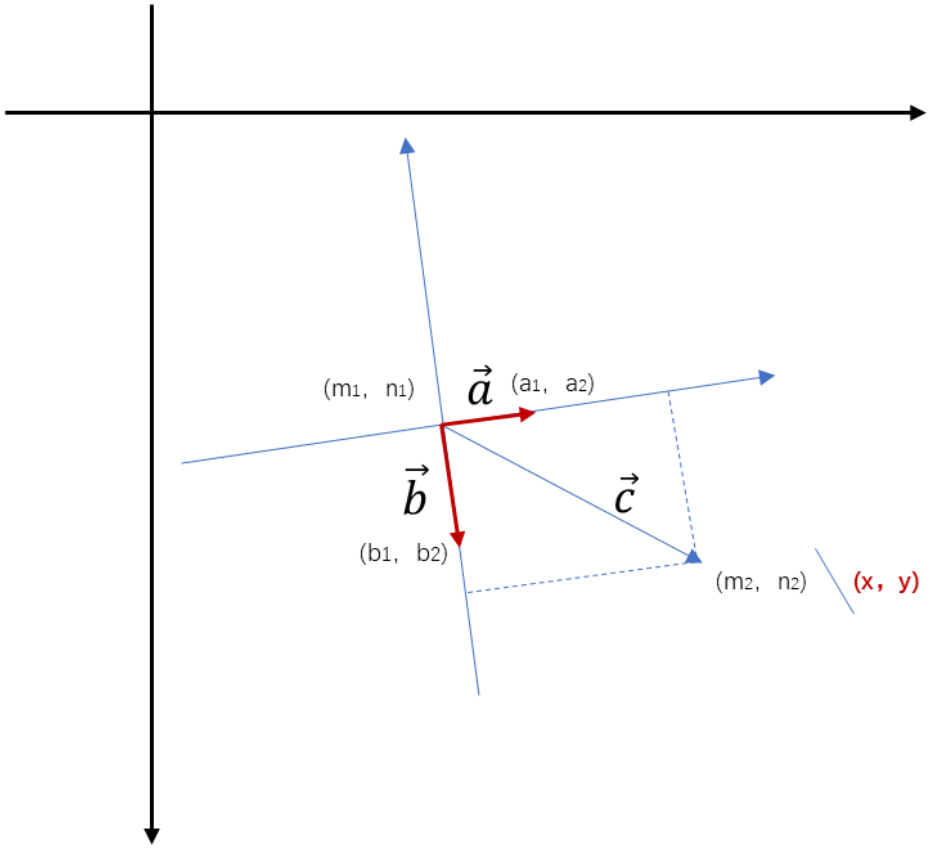


Figure 4. Coordinate system conversion, where the absolute coordinates (m_1,n_1), (m_2, n_2), (a_1, a_2), and (b_1, b_2) can be converted to the relative coordinates (0,0), (x,y), (1,0), and (0,1), respectively.



XSL•FO
RenderX

The following relationships were satisfied between vectors:



After calculation, the converted relative coordinates (x, y) were:



At this time, the value of (x, y) was the coordinate position in the relative coordinate system, which was converted by (m_2, n_2) in the absolute coordinate system.

Key Facial Feature Extraction

Based on the facial expressions and tremors in patients with PD, we evaluated and extracted features from two main dimensions. One was the amplitude magnitude of the patient's facial expression, which can be used to detect whether there is "mask face." The other was judging the patient's facial tremor through the face's key points. When the tremor effects of some patients are obvious, head and elbow vibration will drive the head to perform regular tremors. For the 106 key points on the face, they were divided into two main parts, which were amplitude features of the facial expression and shaking features of the facial small muscle group.

Extraction of the Features of Facial Expression Amplitude

We defined a key point z's position at the ith frame with a vector: $\boldsymbol{p}_{i,z}=(x_{i,z},y_{i,z})$, in which $x_{i,z}$ and $y_{i,z}$ are the relative position's horizontal coordinate and vertical coordinate, respectively. It is called the $\boldsymbol{p}$'s vector position.

For range R_{max}, we defined the key point z's range in the x-axis direction as R_x, then:

$$R_{x_max}= x_{z_max}- x_{z_min} \quad (3)$$

Similarly, the key point's range in the y-axis direction was:

$$R_{y_max}= x_{z_max}- x_{z_min} \quad (4)$$

Intuitively, covariance, Cov(X,Y), represents the expectation of the overall error of two variables, which can reflect the correlation of the patient's expression amplitude changes in two directions, and it is calculated as follows:

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) \quad (5)$$

where $E(X)$ and $E(Y)$ are the expected values of x and y, respectively.

For absolute covariance, Cov(X_abs, Y_abs), we also calculated the covariance between the absolute coordinate X_abs, Y_abs. To some extent, it reflects the shaking of the human head.

Feature Extraction of the Tremor at Facial Key Points

Because patients with Parkinson disease patients cannot control their muscles like people without PD, the key points will tremble when they are trying to maintain facial expressions. To obtain the extent of tremor, we adopted absolute coordinates to measure the patient's vibration. More specifically, we used $\boldsymbol{p}_{i,x} = (x_{iz},y_{iz})$ to denote the position of the key point z at the ith frame. The Euclidean distance between the two locations was defined as:



Moreover, when there were many position vectors, the set $\mathbf{P}$ was defined as:

$$\mathbf{P} = \{\boldsymbol{p}_{a(1,1),\, a(1,2)}, \boldsymbol{p}_{a(2,1)a(2,2)}, \ldots, \boldsymbol{p}_{a(N,1)a(N,2)}\} \quad (7)$$

in which there exists N position vectors. Then, the average position of the set $\mathbf{P}$ was defined as:



Jitter is a good measure of tremor [16]. However, it was created to calculate the tremor of a single signal. If we want to evaluate the positional Jitter of organ z, we take N frames of the organ, and the average position of the N frames is $\boldsymbol{p}$_ave($\mathbf{P}_z$). Then, positional jitter can be defined.

Jitter_abs is the relative Jitter of the key points (ie, the average absolute value of the positional distance between two adjacent frames).



Jitter_PPQ5 represents the adjacent 5 points of Jitter at the key point (ie, the average absolute value of the difference between the position of a certain frame and the average position of the adjacent 5 frames).



where $\mathbf{P}_{z,i,5}=\{\boldsymbol{p}_{i-2,z}, \boldsymbol{p}_{i-1,z}, \boldsymbol{p}_{i,z}, \boldsymbol{p}_{i+1,z}, \boldsymbol{p}_{i+2,z}\}$ denotes the average position of 5 adjacent frames (including the i-th frame) before and after the i-th frame of the key point z.

Jitter_rap represents the adjacent 3 points of jitter at the key point (ie, the average absolute value of the difference between the position of a certain frame and the average position of the adjacent 3 frames).



where $\mathbf{P}_{z,i,5}=\{\boldsymbol{p}_{i-1,z}, \boldsymbol{p}_{i,z}, \boldsymbol{p}_{i+1,z}\}$ is the average position of 3 frames (including the ith frame) before and after the ith frame of the key point z.

Jitter_ddp represents the difference between the adjacent 3 points of the key points' jitter (ie, the difference between the distances of each adjacent 2 frames in the adjacent 3 frames). Then, the average absolute value was calculated as:



When we analyzed each patient's key points, statistical analysis was performed on 106 key points. There were 8 features extracted for each key point, including 4 facial expression change amplitude features (R_{x_max}, R_{y_max}, Cov(X,Y), Cov(X_abs, Y_abs)) and 4 facial tremor features (Jitter, Jitter_PPQ5, Jitter_rap, Jitter_ddp). Since the collected data were not evaluated by a professional doctor using the UPDRS and we only knew whether the person had PD, we performed a diagnostic (classification) experiment. The final data statistics used in our experiment are shown in Table 2.

XSL•FO
RenderX

Table 2. Video data statistics.

Data statistics	Video data
Creation date	3/15/2018
Number of patients with Parkinson disease	33
Number of people without Parkinson disease	31
Number of records	176
Number of features	848
Task	Classification

Results

In this paper, we performed a categorization task for facial expressions using widely used machine learning algorithms such as logistic regression (LR) [17], support vector machine (SVM) [18], decision tree [19], and random forest (RF) [20]. Table 3 shows that the methods based on facial key points can distinguish patients with PD from people without PD. The diagnostic result is relatively good.

Table 3. Experimental results of common machine learning algorithms.

Algorithm	Precision	Recall	F1 value
LR[a]	0.98	0.98	0.98
SVM[b]	0.99	0.99	0.99
DT[c]	0.93	0.93	0.93
RF[d]	0.98	0.98	0.98

[a]LR: logistic regression.

[b]SVM: support vector machine.

[c]DT: decision tree.

[d]RF: random forest.

In order to verify the significance of original features, we conducted hypothesis testing. In this experiment, the threshold value α, which is referred to as the level of significance, was set to .05 and .005 separately.

Table 4 shows the number of key points that reached significance for each of the 8 features types. At a *P* value <.05, all 106 key points were significant for all 4 tremor features at facial key points. At a *P* value <.005, all key points were significant for only 3 tremor features at facial key points. Therefore, we found that tremor features at facial key points may be more important than facial expression amplitude features.

Table 4. Number of points that reached significance for each feature type.

Feature name	Number of key points that reached significance	
	P value <.05	*P* value <.005
R_{x_max}	83	69
R_{y_max}	56	45
Cov(X, Y)	97	87
Cov(X_abs, Y_abs)	13	12
Jitter	106	106
Jitter_PPQ5	106	106
Jitter_rap	106	90
Jitter_ddp	106	106
Total	673	621

The least absolute shrinkage and selection operator (LASSO) is a linear regression method using L1 regularization [21]. LASSO can make part of the learned feature weights be 0, so it achieves the function of sparseness and feature selection simultaneously. In this experiment, we used LASSO to compress the features.

XSL•FO
RenderX

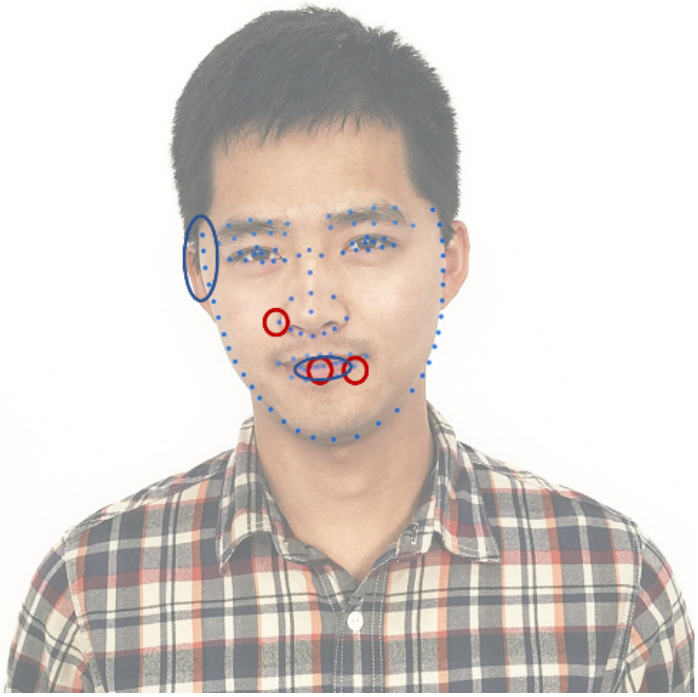
If a feature weight is nonzero in LASSO, this feature will be left; otherwise, this feature will be abandoned. The red line in Figure 5 shows that, as the coefficient α in LASSO increases, the number of features after compression decreases gradually. Then, we used the compressed features to perform the final classification task. As the number of features used in the training model decreases, the accuracy of the training set of the LR and SVM models changes (as shown by the gray and yellow lines, respectively). At an α value of .004, the two models (LR and SVM) had the highest prediction accuracy. However, as the original features were gradually reduced, the accuracy did not decrease linearly, which can prove the remaining features are more important than the discarded features.

Figure 5. The effects of least absolute shrinkage and selection operator (LASSO) feature compression on logistic regression (LR) and support vector machine (SVM) models.



Then, we used LASSO with the best hyperparameters to obtain the most relevant features to the target (PD or not). At the same time, RF was used to sort out the importance of the features. These points are near the corner of the right eye and the lips on a human. In Figure 6, the features chosen by LASSO are marked with red circles, and the features chosen by RF are marked with blue ovals.

Figure 6. The key points that have a large influence on the classification result.



XSL•FO
RenderX

In this paper, we also utilized the advanced deep learning technique long short-term memory (LSTM) [22] to perform the classification. We converted the position changes of the key points in the x axis and y axis with time into time-series data. Then, we applied an order difference to the time series, that is, $\Delta x_i = x_{i-1} - x_i$, $\Delta y_i = y_{i-1} - y_i$. LSTM utilized these new time series in the model training. The results are shown in Table 5. The experimental results are relatively good too. Compared with the pure recurrent neural network technique [23], which has no effect on the classification, LSTM demonstrated it is more practical when dealing with sequential video data. Note that the facial video data used in this paper are limited. LSTM should be able to achieve better results if more data can be acquired.

Table 5. Experimental results of neural network models.

Algorithm	Precision	Recall	F1 value
LSTM[a]	0.86	0.66	0.75
RNN[b]	0.48	0.46	0.47

[a]LSTM: long short-term memory.

[b]RNN: recurrent neural network.

Discussion

Limitations

The amount of data that we collected was not sufficient. We hope to collect more data not only from patients with PD and people without PD but also from patients with other neurological diseases.

Comparison With Prior Work

In the case of PD, to the best of our knowledge, there have only been a few software options, similar to ours, providing patients a convincing diagnosis using the facial recognition technique. These are compared to our work in Table 6.

Table 6. Comparison with a selection of prior work.

Work	Target and result	Data	Feature	Technology
Bandini et al [10]	Found PD[a] patients have lower average facial expression movement distance; facial expression recognition for PD	17 PD patients, 17 healthy control subjects	Average distance of 49 facial key points in the facial expression movement	Face tracing, SVM[b]
Rajnoha et al [11]	Identified PD hypomimia by analyzing static facial images; less accurate compared with video-recording processing method.	50 PD patients, 50 healthy control subjects	128 facial measures (embedding) by CNN[c]	Face detector-based (HOG[d]), CNN, traditional classifiers (eg, random forests, XGBoost)
PARK[e] framework by Langevin et al [12]	PARK instructs and guides users through 6 motor tasks and 1 audio task selected from MDS-UPDRS[f] and records their performance by videos	127 PD patients, 127 healthy control subjects	Facial features: facial action units (AUs); motion features: motion magnitude metric of fingers and hands based on FFT[g]	OpenFace tool version 2, FFT
Our method	Proposed facial landmark features from videos to diagnose PD using facial expressions and achieved outstanding performance	33 PD patients, 31 healthy control subjects, 176 records	848 facial expression amplitude features and tremor features of facial key points; 65 features were left after feature compression	Face ++, traditional classifiers (LR[h], SVM, DT[i], RF[j]), LSTM[k], LASSO[l]

[a]PD: Parkinson disease.

[b]SVM: support vector machine.

[c]CNN: convolutional neural network.

[d]HOG: histogram of oriented gradients.

[e]PARK: Parkinson's Analysis with Remote Kinetic-tasks.

[f]MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale.

[g]FFT: fast fourier transform.

[h]LR: logistic regression.

[i]DT: decision tree.

[j]RF: random forest.

[k]LSTM: long short-term memory.

[l]LASSO: least absolute shrinkage and selection operator.

XSL•FO
RenderX

Conclusions

In this paper, we established a diagnostic model for PD based on facial expressions. In the model, we formulated the diagnostic task into a classification problem. Then, we solved it by using algorithms from the area of traditional machine learning and the LSTM model from the field of deep learning research. When constructing video features, we conducted feature extraction according to the expression amplitude and degree of tremor. Using a fixed time interval and conversion of the coordinate system, the image was intercepted. This method converted the video into frame data. Further, the LSTM model was applied to the diagnosis of PD based on the generated time series data. Experimental results proved our model is effective and can be used as an efficient tool in PD diagnosis.

This article is a preliminary exploration of neurological diseases in the context of machine learning. The proposed method is designed to help patients get more comprehensive treatment and help doctors to understand the real-time dynamics of the disease. At the same time, it also aimed to relieve the problem of registering patients who have difficulty moving and to relieve the pressure on repeated patient diagnoses by doctors. With the development of science and technology, the introduction and application of artificial intelligence will bring more convenient and rapid diagnostic and treatment technologies.

Acknowledgments

We are deeply indebted to people from the Department of Neurology, the First Affiliated Hospital of Dalian Medical University, who supported us in data collection and provided considerable professional advice. In addition, we owe our thanks to all the volunteers, who allowed us to collect their video data. This study was supported by the Program of National Key R&D Program of China (2018YFC0116800), National Natural Science Foundation of China (No. 61772110), CERNET Innovation Project (NGII20170711), and Introducing Talents of Discipline to Universities (Plan 111) (No. B20070).

Authors' Contributions

YQ developed the conceptual framework and research protocol for the study. BJ and LZ conducted the publications review, data collection, and analysis. ZG interpreted the data. BJ drafted the manuscript, and LZ made major revisions. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Lexell J. Evidence for nervous system degeneration with advancing age. J Nutr 1997 May;127(5 Suppl):1011S-1013S [FREE Full text] [doi: 10.1093/jn/127.5.1011S] [Medline: 9164286]
2. Thomson R. Disease briefing: Parkinson's disease. Journal of International Pharmaceutical Research 2015(3):338-345. [doi: 10.13220/j.cnki.jipr.2015.03.009]
3. Martinez-Martín P, Rodriguez-Blazquez C, Paz S, Forjaz MJ, Frades-Payo B, Cubo E, ELEP Group. Parkinson Symptoms and Health Related Quality of Life as Predictors of Costs: A Longitudinal Observational Study with Linear Mixed Model Analysis. PLoS One 2015 Sep;10(12):e0145310-e0145333 [FREE Full text] [doi: 10.1371/journal.pone.0145310] [Medline: 26698860]
4. Kruszka P, Addissie YA, McGinn DE, Porras AR, Biggs E, Share M, et al. 22q11.2 deletion syndrome in diverse populations. Am J Med Genet A 2017 Apr;173(4):879-888 [FREE Full text] [doi: 10.1002/ajmg.a.38199] [Medline: 28328118]
5. Basel-Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz ID, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. Clin Genet 2016 May;89(5):557-563 [FREE Full text] [doi: 10.1111/cge.12716] [Medline: 26663098]
6. Goetz C. The history of Parkinson's disease: early clinical descriptions and neurological therapies. Cold Spring Harb Perspect Med 2011 Sep;1(1):a008862 [FREE Full text] [doi: 10.1101/cshperspect.a008862] [Medline: 22229124]
7. Tickle-Degnen L, Doyle Lyons K. Practitioners' impressions of patients with Parkinson's disease: the social ecology of the expressive mask. Social Science & Medicine 2004 Feb;58(3):603-614 [FREE Full text] [doi: 10.1016/s0277-9536(03)00213-2]
8. Tsanas A, Little M, McSharry P, Spielman J, Ramig L. Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. IEEE Trans. Biomed. Eng 2012 May;59(5):1264-1271 [FREE Full text] [doi: 10.1109/tbme.2012.2183367]
9. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat Prec 2008 Sep 12:1-1 [FREE Full text] [doi: 10.1038/npre.2008.2298.1]
10. Bandini A, Orlandi S, Escalante HJ, Giovannelli F, Cincotta M, Reyes-Garcia CA, et al. Analysis of facial expressions in Parkinson's disease through video-based automatic methods. J Neurosci Methods 2017 Apr 01;281:7-20 [FREE Full text] [doi: 10.1016/j.jneumeth.2017.02.006] [Medline: 28223023]
11. Rajnoha M, Mekyska J, Burget R, Eliasova I, Kostalova M, Rektorova I. Towards Identification of Hypomimia in Parkinson's Disease Based on Face Recognition Methods. 2018 Presented at: 10th International Congress on Ultra Modern

XSL•FO
RenderX

Telecommunications and Control Systems and Workshops (ICUMT); November 5-9, 2018; Moskva · Russian Federation p. 1-4 URL: https://doi.org/10.1109/ICUMT.2018.8631249
12. Langevin R, Ali MR, Sen T, Snyder C, Myers T, Dorsey ER, et al. The PARK Framework for Automated Analysis of Parkinson's Disease Characteristics. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2019 Presented at: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; September 11-13, 2019; London, UK p. 1-22 URL: https://doi.org/10.1145/3328925
13. Madeley P, Ellis AW, Mindham RHS. Facial Expressions and Parkinson's Disease. Behavioural Neurology 1995;8(2):115-119 [FREE Full text] [doi: 10.1155/1995/750425]
14. Newmarch J. Ffmpeg/libav. In: Linux Sound Programming. Berkeley: Apress; 2017:227-234.
15. Jung SG, An J, Kwak H, Salminen J, Jansen BJ. Inferring social media users' demographics from profile pictures: A Face++ analysis on Twitter users. New York, NY, United States: Association for Computing Machinery; 2017 Presented at: Proceedings of the 17th International Conference on Electronic Business; August, 2015; Seoul, Republic of Korea p. 140-145 URL: https://aisel.aisnet.org/iceb2017/22/
16. Boersma P, Weenink D. Praat v. 4.0. 8. A system for doing phonetics by computer. Institute of Phonetic Sciences of the University of Amsterdam 2002;218:1-2.
17. Kleinbaum DG, Dietz K, Gail M, Klein M. Logistic regression. In: Logistic regression. New York, NY: Springer; 2002.
18. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell. Syst. Their Appl 1998 Jul;13(4):18-28 [FREE Full text] [doi: 10.1109/5254.708428]
19. Quinlan J. Simplifying decision trees. International Journal of Man-Machine Studies 1987 Sep;27(3):221-234 [FREE Full text] [doi: 10.1016/s0020-7373(87)80053-6]
20. Liaw A, Wiener M. Classification and regression by random Forest. R news 2002;2:18-22 [FREE Full text]
21. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;58(1):267-288 [FREE Full text] [doi: 10.1111/j.2517-6161.1996.tb02080.x]
22. Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling. In: Thirteenth annual conference of the international speech communication association. 2012 Presented at: Thirteenth annual conference of the international speech communication association; September 9-13, 2012; Portland, Oregon URL: https://www.isca-speech.org/archive/interspeech_2012/i12_0194
23. Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent neural network based language model. 2010 Presented at: Eleventh annual conference of the international speech communication association; September 26-30, 2010; Makuhari, Chiba, Japan URL: https://www.isca-speech.org/archive/interspeech_2010/i10_1045

Abbreviations

CNN: convolutional neural network
DT: decision tree
FFT: fast fourier transform
HOG: histogram of oriented gradients
LASSO: least absolute shrinkage and selection operator
LR: logistic regression
LSTM: long short-term memory
MDS-UPDRS: Movement Disorder Society Unified Parkinson Disease Rating Scale
PARK: Parkinson's Analysis with Remote Kinetic-tasks
PD: Parkinson disease
RF: random forest
RNN: recurrent neural network
SVM: support vector machine

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by P Zhang, A Benis; comments to author 03.04.20; revised version received 15.04.20; accepted 19.04.20; published 10.07.20.

Please cite as:
Jin B, Qu Y, Zhang L, Gao Z
Diagnosing Parkinson Disease Through Facial Expression Recognition: Video Analysis
J Med Internet Res 2020;22(7):e18697
URL: https://www.jmir.org/2020/7/e18697
doi:10.2196/18697
PMID:32673247

XSL•FO
RenderX

©Bo Jin, Yue Qu, Liang Zhang, Zhan Gao. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterization and Correction of Bias Due to Nonparticipation and the Degree of Loyalty in Large-Scale Finnish Loyalty Card Data on Grocery Purchases: Cohort Study

Anna-Leena Vuorinen[1,2], PhD; Maijaliisa Erkkola[3], PhD; Mikael Fogelholm[3], PhD; Satu Kinnunen[3], BSc; Hannu Saarijärvi[4], PhD; Liisa Uusitalo[3], PhD; Turkka Näppilä[5], PhD; Jaakko Nevalainen[1], PhD

[1]Faculty of Social Sciences (Health Sciences), Tampere University, Tampere, Finland

[2]VTT Technical Research Centre of Finland Ltd, Tampere, Finland

[3]Department of Food and Nutrition, University of Helsinki, Helsinki, Finland

[4]Faculty of Management and Business, Tampere University, Tampere, Finland

[5]Tampere University Library, Tampere University, Tampere, Finland

Corresponding Author:
Anna-Leena Vuorinen, PhD
Faculty of Social Sciences (Health Sciences)
Tampere University
Arvo Ylpön katu 34
FI-33014
Tampere,
Finland
Phone: 358 408485966
Email: anna-leena.vuorinen@vtt.fi

Abstract

Background: To date, the evaluation of diet has mostly been based on questionnaires and diaries that have their limitations in terms of being time and resource intensive, and a tendency toward social desirability. Loyalty card data obtained in retailing provides timely and objective information on diet-related behaviors. In Finland, the market is highly concentrated, which provides a unique opportunity to investigate diet through grocery purchases.

Objective: The aims of this study were as follows: (1) to investigate and quantify the selection bias in large-scale (n=47,066) loyalty card (LoCard) data and correct the bias by developing weighting schemes and (2) to investigate how the degree of loyalty relates to food purchases.

Methods: Members of a loyalty card program from a large retailer in Finland were contacted via email and invited to take part in the study, which involved consenting to the release of their grocery purchase data for research purposes. Participants' sociodemographic background was obtained through a web-based questionnaire and was compared to that of the general Finnish adult population obtained via Statistics Finland. To match the distributions of sociodemographic variables, poststratification weights were constructed by using the raking method. The degree of loyalty was self-estimated on a 5-point rating scale.

Results: On comparing our study sample with the general Finnish adult population, in our sample, there were more women (65.25%, 30,696/47,045 vs 51.12%, 2,273,139/4,446,869), individuals with higher education (56.91%, 20,684/36,348 vs 32.21%, 1,432,276/4,446,869), and employed individuals (60.53%, 22,086/36,487 vs 52.35%, 2,327,730/4,446,869). Additionally, in our sample, there was underrepresentation of individuals aged under 30 years (14.44%, 6,791/47,045 vs 18.04%, 802,295/4,446,869) and over 70 years (7.94%, 3,735/47,045 vs 18.20%, 809,317/4,446,869), as well as retired individuals (23.51%, 8,578/36,487 vs 31.82%, 1,414,785/4,446,869). Food purchases differed by the degree of loyalty, with higher shares of vegetable, red meat & processed meat, and fat spread purchases in the higher loyalty groups.

Conclusions: Individuals who consented to the use of their loyalty card data for research purposes tended to diverge from the general Finnish adult population. However, the high volume of data enabled the inclusion of sociodemographically diverse subgroups and successful correction of the differences found in the distributions of sociodemographic variables. In addition, it seems that food purchases differ according to the degree of loyalty, which should be taken into account when researching loyalty

XSL•FO
RenderX

card data. Despite the limitations, loyalty card data provide a cost-effective approach to reach large groups of people, including hard-to-reach population subgroups.

(J Med Internet Res 2020;22(7):e18059) doi:10.2196/18059

KEYWORDS

loyalty card data; diet; selection bias; weighting; raking; food

Introduction

Diet has a substantial impact on human health. Poor dietary habits are associated with obesity and a wide range of chronic diseases, including type 2 diabetes, cancer, and cardiovascular diseases [1,2]. Suboptimal diet is responsible for more deaths than any other risk factor globally [3]. It is therefore imperative to collect timely and valid information on diet and individual risk factors.

To date, the evaluation of diet has mostly been based on questionnaires and diaries [4]. Although valuable in research, data collection with such instruments, particularly food diaries, is time and resource intensive, and the information is gained with a considerable delay. They also suffer from participant tendency toward social desirability [5,6]. Moreover, the information gained through questionnaires is subject to recall bias with participants not reporting all foods consumed [4]. Another limitation with dietary surveys as well as health surveys in general is selection bias, which manifests as healthy, socioeconomically advantaged, middle-aged women being the most likely to enroll in these studies [7-9].

The continued development of innovative digital tools and digital data repositories provides novel opportunities for epidemiological research [10-13]. Web-based data collection instruments [13,14] and consumer-generated data are increasingly being used for health research purposes [15-19]. While such novel data collection methods and tools may overcome some of the problems faced with traditional methods, some of the limitations remain, of which selection bias is a major concern [13,20]. Namely, those who generate the data are frequently highly selected and likely to differ from the general population representing wealthy and healthy individuals. For instance, smartphone users, and subsequently mobile health app and social media users, are younger, better educated, and represent wealthier individuals than those in the general population [21-23]. However, automated data collection, which is a typical feature for these instruments and tools, provides objective measures on individuals' health behaviors and thus decreases information bias.

Food purchase data have invoked interest as a novel approach to enrich diet and nutrition research efforts [24-26]. So far, most of the published studies have used panel-based data, with all grocery purchase receipts scanned at home [26]. While such studies are frequently large and may include data from multiple sources, they are limited by recording discrepancies [27]. In addition, receipt scanning requires consistent efforts and long-term engagement from the participants [28,29]. In this study, we used loyalty card data (ie, individual-level grocery transaction data generated by retail food chains). Importantly, loyalty card data contain information about what, where, when, and who has bought, thus enabling longitudinal tracking of the purchase behaviors of a single customer or a household over time. Objective measures of food purchases have been shown to correlate with one's food intake and overall diet quality [28]. Loyalty card data also accumulates automatically in retailers' information technology systems, producing objective and up-to-date information in a cost-effective manner. However, loyalty card data have shortfalls that could impede the usefulness for research. First, consumers may distribute their purchases among different retailers. Therefore, loyalty card data from a single retailer most likely does not include all food purchases conducted by consumers. However, in Finland, the market is highly centralized with the three biggest market chains claiming over 90% of the market share, and the largest operator having a market share as high as 47% [30]. Such centralization provides a unique opportunity to investigate heterogeneous populations through a single retailer.

The aims of this study were as follows: (1) to investigate and quantify selection bias in Finnish large-scale loyalty card (LoCard) data and further develop a means to correct this bias by characterizing the loyalty card data consenters and comparing their sociodemographic background to that of the general Finnish adult population and (2) to assess how the degree of loyalty relates to food purchases by investigating the self-perceived degree of loyalty (share of total grocery purchases in retailers' shops and supermarkets) and its association with food purchases. The overall purpose of this research was to increase the understanding of how loyalty card data should be understood and subsequently analyzed in dietary and health research.

Methods

Study Design and Participation

The LoCard data used in this study were obtained from S Group, which is the largest commercial operator of retail grocery stores in Finland. According to S Group, their full coverage is 2.4 million households, meaning that 88% of households in Finland have registered purchases in their databases. The members of S Group's loyalty card program are provided with an electronic customer card to be used when making purchases, and customers are rewarded for their purchases by getting a maximum 5% financial bonus that is refunded to them on a monthly basis. Individuals of the same household may link their purchases to the same loyalty account. In this study, only purchases of the household's main cardholder were used.

Members of S Group's loyalty card program (primary cardholders) across Finland were contacted via email and were invited to take part in the study, which involved consenting to

XSL•FO
RenderX

the release of their grocery purchase data to be used for research purposes and voluntarily responding to the study questionnaire. Members who did not have an email address declared or who had prohibited the retailer from contact them with any marketing or research-related material were excluded. Cardholders under 18 years of age were also excluded. All invitations were sent by S Group as they had customers' contact information.

The grocery purchase data used in this study covered the period from January 1, 2017, to December 31, 2018. Each purchase was associated with item description, time stamp, quantity (ie, weight, volume, or number of packages), and expenditure on the item.

Background Variables

All consenting participants were asked to fill out a web-based background questionnaire that included the following sociodemographic variables: education, marital status, size of the household, number and age of children, occupational status, income, and perceived health. The background data were complemented with information on participant sex, age, and postal code obtained from the retailer's electronic database.

Degree of Loyalty

As part of the baseline questionnaire, all participants were asked to estimate their degree of loyalty as a share of purchases made in the retailer's shops and supermarkets on a five-item ordinal scale. The response categories were as follows: "0%-20%," "21%-40%," "41%-60%," "61%-80%," and "81%-100%."

Food Variables and Food Groups

The LoCard grocery purchase data required preprocessing to be usable in further analyses. First, we identified food groups from all the grocery product groups. Second, we regrouped the identified food groups into new groups that were formed on the basis of the commonly used food groupings in nutritional studies [31] and earlier findings on the associations between dietary components and health [32,33]. For instance, skimmed liquid milk and buttermilk were aggregated into "skimmed milk & sour milk" and foods and mixed dishes with red or processed meat as the main ingredient were aggregated into "red meat & processed meat."

Out of 4234 grocery product groups, 865 (20.4%) were assigned into one of the new food groups used in this study. In addition, 42 food groups were left out as they involved either (1) a mixed dish or food group with no definite primary ingredient or (2) a rarely purchased product. The food groups used in this study included "vegetables," "skimmed milk & sour milk," "sugar-sweetened beverages," "rye bread," "red meat & processed meat," "fat spreads," and "sweets & chocolate." These groups were used as indicators for evaluating the nutritional quality of household food purchases. A detailed description about the grouping of the food purchase data is included in Multimedia Appendix 1.

Reference Material

Population statistics on the general adult population were obtained from Statistics Finland using StatFin databases that can be freely accessed [34]. The databases include tabulated data on Finnish citizens and Finland in general that are collected on a yearly basis. Data from 2017 were used because of the availability of the latest data tables for all sociodemographic variables used in the analyses. For this study, individuals aged at least 18 years were included.

The FinHealth survey is a national population health study on Finnish citizens. The study encompasses a series of cross-sectional population surveys carried out every 5 years in Finland. The latest FinHealth survey was carried out at 50 localities in 2017, with a participation rate of 71% among those invited for the study [35]. The purpose of the FinHealth study is to collect up-to-date information about the health and well-being of adults residing in Finland and on the factors influencing their health and well-being. Each survey invites 10,000 randomly selected individuals aged over 18 years. The study consists of physical examinations and study questionnaires. The latest report (values used in this study) is restricted to adults aged 30 years or older to make the results comparable with earlier FinHealth studies. A subgroup of the participants was also invited to undergo a nutrition review; the FinDiet survey is a substudy (n=1655) of the FinHealth survey, which monitors the nutrition and dietary habits of the Finnish population [36].

Statistical Methods

Analysis of and Correction for Selection Bias

The sociodemographic characteristics of the LoCard study participants were first compared with the characteristics of the Finnish adult population and participants of the FinHealth study to identify traits in LoCard participants that deviated from traits in the general Finnish adult population.

Second, we constructed poststratification weights for the LoCard participants to match their sociodemographic distributions with the adult Finnish population distributions as closely as possible. The individual weights were calculated using the raking function available in the *survey* package in R [37]. The raking function uses iterative proportional fitting (IPF), which is a technique that can be used to adjust a distribution reported in one dataset by totals reported in another. For a given two-way contingency table, the IPF proportionally adjusts each row of the sample distribution in the two-way contingency table to have its total equal the reference population row distribution and adjusts each column of the sample distribution to have its total equal the column total in the reference table [38].

The advantage of the raking function is that the algorithm allows multiple two-dimensional (or higher dimensional) tables to be matched simultaneously [37]. For example, instead of matching age, sex, and education univariate distributions separately, we can match all bivariate distributions (ie, age and education, sex and education, and sex and age) simultaneously. The adjustment process is repeated iteratively until the weights converge for each table used in the analysis. The raking function requires that the two contingency tables have the same classes for the row and column variables and no zero values in any of the cells.

The following two-way tables were available for both the LoCard data and the Finnish adult population: sex and age, sex and education, sex and marital status, sex and occupational status, age and education, age and marital status, age and

XSL•FO
RenderX

occupational status, and education and occupational status. All tables were subsequently used to construct the poststratification weights. In addition, the distribution of children aged under 18 years living in the household was used alone because corresponding two-dimensional tables with any of the background variables were not available in Statistics Finland. In total, eight two-way tables and a single one-way table were used in the construction of the weights. Finally, the obtained weights were trimmed to avoid extreme values and instability by setting a minimum value of 0.1 and a maximum value of 10. Without trimming, the poststratification weights ranged from 0.04 to 32.7, and there was a single extremely high weight of 82.4.

Owing to missing data, the poststratification weights were constructed in two phases. First, the weights were calculated as described above for participants for whom all baseline characteristics used in the matching were available. These data were available for 36,094 individuals. Participants with missing data for any of these variables (n=10,972) obtained their weights in the second phase, where the poststratification weights were calculated for the whole LoCard sample using sex and age variables only. This information was available for 47,045 participants. Finally, the combined weights were rescaled to add up to 47,045. Twenty-one participants without data on sex and/or age remained without weights.

The selected food group variables were analyzed to describe the volume and money (€) spent on their purchases over the 2-year period (2017-2018). For descriptive purposes, median values and IQRs were reported for each variable because the distributions were strongly skewed to the right, and there was an excess number of zero values in some of the food variables. The same variables were used to demonstrate how the poststratification weights affected the results.

Degree of Loyalty

To validate the self-assessed degree of loyalty, we conducted the recency, frequency, and monetary (RFM) value analysis using the transaction data of all participants and compared the RFM scores across the five degree of loyalty groups. RFM analysis is a behavior-based technique used to segment customers by examining their transaction history from three dimensions (how recently a customer made purchases, how often they purchased, and how much they purchased). RFM analysis is also widely used in customer relationship management. Based on these three dimensions, the RFM score is generated for each individual, with a higher score indicating higher loyalty. The analysis was conducted using the *rfm* package in R [39]. In addition, total volume and total money (€) spent on food purchases were calculated for each degree of loyalty group to investigate how closely the self-reported degree of loyalty relates to volume and money spent on the purchases.

To assess the impact of the degree of loyalty on food purchasing profiles, the selected food group variables were compared among the five degree of loyalty groups. The Kruskal-Wallis test was applied for differences across the groups.

The association between the degree of loyalty and background characteristics was analyzed by comparing the distribution of each sociodemographic variable among the five degree of loyalty groups. The differences across the groups were tested using the chi-square test.

Ethical Aspects

The study was approved by the University of Helsinki Review Board in the Humanities and Social and Behavioral Sciences (Statement 21/2018). Informed consent was electronically obtained from all participants included in the study when they were invited via email to release their loyalty card data and fill out the background questionnaire. The data were pseudonymized by S Group before the researchers could obtain the data.

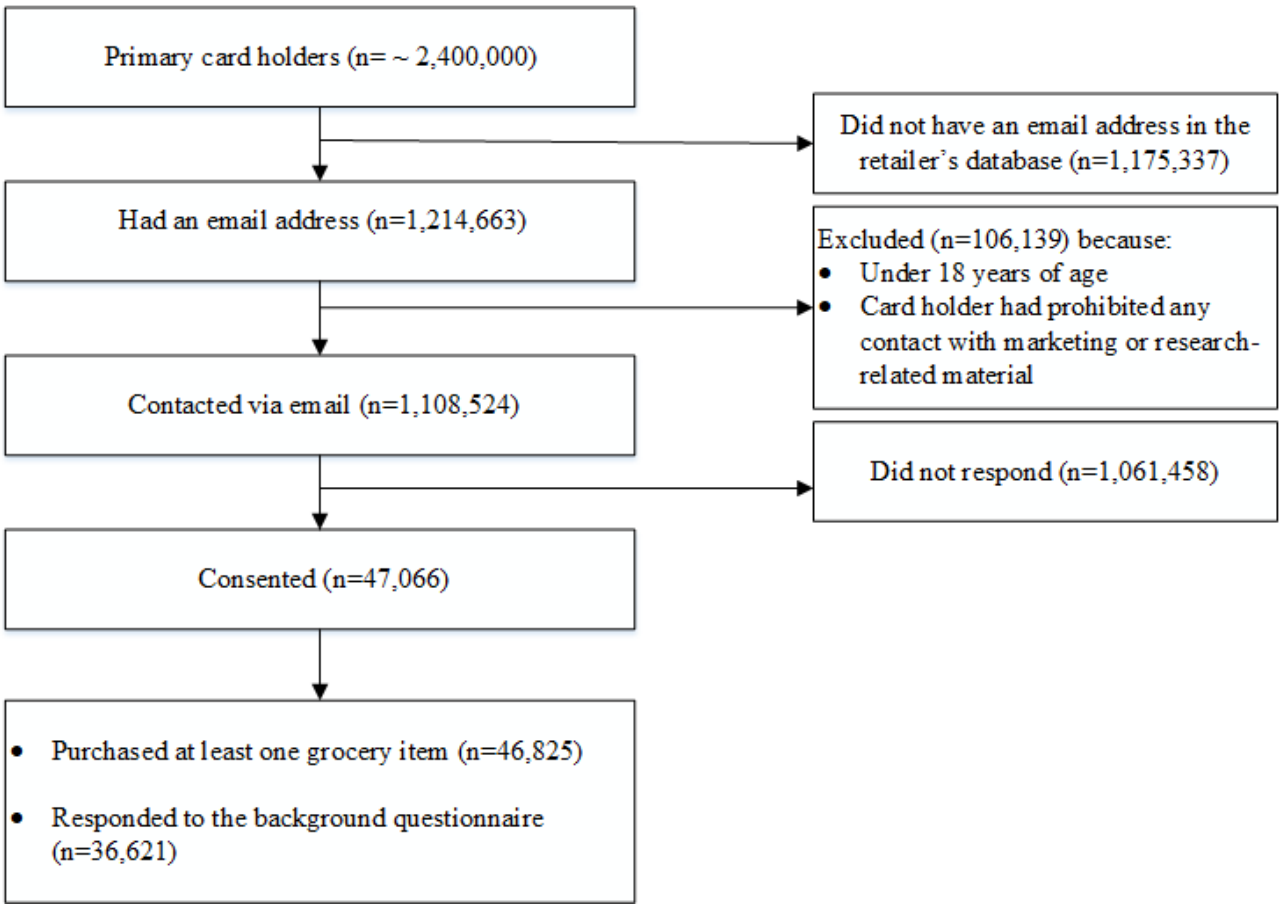
Results

Recruitment

S Group had approximately 2.4 million primary loyalty card owners, and all of them were assessed for eligibility (Figure 1). Approximately half (1,214,663, 51%) of the loyalty card owners were contacted, and of these, 47,066 (4%) consented to participate. We did not have information on the number of valid email addresses or what proportion of emails reached the card owners (eg, by passing through trash email filters). Among the participants, 36,621 (78%) responded to the background questionnaire. Nearly all participants (46,825, 99.5%) purchased at least one grocery item from 2017 to 2018.

XSL•FO
RenderX

Figure 1. Participant recruitment and eligibility flow chart.



Participant Characteristics

Table 1 shows the participant characteristics compared with those of the Finnish adult population and the FinHealth study participants. Discrepancies were found in sex, age, education, and occupational status when compared with the general Finnish adult population. Namely, there were more women, more individuals with a higher education, and more employed individuals in the LoCard sample. On the contrary, individuals aged under 30 years and over 70 years (correspondingly, retired individuals) were underrepresented in the LoCard sample. Selectivity associated with education was strong in the LoCard sample. The proportion of individuals having a basic education level was clearly lower in the LoCard sample (6% of participants had basic education) than in the Finnish adult population (25% had basic education). There were no major differences in the distribution of marital status. However, there were fewer individuals living in a household with children aged under 18 years in the LoCard sample. The LoCard sample was widely distributed across Finland and comparable to the geographical distribution of Finnish citizens (Multimedia Appendix 2 and Multimedia Appendix 3).

On comparing the LoCard sample to the FinHealth study participants, there were differences in sex, education, and marital status distributions, with more women and individuals with higher education and fewer married individuals in the LoCard sample. The age distributions were not comparable owing to the fact that the FinHealth study included only individuals aged at least 30 years. Distortion in the distribution of occupational status was similar in the two studies compared with the Finnish adult population.

The reweighted distributions of the sociodemographic variables demonstrated that the constructed poststratification weights corrected the deviations successfully, and thereafter, the sociodemographic distributions of the LoCard sample matched well with the Finnish adult population.

XSL•FO
RenderX

Table 1. LoCard participant characteristics compared with those of the general Finnish adult population and participants of the FinHealth study.

Characteristic	Finnish general adult population (N=4,446,869)	FinHealth study[a] (N=6545)	LoCard sample (N=47,066)[b]	Weighted LoCard sample[c] (N=47,045)
Sex (women), n (%)	2,273,139 (51.12%)	3496 (53.42%)	30,696 (65.25%)	23,837 (50.67%)
Age (years), mean (SD)	50.23 (19.06)	—[d]	47.10 (15.21)	49.5 (0.14)
Age distribution (years), n (%)				
≤29	802,295 (18.04%)	N/A[e]	6791 (14.44%)	8532 (18.14%)
30-39	702,767 (15.80%)	Men, 483 (15.8%); women, 539 (15.4%)	9982 (21.22%)	7505 (15.95%)
40-49	660,703 (14.86%)	Men, 530 (17.4%); women, 561 (16.1%)	9503 (20.20%)	6986 (14.85%)
50-59	734,554 (16.52%)	Men, 608 (20.0%); women, 661 (18.9%)	9154 (19.45%)	7715 (16.40%)
60-69	737,233 (16.58%)	Men, 727 (23.8%); women, 774 (22.4%)	7880 (16.75%)	7734 (16.44%)
≥70	809,317 (18.20%)	Men, 701 (23.0%); women, 961 (27.5%)	3735 (7.94%)	8572 (18.22%)
Marital status, n (%)				
Presently married	1,990,928 (44.77%)	Men, 58.0%; women, 52.3%	17,240 (47.32%)	16,254 (45.12%)
Cohabiting	—	Men, 16.7%; women, 14.4%	7408 (20.33%)	—
Single	1,599,827 (35.98%)[f]	Men, 13.3%; women, 10.3%	6412 (17.60%)	12,762 (35.43%)[f]
Divorced or separated	574,620 (12.92%)	Men, 8.7%; women, 12.2%	4331 (11.89%)	4713 (13.08%)
Widowed	281,494 (6.33%)	Men, 3.4%; women, 10.9%	1040 (2.86%)	2295 (6.37%)
Household, mean number of members (SD)	2.8 (not available)	—	2.36 (1.25)	2.42 (0.01)
Children aged under 18 years living in the household, n (%)	566,242 (38.48%)	31[g] (31.52%)	11,705 (32.08%)	13,567 (37.61%)
Education, n (%)				
Primary school or less	1,112,261 (25.01%)	Men, 23.2%; women, 21.0%	2259 (6.21%)	7881 (23.54%)
Upper secondary school	1,902,332 (42.78%)	Men, 38.3%; women, 29.1%	13,405 (36.88%)	15,534 (43.25%)
Bachelor's degree or equivalent	955,395 (21.49%)	Men, 38.5%[h]; women, 49.9%[h]	11,787 (32.43%)	8453 (21.94%)
Master's degree or higher	476,881 (10.72%)	—	8897 (24.48%)	4049 (11.27%)
Occupational status, n (%)				
Employed	2,327,730 (52.35%)	Men, 65.9%; women, 62.3%	22,086 (60.53%)	19,027 (52.75%)
Unemployed	296,191 (6.66%)	Men, 8.5%; women, 7.0%	1637 (4.49%)	2417 (6.70%)
Student	230,489 (5.18%)	Men, 2.4%; women, 3.5%	1824 (5.00%)	1619 (4.49%)

XSL•FO
RenderX

Characteristic	Finnish general adult population (N=4,446,869)	FinHealth study[a] (N=6545)	LoCard sample (N=47,066)[b]	Weighted LoCard sample[c] (N=47,045)
Retired	1,414,785 (31.82%)	Men, 21.3%; women, 20.0%	8578 (23.51%)	11,600 (32.16%)
Parental leave		Men, 0.2%; women, 4.2%	1255 (3.44%)	—
Other	177,674 (4.00%)	Men, 1.8%; women, 3.0%	1107 (3.03%)	1411 (3.91%)[i]
Degree of loyalty, n (%)				
0%-20%	—	—	2283 (6.25%)	2132 (5.90%)
21%-40%	—	—	4670 (12.79%)	4160 (11.52%)
41%-60%	—	—	6155 (16.85%)	5828 (16.14%)
61%-80%	—	—	9224 (25.25%)	8962 (24.82%)
81%-100%	—	—	14,194 (38.86%)	15,031 (41.62%)

[a]FinHealth study included individuals aged ≥30 years, which makes the age distribution not comparable to other data listed in the table.

[b]Data for the following numbers of participants were missing in the LoCard sample: sex, 21; age, 21; marital status, 10,635; household, 10,689; children aged under 18 years, 10,576; education, 10,718; occupational status, 10,579; degree of loyalty, 10,540.

[c]Weighted LoCard sample refers to the descriptive statistics calculated using the poststratification weights of the LoCard participants.

[d]Not available.

[e]N/A: not applicable.

[f]Cohabitating included in this category.

[g]Households with three or more persons.

[h]Bachelor's degree or higher.

[i]Parental leave included.

Food Purchase

Table 2 shows the purchases of selected food groups in the original LoCard sample and in the weighted LoCard sample. Over 95% of the participants had purchased at least one food product in all food groups, except skimmed milk & sour milk. Skimmed milk & sour milk had been purchased by 74% of the participants. Among them, the median expenditure and the median weight were €23.0 (€1=US $1.13 in 2017) and 23.5 kg, respectively, during the 2-year follow-up.

After applying the poststratification weights, there was an increase in the purchase of red meat & processed meat and small increases in sugar-sweetened beverages and fat spreads. The purchase of vegetables and sweets & chocolate decreased as a result of reweighting. The largest change was seen in red meat & processed meat; the weighted amount of purchase increased from €387 to €417 (cost) and from 48 kg to 54 kg (weight), corresponding to relative percentage increases of 7.8% and 12.6%, respectively.

XSL•FO
RenderX

Table 2. Purchase of selected food groups (measured in € and kg) in the original LoCard sample and in the weighted LoCard sample.

Food group	Original LoCard sample[a] (N=47,066)				Weighted LoCard sample[a] (N=47,045)			
	€, median [IQR]	€ (%)[b], median [IQR]	kg, median [IQR]	kg (%)[c], median [IQR]	€, median [IQR]	€ (%)[b], median [IQR]	kg, median [IQR]	kg (%)[c], median [IQR]
Vegetables	284.3 [124.9-520.1]	7.7 [5.4-10.5]	76.6 [33.2-144.4]	8.2 [5.4-11.7]	263.7 [107.3-487.6]	7.2 [4.8- 9.9]	73.4 [29.6-139.6]	7.6 [4.8- 11.0]
Skimmed milk & sour milk	6.9 [0-60.9]	0.2 [0-1.7]	7.0 [0-65.6]	0.9 [0-7.3]	6.6 [0-60.5]	0.2 [0-1.8]	6.5 [0-66.0]	0.9 [0-7.2]
Sugar-sweetened beverages	45.3 [15.1-111.9]	1.3 [0.6-2.7]	23.5 [7.5-63.4]	2.8 [1.1-6.0]	47.4 [15.0-120.2]	1.4 [0.6-3.0]	25.6 [7.7-69.8]	3.0 [1.1-6.7]
Rye bread	50.7 [18.1-112.7]	1.5 [0.8-2.5]	12.9 [4.6-28.9]	1.5 [0.8-2.4]	50.5 [17.3-114.9]	1.5 [0.7-2.5]	13.2 [4.6-29.8]	1.5 [0.7-2.5]
Red meat & processed meat	386.5 [153.6-778.1]	11.3 [7.5-15.2]	47.5 [18.3-98.2]	5.3 [3.4-7.5]	416.8 [170.1-816.7]	12.1 [8.2-16.0]	53.5 [21.4-106.1]	5.7 [3.8-8.0]
Fat spreads	53.1 [20.1-114.8]	1.5 [0.9-2.3]	10.1 [3.8-21.6]	1.1 [0.7-1.7]	56.3 [20.3-122.0]	1.6 [0.9-2.5]	10.9 [4.0-23.2]	1.2 [0.7-1.8]
Sweets & chocolate	119.2 [48.9 - 243.4]	3.5 [1.9-5.8]	10.3 [4.1-21.6]	1.2 [0.6-2.0]	109.9 [42.1-232.7]	3.2 [1.7-5.5]	9.5 [3.6-20.8]	1.1 [0.5-1.9]

[a]Purchases are aggregated over a 2-year period from January 1, 2017, to December 31, 2018 (€1=US $1.13 in 2017).

[b]Share of the food group purchase among all grocery purchases measured in euros.

[c]Share of the food group purchase among all grocery purchases measured in kilograms.

Degree of Loyalty

Table 1 shows the self-assessed degree of loyalty. Almost 40% (14,194/36,526) of the participants reported that they made 80% or more of their food purchases at S Group shops and supermarkets, and 64% (23,418/36,526) reported making at least 60% of their purchases at the retailer's shops and supermarkets.

The RFM scores were significantly different among the five degree of loyalty groups, with the lowest scores in the lowest degree of loyalty group and a steady increasing trend toward the highest degree of loyalty group (F_4=4625.5, P<.001). The poststratification weights also differed significantly across the five groups (F_4=24.1, P<.001), indicating that the degree of loyalty was associated with individuals' sociodemographic characteristics. However, the observed differences were rather small, with a maximum difference of six percentage points between the groups (Table 3). In the highest degree of loyalty group, there were slightly more young and married participants, and the percentage of households with children was higher, whereas the percentage of divorced or separated participants and those with a master's degree declined with the degree of loyalty.

XSL•FO
RenderX

Table 3. LoCard participant characteristics and RFM scores across the five degree of loyalty groups.

Characteristic	Degree of loyalty				
	0%-20% (n=2283)	21%-40% (n=4670)	41%-60% (n=6155)	61%-80% (n=9224)	81%-100% (n=14,194)
RFM[a] analysis score, median [IQR]	182.5 [111.0-321.0]	311.0 [122.0-442.0]	335.0 [221.0-522.0]	432.0 [244.0-534.0]	445.0 [324.0-545.0]
Sex (women), n (%)	1472 (64.6%)	3176 (68.1%)	4150 (67.5)%	6101 (66.2)%	9317 (65.7%)
Age, n (%)					
≤29	227 (10.0%)	571 (12.2%)	866 (14.1%)	1284 (13.9%)	2089 (14.7%)
30-39	406 (17.8%)	935 (20.0%)	1232 (20.0%)	1993 (21.6%)	3107 (21.9%)
40-49	512 (22.5%)	1003 (21.5%)	1327 (21.6%)	1838 (19.9%)	2779 (19.6%)
50-59	536 (23.5%)	1035 (22.2%)	1273 (20.7%)	1763 (19.1%)	2664 (18.8%)
60-69	411 (18%)	797 (17.1%)	1003 (16.3%)	1585 (17.2%)	2409 (17.0%)
≥70	188 (8.2%)	325 (7.0%)	450 (7.3%)	759 (8.2%)	1143 (8.1%)
Marital status, n (%)					
Presently married	1021 (45.0%)	2039 (43.9%)	2733 (44.5%)	4379 (47.6%)	7056 (49.8%)
Cohabitating	437 (19.3%)	982 (21.1%)	1321 (21.5%)	1862 (20.3%)	2803 (19.8%)
Single	416 (18.3%)	910 (19.6%)	1187 (19.3%)	1593 (17.3%)	2303 (16.3%)
Divorced or separated	323 (14.2%)	600 (12.9%)	729 (11.9%)	1090 (11.9%)	1588 (11.2%)
Widowed	73 (3.2%)	116 (2.5%)	166 (2.7%)	271 (3.0%)	412 (2.9%)
Household, n (%)					
Children aged under 18 years living in the household	649 (28.5%)	1398 (30.0%)	1846 (30.1%)	2995 (32.5%)	4814 (34.0%)
Education, n (%)					
Primary school or less	144 (6.3%)	231 (5.0%)	335 (5.5%)	525 (5.7%)	1023 (7.3%)
Upper secondary school	756 (33.3%)	1645 (35.4%)	2201 (36.0%)	3343 (36.4%)	5451 (38.6%)
Bachelor's degree or equivalent	735 (32.4%)	1538 (33.1%)	2046 (33.4%)	3020 (32.9%)	4446 (31.5%)
Master's degree or higher	635 (28.0%)	1228 (26.5%)	1538 (25.1%)	2302 (25.1%)	3189 (22.6%)
Occupational status, n (%)					
Employed	1348 (59.2%)	2892 (62.2%)	3761 (61.2%)	5544 (60.2%)	8533 (60.2%)
Unemployed	124 (5.5%)	219 (4.7%)	284 (4.6%)	408 (4.4%)	602 (4.3%)
Student	125 (5.5%)	236 (5.1%)	329 (5.4%)	461 (5.0%)	673 (4.7%)
Retired	563 (24.7%)	1018 (21.9%)	1381 (22.5%)	2207 (24.0%)	3391 (23.9%)
Parental leave	51 (2.2%)	137 (2.9%)	174 (2.8%)	318 (3.5%)	575 (4.1%)
Other	65 (2.9%)	151 (3.2%)	214 (3.5%)	273 (3.0%)	402 (2.8%)

[a]RFM: recency, frequency, and monetary.

Table 4 shows food purchases in the degree of loyalty groups, and all showed significant associations (P<.001 for all food groups, except sweets & chocolate [P=.007]). The result was expected owing to the large sample size. The shares of vegetable, red meat & processed meat, and fat spread purchases increased as the degree of loyalty increased. In the other food groups, there were no major differences across the degree of loyalty groups.

Additionally, Table 4 shows that the quantity and expenditure regarding food groups increased steadily with the self-assessed degree of loyalty, suggesting that the self-assessment can be relied upon.

XSL•FO
RenderX

Table 4. Purchases (in € and kg) of selected food groups across the five degree of loyalty groups.

Food group	Degree of loyalty																			
	0%-20% (n=2216)				21%-40% (n=4611)				41%-60% (n=6119)				61%-80% (n=9168)				81%-100% (n=14,133)			
	€[a,b]	€%[b,c]	kg[b]	kg%[b,d]	€[b]	€%[b,c]	kg[b]	kg%[b,d]	€[b]	€%[b,c]	kg[b]	kg%[b,d]	€[b]	€%[b,c]	kg[b]	kg%[b,d]	€[b]	€%[b,c]	kg[b]	kg%[b,d]
Vegetables	58.0	6.6	15.2	7.1	131.1	7.4	35.0	7.9	232.0	7.9	60.6	8.4	344.2	8.1	93.7	8.7	441.9	7.9	122.5	8.5
Skimmed milk & sour milk	1.8	0.2	2.0	0.8	3.6	0.2	3.0	0.9	5.6	0.2	5.0	0.8	8.8	0.2	8.0	0.9	14.0	0.3	13.5	1.1
Sugar-sweetened beverages	10.9	1.3	5.5	2.7	23.7	1.4	12.2	2.9	37.2	1.3	19.1	2.8	52.2	1.3	27.4	2.7	67.9	1.3	35.7	2.7
Rye bread	10.9	1.3	2.7	1.3	24.1	1.4	6.2	1.4	41.9	1.5	10.4	1.5	61.9	1.5	15.5	1.5	82.9	1.5	21.4	1.5
Red meat & processed meat	80.4	9.9	9.9	5.0	186.4	10.7	22.8	5.2	297.8	10.9	36.8	5.2	457.4	11.3	55.2	5.3	621.5	11.6	76.9	5.4
Fat spreads	10.2	1.2	1.8	0.9	23.5	1.3	4.4	1.0	41.0	1.4	7.8	1.1	65.4	1.6	12.3	1.2	87.2	1.6	16.7	1.2
Sweets & chocolate	32.3	3.7	2.8	1.3	64.1	3.7	5.4	1.3	95.4	3.4	8.3	1.2	134.6	3.3	11.6	1.1	183.9	3.5	15.6	1.2
Total amount of grocery purchases, median [IQR]	873.4 [442.0-1602.5]		215.9 [105.6-393.2]		1883.8 [1095.7-2924.2]		460.7 [258.5-756.1]		2958.6 [1840.0-4528.8]		734.0 [452.8-1169.5]		4320.2 [2726.5-6479.0]		1095.3 [671.5-1698.1]		5680.1 [3616.4-8531.8]		1462.7 [913.0-2250.6]	

[a]€1=US $1.13 in 2017.

[b]Median value.

[c]Share of the food group purchase among all grocery purchases measured in euros.

[d]Share of the food group purchase among all grocery purchases measured in kilograms.

Discussion

Principal Findings

The findings of this study showed that individuals who consented to the release of their loyalty card data for research purposes tended to diverge from the general Finnish adult population. Similar to many other health and nutrition studies, including those encompassing electronic data collection tools [7,13,35,40,41], the LoCard participants manifested volunteer bias, with employed individuals, middle-aged individuals, women, and individuals with higher education being overrepresented in the sample. The LoCard sample included fewer retired individuals, fewer individuals with basic education, and fewer individuals who had children aged under 18 years living in the household. Compared with the Finnish national FinHealth and FinDiet studies, the selection mechanism appeared to be somewhat different in the LoCard sample. While employed individuals were overrepresented in all these three studies, the gender and education biases were stronger in the LoCard sample. Moreover, the LoCard sample had a rather similar distribution of marital status as among Finnish adults, whereas in the FinHealth study, married individuals were overrepresented [35,36,42].

However, the size (n=47,066) and heterogeneity of the LoCard sample enabled a successful correction of the differences seen in the sociodemographic variables. We developed the poststratification weights using all sociodemographic background variables available with the two-way joint distributions to correct the background distributions of the LoCard participants to make them closer to the Finnish adult population. The large sample size provided a sufficient number of participants for hard-to-reach population subgroups, and thus, it was possible to construct the poststratification weights for them as well. The highest weights were seen for unmarried men aged under 30 years, who indeed are often underrepresented or not enrolled in health studies [41].

Of the 1.1 million loyalty card holders contacted, approximately 4% (n=47,066) took part in the LoCard study. Although low, the participation rate was similar to that for other massive data collection methods [7]. The advantage of the use of digital tools is that they reach a large number of potential study participants with relatively low effort in data collection. After all, we reached substantially more individuals than in the majority of dietary studies using traditional data collection methods with minimum human involvement in data collection. A likely reason for the low participation rate was that the participants were contacted via email, which may not have reached them (invalid email address or contact email classified as “junk email”) or may have limited their participation and induced selection bias. Although 88% of households in Finland have an internet connection and 83% use email [43], email use varies according to

XSL•FO
RenderX

sociodemographic profiles and is relatively low at 62% among individuals aged over 65 years and among individuals with basic education [44]. This may partly explain the baseline characteristics of the LoCard sample. However, it has been shown that the use of digital tools in recruitment and data collection does not increase the selection bias, but the traits of participants in health studies are rather similar regardless of the recruitment method used [13,40]. Moreover, it is likely that many simply ignored an email coming from a commercial party.

Important aspects are whether and when informed consent from loyalty card owners is needed. Recently, Aiello et al [45] published an interesting ecological study on the associations between loyalty card food purchase data and prescription records that were used as a proxy for real disease profiles in London. Their dataset included 1.6 million loyalty card users, and they used the anonymized data without the consent of the individuals. In our study, consent and a positive reply were required for two reasons. First, transparent use of loyalty card data on customers for a common good builds trust among them, researchers, and the company, and reduces the likelihood of negative publicity. Second, contact was needed to obtain information about participants' background characteristics for use in further analyses. A future scenario could involve a consent request when the customer becomes a member of the loyalty card program. This would create an ethically sound and transparent research protocol for the use of customer data.

Poststratification weights were further applied in evaluating the purchases of the main food groups. The corrections demonstrated small changes in some food groups; the purchase of vegetables and sweets & chocolate decreased after the correction, whereas the purchase of red meat & processed meat, sugar-sweetened beverages, and fat spreads increased. The sociodemographic profiles of the LoCard participants and bias related to them might, at least partly, explain these results. The FinDiet study showed that women, who were overrepresented in the LoCard sample and thus had smaller weights, tended to consume more fruits and vegetables than men [36]. In line with this, after applying the poststratification weights, the purchase of vegetables decreased. It has also been shown that socioeconomically advantaged individuals, who likewise were overrepresented in the LoCard sample, consumed healthy foods, such as fruits and vegetables and low-fat dairy products, more frequently [46]. Moreover, the increased amount of red meat & processed meat purchase is likely related to male participants who tend to consume more meat [36,47]. In line with the overall findings of this study, the NutriNet-Santé study showed that the consumption of fruits and vegetables was higher and the consumption of meat was lower in the cohort than in the general population in France [13].

The degree of loyalty was fairly high in the LoCard sample with 64% (23,418/36,526) of the participants reporting making over 60% of their grocery purchases at the retailer's shops and supermarkets. The food purchases were rather similar in the higher loyalty group (60% or higher), whereas individuals making less purchases in the retailer's grocery stores showed some differences. In particular, individuals reporting the lowest degree of loyalty tended to buy fewer vegetables and fat spreads and fewer red meat & processed meat products. Although some variation was seen, the differences across the loyalty groups were smaller than expected. One reason could be the food groups selected for the current analyses. There could be other products, such as alcohol and tobacco, that are differently purchased. These results, together with the differences seen in the sociodemographic variables between the loyalty groups, underline the importance of estimating and accounting for the degree of loyalty in future studies using loyalty card data. A direct way to address the problem of coincidental purchases is to focus on a subsample with at least 60% loyalty. It is also important to note that loyalty card data can enable research on longitudinal trends in food purchases, which can be performed regardless of the degree of loyalty.

Limitations

Although we used a large set of matching variables for developing the poststratification weights, some limitations concerning these remain. First, we were not able to compare or account for possible differences in income, as there was no comparable reference data available in Statistics Finland. Therefore, it remains unclear whether the LoCard sample was representative in terms of income, which is an important contributor to food purchase. The higher education level of the LoCard participants and the lower prevalence of young and retired individuals clearly suggest that the income levels might be overestimated in our sample. Second, although we matched families with children, the number of children and their ages, which can clearly affect a household's food purchases, were not used in weighting. Importantly, we were able to correct the differences only in the observed sociodemographic variables, and thus, unidentifiable selection bias cannot be ruled out. This may include factors that would be associated with willingness to participate, such as special dietary restrictions and socially excluded people. In particular, among those participants who did not have complete background information and whose poststratification weights were thus based on sex and age only, the risk for unidentifiable selection bias could be even higher.

It is important to note that grocery purchases reflect consumption on a household level, which may consist of more than one person, and not everybody might eat the same foods. Thus, accuracy of loyalty card data in investigating individual diet may not be as high as that obtained with traditional data collection methods. However, good compatibility between respondent-collected household-level food purchase data and individual-level dietary data has been demonstrated [28,48]. Moreover, foods purchased from stores do not necessarily indicate foods consumed owing to many different reasons. These include, for example, foods that are not included in loyalty card data, such as dinner foods at restaurants or lunch foods at work. Cardholders may also buy foods that are consumed by others, for example, grandchildren, other relatives or close friends invited for dinner, and pets. Some foods are not consumed at all, resulting in food wastage [49,50].

Finally, the degree of loyalty remains a challenge. In this study, the degree of loyalty was self-estimated, and it is difficult to estimate possible bias related to this self-report. However, we showed that the RFM scores increased steadily with the groups of loyalty, indicating that higher frequency, higher engagement,

XSL•FO
RenderX

and more money spent on grocery purchases were associated with a higher degree of loyalty. Moreover, a positive correlation was found between the proportional increases in money spent (€) and weight (kg) regarding food purchases and the degree of loyalty. These results suggest that this self-assessment seemed to provide a feasible estimate of the true values. In another study, the researchers defined loyalty through the frequency of purchases made in the supermarket combined with the amount of money spent on purchases. However, in this study, the degree of loyalty was not specifically defined [51].

Despite its limitations, we see real potential in the use of these automatically collected longitudinal food purchase data in the population-based assessment of dietary patterns, which are important determinants of health and carbon footprint [52]. Loyalty card data provide a cost-effective tool to reach large groups of individuals with minimum data collection efforts and to investigate diet-related behaviors with less information bias. Linking these data with other health data (such as electronic health records and health registers) would provide new opportunities to understand diet and related outcomes. However, such research settings include privacy concerns that need to be carefully addressed to guarantee individual anonymity and consent. In addition, loyalty card data enable the monitoring of longitudinal trends in food purchases including timely monitoring and evaluation of the impact of population-level steering instruments such as taxation.

Conclusions

Individuals who consented to the use of their loyalty card data for research purposes tended to differ from individuals in the general Finnish adult population. The sociodemographic distributions were toward similar characteristics, as is frequently seen in health and nutrition studies. However, the high volume of data enabled the inclusion of sociodemographically heterogeneous subgroups, potentially including hard-to-reach subgroups, and further correction of the differences so that distributions matched well with those of the general Finnish adult population. A potential confounder in studies using loyalty card data is the degree of loyalty, which in this study, was associated with food-purchasing profiles and also the participants' background characteristics. This underlines the importance of obtaining sufficient background information when using loyalty card data for health research.

Despite the limitations, loyalty card data provide a cost-effective approach for large groups of individuals with minimum data collection effort and for the investigation of diet-related behaviors on a large scale with less information bias. Importantly, loyalty card data enable the monitoring of longitudinal trends in grocery purchases.

Acknowledgments

We thank S Group for collaboration. We are also grateful to the loyalty card holders who provided consent for the use of their loyalty card data in this research project. This work was funded by Tampere University, the University of Helsinki, the Finnish Foundation for Alcohol Studies, and EIT Food ("Towards a smarter shopping list" #20041). EIT Food is the Innovation Community on Food of the European Innovation and Technology (EIT), a body of the EU, under Horizon 2020, the EU Framework Programme for Research and Innovation.

Authors' Contributions

ME, JN, HS, LU, and MF participated in data collection and transfer. All authors participated in the design of the study. TN, ME, SK, and JN performed data management. A-LV and JN planned the data analyses, which A-LV conducted. A-LV wrote the original draft. All authors participated in reviewing and writing the drafts, including approval of the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Grouping of the food purchase data.
[DOCX File , 16 KB - jmir_v22i7e18059_app1.docx]

Multimedia Appendix 2
The percentage of individuals living in each of the 19 regions in Finland, in the LoCard sample, and in the weighted LoCard sample.
[DOCX File , 15 KB - jmir_v22i7e18059_app2.docx]

Multimedia Appendix 3
Regions of Finland.
[PNG File , 33 KB - jmir_v22i7e18059_app3.png]

References

XSL•FO
RenderX

1. Schwingshackl L, Knüppel S, Michels N, Schwedhelm C, Hoffmann G, Iqbal K, et al. Intake of 12 food groups and disability-adjusted life years from coronary heart disease, stroke, type 2 diabetes, and colorectal cancer in 16 European countries. Eur J Epidemiol 2019 Aug;34(8):765-775 [FREE Full text] [doi: 10.1007/s10654-019-00523-4] [Medline: 31030306]
2. Global action plan for the prevention and control of NCDs 2013-2020. World Health Organization. 2013. URL: https://www.who.int/nmh/publications/ncd-action-plan/en/ [accessed 2020-06-10]
3. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
4. Willett W. Nutritional Epidemiology. Oxford: Oxford University Press; 2013.
5. Lovefrove J, Hodson L, Sharma S, Lanham-New S. Nutrition Research Methodologies. Hoboken, New Jersey: Wiley-Blackwell; 2015.
6. Murakami K, Livingstone M. Prevalence and characteristics of misreporting of energy intake in US children and adolescents: National Health and Nutrition Examination Survey (NHANES) 2003-2012. Br J Nutr 2016 Jan 28;115(2):294-304. [doi: 10.1017/S0007114515004304] [Medline: 26525591]
7. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol 2017 Nov 01;186(9):1026-1034 [FREE Full text] [doi: 10.1093/aje/kwx246] [Medline: 28641372]
8. Strandhagen E, Berg C, Lissner L, Nunez L, Rosengren A, Torén K, et al. Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. Eur J Epidemiol 2010 Mar;25(3):163-172. [doi: 10.1007/s10654-010-9427-7] [Medline: 20127393]
9. Kesse-Guyot E, Andreeva V, Castetbon K, Vernay M, Touvier M, Méjean C, et al. Participant profiles according to recruitment source in a large Web-based prospective study: experience from the Nutrinet-Santé study. J Med Internet Res 2013 Sep 13;15(9):e205 [FREE Full text] [doi: 10.2196/jmir.2488] [Medline: 24036068]
10. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
11. Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: Epidemiology in the era of big data. Epidemiology 2015 May;26(3):390-394 [FREE Full text] [doi: 10.1097/EDE.0000000000000274] [Medline: 25756221]
12. Tin ST, Mhurchu CN, Bullen C. Supermarket sales data: feasibility and applicability in population food and nutrition monitoring. Nutr Rev 2007 Jan;65(1):20-30. [doi: 10.1111/j.1753-4887.2007.tb00264.x] [Medline: 17310856]
13. Kesse-Guyot E, Assmann K, Andreeva V, Castetbon K, Méjean C, Touvier M, et al. Lessons Learned From Methodological Validation Research in E-Epidemiology. JMIR Public Health Surveill 2016 Oct 18;2(2):e160 [FREE Full text] [doi: 10.2196/publichealth.5880] [Medline: 27756715]
14. Méjean C, Szabo de Edelenyi F, Touvier M, Kesse-Guyot E, Julia C, Andreeva VA, et al. Motives for participating in a web-based nutrition cohort according to sociodemographic, lifestyle, and health characteristics: the NutriNet-Santé cohort study. J Med Internet Res 2014 Aug 07;16(8):e189 [FREE Full text] [doi: 10.2196/jmir.3161] [Medline: 25135800]
15. Pietilä J, Helander E, Korhonen I, Myllymäki T, Kujala UM, Lindholm H. Acute Effect of Alcohol Intake on Cardiovascular Autonomic Regulation During the First Hours of Sleep in a Large Real-World Sample of Finnish Employees: Observational Study. JMIR Ment Health 2018 Mar 16;5(1):e23 [FREE Full text] [doi: 10.2196/mental.9519] [Medline: 29549064]
16. Helander EE, Vuorinen A, Wansink B, Korhonen IK. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
17. Sperrin M, Rushton H, Dixon WG, Normand A, Villard J, Chieh A, et al. Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England. J Med Internet Res 2016 Jan 21;18(1):e17 [FREE Full text] [doi: 10.2196/jmir.4767] [Medline: 26794900]
18. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013;9(10):e1003256 [FREE Full text] [doi: 10.1371/journal.pcbi.1003256] [Medline: 24146603]
19. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela J, Keshavan M. Smartphone-Based Tracking of Sleep in Depression, Anxiety, and Psychotic Disorders. Curr Psychiatry Rep 2019 Jun 04;21(7):49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
20. Ekman A, Litton J. New times, new needs; e-epidemiology. Eur J Epidemiol 2007;22(5):285-292. [doi: 10.1007/s10654-007-9119-0] [Medline: 17505896]
21. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
22. Guan L, Peng T, Zhu JJ. Who is Tracking Health on Mobile Devices: Behavioral Logfile Analysis in Hong Kong. JMIR Mhealth Uhealth 2019 May 23;7(5):e13679 [FREE Full text] [doi: 10.2196/13679] [Medline: 31120429]
23. Hargittai E. Potential Biases in Big Data: Omitted Voices on Social Media. Social Science Computer Review 2018 Jul 30;38(1):10-24 [FREE Full text] [doi: 10.1177/0894439318788322]

XSL•FO
RenderX

24. Nevalainen J, Erkkola M, Saarijärvi H, Näppilä T, Fogelholm M. Large-scale loyalty card data in health research. Digit Health 2018;4:2055207618816898 [FREE Full text] [doi: 10.1177/2055207618816898] [Medline: 30546912]
25. Mamiya H, Moodie EE, Buckeridge DL. A novel application of point-of-sales grocery transaction data to enhance community nutrition monitoring. AMIA Annu Symp Proc 2017;2017:1253-1261 [FREE Full text] [Medline: 29854194]
26. Bandy L, Adhikari V, Jebb S, Rayner M. The use of commercial food purchase data for public health nutrition research: A systematic review. PLoS One 2019;14(1):e0210192 [FREE Full text] [doi: 10.1371/journal.pone.0210192] [Medline: 30615664]
27. Einav L, Leibtag E, Nevo A. Recording discrepancies in Nielsen Homescan data: Are they present and do they matter? Quant Mark Econ 2009 Aug 25;8(2):207-239. [doi: 10.1007/s11129-009-9073-0]
28. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? Int J Behav Nutr Phys Act 2017 Apr 11;14(1):46 [FREE Full text] [doi: 10.1186/s12966-017-0502-2] [Medline: 28399887]
29. French SA, Wall M, Mitchell NR, Shimotsu ST, Welsh E. Annotated receipts capture household food purchases from a broad range of sources. Int J Behav Nutr Phys Act 2009 Jul 01;6:37 [FREE Full text] [doi: 10.1186/1479-5868-6-37] [Medline: 19570234]
30. Finnish Grocery Trade Association. Finnish Grocery Trade 2019. Helsinki: Finnish Grocery Trade Association; 2019.
31. Nutrition Unit of the National Institute for Health and Welfare (THL). Food Composition Database Release 20. Fineli. 2019. URL: https://fineli.fi [accessed 2019-01-01]
32. Fogelholm M, Anderssen S, Gunnarsdottir I, Lahti-Koski M. Dietary macronutrients and food consumption as determinants of long-term weight change in adult populations: a systematic literature review. Food Nutr Res 2012;56 [FREE Full text] [doi: 10.3402/fnr.v56i0.19103] [Medline: 22893781]
33. Fardet A, Boirie Y. Associations between food and beverage groups and major diet-related chronic diseases: an exhaustive review of pooled/meta-analyses and systematic reviews. Nutr Rev 2014 Dec;72(12):741-762. [doi: 10.1111/nure.12153] [Medline: 25406801]
34. StatFin statistical database. URL: http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/ [accessed 2019-06-26]
35. Koponen P, Borodulin K, Lundqvist A, Sääksjärvi K, Koskinen S, editors. Terveys, toimintakyky ja hyvinvointi Suomessa FinTerveys 2017-tutkimus. Helsinki: Terveyden ja hyvinvoinnin laitos (THL); 2018.
36. Valsta L, Kaartinen N, Tapanainen H, Männistö S, Sääksjärvi K, editors. Ravitsemus Suomessa - FinRavinto 2017 -tutkimus / Nutrition in Finland - The National FinDiet 2017 Survey,?. Helsinki: Terveyden ja hyvinvoinnin laitos (THL); 2018.
37. Lumley T. Analysis of Complex Survey Samples. J. Stat. Soft 2004;9(8):1-19 R package verson 2.2. [doi: 10.18637/jss.v009.i08]
38. Deming WE, Stephan FF. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. Ann. Math. Statist 1940 Dec;11(4):427-444. [doi: 10.1214/aoms/1177731829]
39. Hebbali A. rfm: Recency, Frequency and Monetary Value Analysis. R package. 2020. URL: https://cran.r-project.org/web/packages/rfm/index.html [accessed 2020-06-16]
40. Ekman A, Dickman PW, Klint A, Weiderpass E, Litton J. Feasibility of using web-based questionnaires in large population-based epidemiological studies. Eur J Epidemiol 2006;21(2):103-111. [doi: 10.1007/s10654-005-6030-4] [Medline: 16518678]
41. Andreeva VA, Salanave B, Castetbon K, Deschamps V, Vernay M, Kesse-Guyot E, et al. Comparison of the sociodemographic characteristics of the large NutriNet-Santé e-cohort with French Census data: the issue of volunteer bias revisited. J Epidemiol Community Health 2015 Sep;69(9):893-898. [doi: 10.1136/jech-2014-205263] [Medline: 25832451]
42. Reinikainen J, Tolonen H, Borodulin K, Härkänen T, Jousilahti P, Karvanen J, et al. Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. Eur J Public Health 2018 Apr 01;28(2):237-243. [doi: 10.1093/eurpub/ckx151] [Medline: 29036286]
43. Official Statistics of Finland. Väestön Tieto- Ja Viestintätekniikan Käyttö [Verkkojulkaisu]. Liitetaulukko 5. Kotitaloudessa internetyhteys 2017, %-osuus talouksista. Helsinki: Statistics Finland; 2017.
44. Official Statistics of Finland. Väestön tieto- ja viestintätekniikan käyttö [verkkojulkaisu]. Liitetaulukko 20. Internetin käyttö puheluihin, pikaviestintään älypuhelimella, sähköpostiin ja pilvitallentamiseen. Helsinki: Statistics Finland; 2017.
45. Aiello LM, Schifanella R, Quercia D, Del Prete L. Large-scale and high-resolution analysis of food purchases and health outcomes. EPJ Data Sci 2019 Apr 30;8(1):14 [FREE Full text] [doi: 10.1140/epjds/s13688-019-0191-y]
46. Giskes K, Avendano M, Brug J, Kunst AE. A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. Obes Rev 2010 Jun;11(6):413-429. [doi: 10.1111/j.1467-789X.2009.00658.x] [Medline: 19889178]
47. Fogelholm M, Kanerva N, Männistö S. Association between red and processed meat consumption and chronic diseases: the confounding role of other dietary factors. Eur J Clin Nutr 2015 Sep;69(9):1060-1065. [doi: 10.1038/ejcn.2015.63] [Medline: 25969395]
48. Becker W. Comparability of household and individual food consumption data--evidence from Sweden. Public Health Nutr 2001 Oct;4(5B):1177-1182. [Medline: 11924944]

XSL•FO
RenderX

49. Saarijärvi H. Customer Value Co-Creation through Reverse Use of Customer Data. Tampere: Acta Universitatis Tamperensis; 2011.
50. Katajajuuri J, Silvennoinen K, Hartikainen H, Heikkilä L, Reinikainen A. Food waste in the Finnish food chain. Journal of Cleaner Production 2014 Jun;73:322-329. [doi: 10.1016/j.jclepro.2013.12.057]
51. Hansel B, Roussel R, Diguet V, Deplaude A, Chapman MJ, Bruckert E. Relationships between consumption of alcoholic beverages and healthy foods: the French supermarket cohort of 196,000 subjects. Eur J Prev Cardiol 2015 Feb;22(2):215-222. [doi: 10.1177/2047487313506829] [Medline: 24065742]
52. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet 2019 Feb 02;393(10170):447-492. [doi: 10.1016/S0140-6736(18)31788-4] [Medline: 30660336]

Abbreviations

IPF: iterative proportional fitting
RFM: recency, frequency, and monetary

Edited by G Eysenbach; submitted 31.01.20; peer-reviewed by M Sperrin, M Morris; comments to author 01.04.20; revised version received 18.04.20; accepted 14.05.20; published 15.07.20.

Please cite as:
Vuorinen AL, Erkkola M, Fogelholm M, Kinnunen S, Saarijärvi H, Uusitalo L, Näppilä T, Nevalainen J
Characterization and Correction of Bias Due to Nonparticipation and the Degree of Loyalty in Large-Scale Finnish Loyalty Card Data on Grocery Purchases: Cohort Study
J Med Internet Res 2020;22(7):e18059
URL: http://www.jmir.org/2020/7/e18059/
doi:10.2196/18059
PMID:32459633

©Anna-Leena Vuorinen, Maijaliisa Erkkola, Mikael Fogelholm, Satu Kinnunen, Hannu Saarijärvi, Liisa Uusitalo, Turkka Näppilä, Jaakko Nevalainen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ontological Organization and Bioinformatic Analysis of Adverse Drug Reactions From Package Inserts: Development and Usability Study

Xiaoying Li[1*], PhD; Xin Lin[1*], MA; Huiling Ren[1], MA; Jinjing Guo[1], MA

Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Huiling Ren, MA
Institute of Medical Information
Chinese Academy of Medical Sciences
69 Dongdan N St
Beijing, 100005
China
Phone: 86 10 52328911
Email: ren.huiling@imicams.ac.cn

Abstract

Background: Licensed drugs may cause unexpected adverse reactions in patients, resulting in morbidity, risk of mortality, therapy disruptions, and prolonged hospital stays. Officially approved drug package inserts list the adverse reactions identified from randomized controlled clinical trials with high evidence levels and worldwide postmarketing surveillance. Formal representation of the adverse drug reaction (ADR) enclosed in semistructured package inserts will enable deep recognition of side effects and rational drug use, substantially reduce morbidity, and decrease societal costs.

Objective: This paper aims to present an ontological organization of traceable ADR information extracted from licensed package inserts. In addition, it will provide machine-understandable knowledge for bioinformatics analysis, semantic retrieval, and intelligent clinical applications.

Methods: Based on the essential content of package inserts, a generic ADR ontology model is proposed from two dimensions (and nine subdimensions), covering the ADR information and medication instructions. This is followed by a customized natural language processing method programmed with Python to retrieve the relevant information enclosed in package inserts. After the biocuration and identification of retrieved data from the package insert, an ADR ontology is automatically built for further bioinformatic analysis.

Results: We collected 165 package inserts of quinolone drugs from the National Medical Products Administration and other drug databases in China, and built a specialized ADR ontology containing 2879 classes and 15,711 semantic relations. For each quinolone drug, the reported ADR information and medication instructions have been logically represented and formally organized in an ADR ontology. To demonstrate its usage, the source data were further bioinformatically analyzed. For example, the number of drug-ADR triples and major ADRs associated with each active ingredient were recorded. The 10 ADRs most frequently observed among quinolones were identified and categorized based on the 18 categories defined in the proposal. The occurrence frequency, severity, and ADR mitigation method explicitly stated in package inserts were also analyzed, as well as the top 5 specific populations with contraindications for quinolone drugs.

Conclusions: Ontological representation and organization using officially approved information from drug package inserts enables the identification and bioinformatic analysis of adverse reactions caused by a specific drug with regard to predefined ADR ontology classes and semantic relations. The resulting ontology-based ADR knowledge source classifies drug-specific adverse reactions, and supports a better understanding of ADRs and safer prescription of medications.

(J Med Internet Res 2020;22(7):e20443) doi:10.2196/20443

KEYWORDS

ontology; adverse drug reactions; package inserts; information retrieval; natural language processing; bioinformatics; drug; adverse events; machine-understandable knowledge; clinical applications

XSL•FO
RenderX

Introduction

Overview

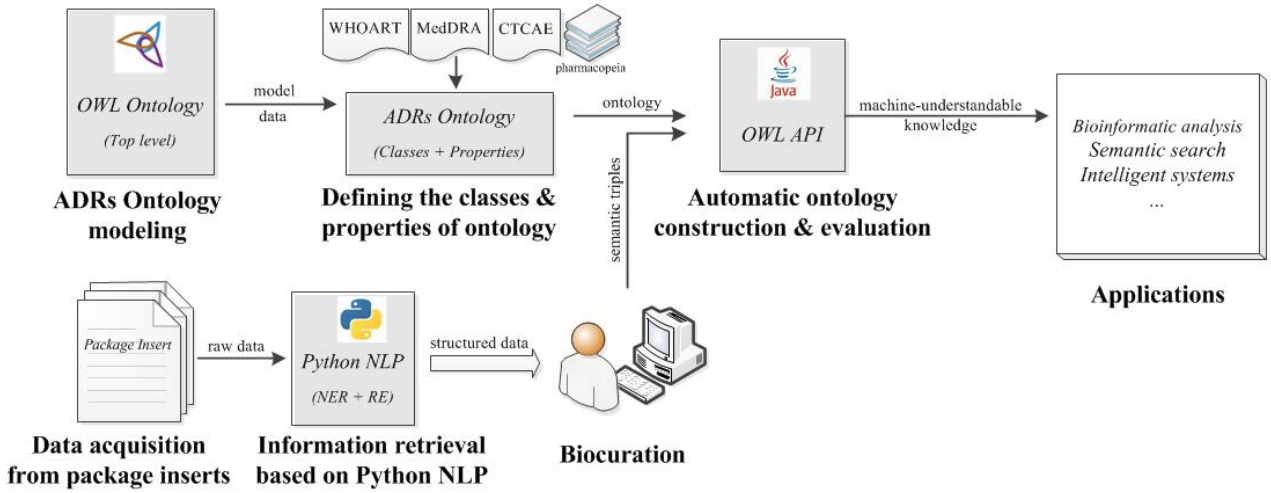
Chemicals and drugs have made a great contribution to human health care. At the same time, they are rarely free from occasional adverse drug reactions (ADRs) [1], which are defined by the World Health Organization (WHO) as any noxious, unintended, and undesired effects of a drug that occur at doses used for the prevention, diagnosis, and treatment of a disorder. A significant number of ADRs occur each year. ADRs are the sixth leading cause of death worldwide, and the fourth primary cause of death in the United States and Canada, behind cardiovascular disease, malignant neoplasm, and stroke [2]. Although the actual incidence of ADRs is difficult to access precisely, it is known that ADRs have a considerable impact upon both health care and pharmaceutical manufacturers.

Package inserts (sometimes called patient information leaflets) are the primary official papers which accompany most prescribed drugs and over-the-counter medications. Although different countries have diverse requirements for the obligatory contents, the package inserts serve at least two main purposes. They contain informative details regarding the generic names of drugs, active ingredients, indication for use, instructions for use, special warnings, contraindications, and statistical values from clinical trials, including the percentage of people who had side effects, the types of side effects, and additional precautions. Furthermore, the package insert is an easy reference for physicians when prescribing medications, and can help them avoid prescribing drugs that may be contraindicated. The inserts also serve as an easy reference for patients. However, the informative package inserts are generally semistructured and cannot be understood easily by machines. As the number of newly licensed drugs increases, the demand for automatic technology for semantic integration and linkages, as well as bioinformatic analysis of the information (including ADRs) enclosed in package inserts, has become an urgent issue both in biomedical research and the pharmaceutical industry.

In information science, an ontology is the formal, explicit specification of a shared conceptualization of a domain [3]. Generally, biomedical ontologies not only represent the essential properties of biomedical entities and their correlations to other biomedical concepts, but also provide a standardized vocabulary and formalized knowledge source for the biomedical community. Hundreds of biomedical ontologies have been elaborately built to support scientific discovery and the analysis of biomedical data [4-8]. Moreover, systematically evaluated ontology is one of the two fundamental sources of background knowledge for artificial intelligence algorithms in biomedicine (the other is the knowledge graph).

In this paper, we propose an ontological organization of traceable ADR information extracted from licensed package inserts, which aims to provide machine-understandable knowledge for bioinformatics analysis, semantic retrieval, and intelligent clinical applications (Figure 1). This entails the following: (1) Present a generic ADR ontology model from two dimensions (and nine subdimensions) covering the essential ADR information and medication instructions. (2) Customize a Python natural language processing (NLP) method to automatically retrieve the identified information enclosed in package inserts. (3) Collect the approved package inserts of quinolone drugs and build a specialized ontology for algorithm verification and validation. (4) Bioinformatically analyze the adverse reactions caused by quinolones based on the obtained ADR ontology, and discuss potential applications including semantic retrieval and a clinical decision-making system.

Figure 1. Study framework. ADRs: adverse drug reactions; API: application programming interface; CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary for Regulatory Activities; NER: named entity recognition; NLP: natural language processing; OWL: Ontology Web Language; RE: relation extraction; WHOART: WHO Adverse Reactions Terminology.



XSL•FO
RenderX

Related Works

Identification of ADRs

In recent decades, many studies have identified ADRs through diverse channels of information, including patient reports [9-14], electronic health records [15-18], PubMed literature [19,20], and social media [21-26]. Briefly, a patient report is a traditional source of information, where the users of a drug spontaneously report medication side effects to health authorities; electronic health records contain comprehensive medications and procedures as recorded by physicians; PubMed collects rich and up-to-date published clinical trials and other types of biomedical publications concerning drugs' adverse reactions; and social media represents a new data source of patient experiences with drugs and could be characterized by its high volume and quick availability. Although patients, health professionals, research scientists, and even the public have increasingly contributed to ADR reporting, the role of pharmaceutical companies in reporting ADRs cannot be neglected; package inserts play a significant part in medication safety.

Ontologies of ADRs

Several studies about building ontologies for ADRs for different applications have already been carried out. For instance, the Adverse Drug Reaction Classification System (ADReCS) was developed as a comprehensive ADR ontology database, which enabled standardization and provided hierarchical classification of ADR terms for a molecular understanding of drug safety in the laboratory, and use in bioinformatics and systems biology for toxicological research [27]. Additionally, an ontology of ADRs (OADRs) was built to describe the semantics of ADR terms for automated signal generation in pharmacovigilance [28].

Adverse drug events (ADEs) refer to the injuries resulting from medical interventions related to drugs [1], which include medication errors, ADRs, allergic reactions, overdoses, and other events associated with the prescription, preparation, dispensation, or administration of medications. Therefore, research efforts concentrated on ontology-based representation and analysis of ADEs indirectly related to our work. Among them, the Ontology of Adverse Events (OAE) has recently garnered research attention; it represents numerous adverse events related to medical intervention, time at medical intervention, pathological bodily process, patient information (especially patient age), and other adverse event–related terms imported from existing ontologies, as well as clinical adverse event reports [29]. To better analyze adverse events related to vaccines and support safety studies of vaccines, the authors further expanded OAE and developed the Ontology of Vaccine Adverse Events (OVAE) by analyzing the adverse events recorded in the official packet inserts of licensed vaccines [30]. All the ontologies mentioned above provide a fundamental basis on which to conduct our study and will be compared with the proposed ADR ontology from several perspectives.

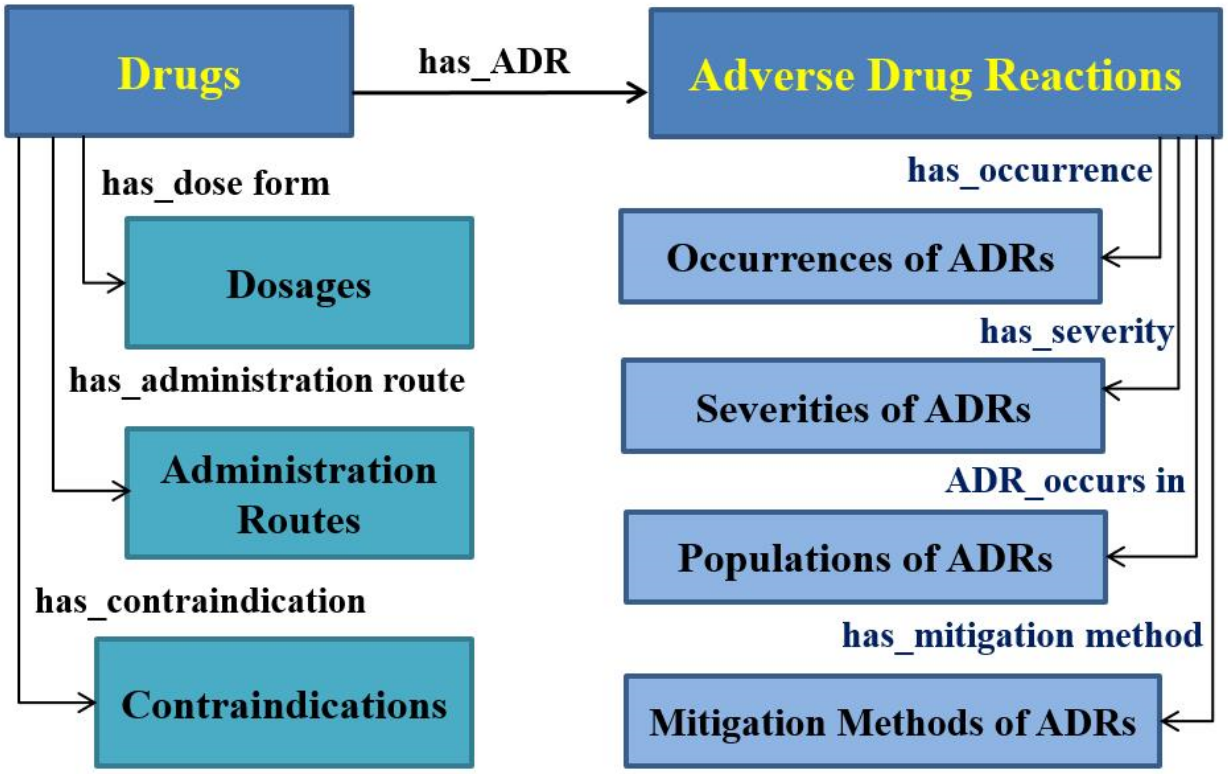
Methods

Ontology Modeling

We built the ADR ontology using two dimensions: ADR information and drug-related medication instructions, which were further separated into nine subdimensions. The former covers the following: (1) adverse drug reactions, presenting diverse types of ADRs; (2) the occurrence of ADRs, describing the frequencies of ADRs after the administration of a drug in a population; (3) the severity of ADRs, describing a general measure of the subsequent risks of potential ADRs; (4) populations affected by specific ADRs, noting the individual human patients associated with an adverse reaction after the administration of a drug; and (5) the mitigation methods of ADRs, referring to any measure that shortens the duration of an adverse reaction or reduces its severity. The drug-related medication instructions contain generic information about rational drug use: (6) drug names, describing different pharmaceutical agents; (7) dosage forms of drugs, collecting the complete form of the pharmaceutical preparation used to administer the prescribed dose of medication; (8) administration routes of drugs, consisting of the various ways of administering a drug to a patient to allow the chemical to be absorbed into the blood and delivered to the target tissue; and (9) contraindications of drugs, predefining a condition or factor associated with a recipient that makes the use of a specific drug improper or inadvisable. These nine branches determine the fundamental concepts and classes of our ontology. Figure 2 demonstrates the nine classes of proposed ADR ontology as well as the eight semantic relations among them. The generic model will help to generate specialized ADR ontologies on request as illustrated in the next section.

XSL•FO
RenderX

Figure 2. The generic model of ADRs ontology. ADRs: adverse drug reactions.



Classes of Ontology and Their Hierarchy and Properties

The biomedical terms denoting nine classes within the ADR ontology are much more comprehensive than shown in Figure 1. With reference to controlled biomedical vocabularies and well-developed pharmacopeia, we integrated many synonymous terms (both in English and Chinese) into various concepts and organized them in a hierarchical way. Generally, the WHO Adverse Reactions Terminology (WHOART) is reused to generate the ADR class and its hierarchy in four levels: system/organ classes (SOC), high level terms (HLT), preferred terms (PT), and included terms (IT). Among them, SOC and HLT consist of broad grouping terms, while PT represents more specific adverse reactions and IT are entry terms (synonyms) for PT. The large number of PT and IT as well as their synonyms from the Medical Dictionary for Regulatory Activities (MedDRA) will become the essential vocabulary to recognize the named entities of ADRs in package inserts. Furthermore, the Council for International Organization of Medical Science (CIMOS) has recommended five terms derived from different percentages to classify the occurrences of ADRs: Very Rare (<0.01%), Rare (≥0.01%, <0.1%), Uncommon (≥0.1%, <1%), Common (≥1%, <10%), and Very Common (≥10%). This forms the basis for the classification of ADR frequencies in our work. The Common Terminology Criteria for Adverse Events (CTCAE) is a standard classification and severity grading scale for adverse events in clinical trials and oncology settings. We adopted CTCAE for labeling the severities of ADRs with five levels: Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe but not immediately life-threatening), Grade 4 (life-threatening) and Grade 5 (death caused by ADRs). The other classes that are closely related to the pharmaceutical agents and their medication instructions will be acquired and represented according to the classic pharmacopeia for ease of use.

In addition, two kinds of data properties will be incorporated into the ADR ontology, including a general description of the drugs (such as the active ingredient, injection excipients, drug specifications, antibacterial mechanism, interaction drugs); the key codes and definitions from referenced vocabularies containing ARecNO (the PT code) and the SOC code from WHOART, the MedDRA code, the NCIt code; and definitions from CTCAE. Furthermore, the origin of the package inserts will be recorded in the ontology's annotation property for ontology data identification and traceability.

Information Retrieval From Package Inserts

From the viewpoint of ontology construction, the ADR information and medication instructions extracted from package inserts will become the instances of object properties within the ADR ontology. Due to the large amount of data enclosed in package inserts, manual extraction would be a labor-intensive process. We developed a Python NLP-based algorithm to automatically retrieve ADRs and drug-related information, which consists of two steps: named entity recognition (NER) and relation extraction (RE).

Briefly, NER will recognize a string of text as an entity (eg, an adverse reaction) that is already defined in our ADR ontology. The Jieba word segmentation model implemented in Python is adopted to segment the words enclosed in package inserts, while the names of classes and their synonyms from the proposed ontology will function as a domain vocabulary to improve performance. RE is a process that determines whether two

XSL•FO
RenderX

entities have a specific relationship (eg, the "has_ADR" causality between a particular drug and an adverse reaction). Since formal package inserts have already been separated into several titled sections, these titles will be used to implement the RE task. Specifically, the titles "ADVERSE REACTIONS" and "WARNINGS AND PRECAUTIONS" are converted into semantic relations about ADR information, while the medication instructions for a particular drug would be extracted from the sections titled "DOSAGE AND ADMINISTRATION," "CONTRAINDICATIONS," and "USE IN SPECIFIC POPULATIONS."

The information automatically retrieved from package inserts will be passed to downstream biocuration for data identification. The major criteria of the manual review process emphasize the following two points: (1) whether the medication instructions and information about ADRs caused by a specific drug was accurate and complete, without missing data or mistakes; and (2) the frequency and severity of a drug-ADR triple must be explicit; vague descriptions were not recorded. Eventually, the human biocurated semantic information is used to build the ADR ontology.

Automated Ontology Construction and Evaluation

The Ontology Web Language (OWL) is a widely used programing language for defining and instantiating web-based ontologies. It provides a machine-understandable schema to describe classes and their semantic relations in a specific domain. In this work, the OWL application programming interface, a Java interface and implementation for OWL, is used to build the ADR ontology automatically based on content data primarily obtained from package inserts and well-established vocabularies (eg, WHOART, MedDRA, CTCAE). Moreover, to ensure high-quality results, the ADR ontology will be evaluated and validated by checking the clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment, which are the fundamental principles of building a domain ontology [31].

Results

Data Collection and Ontology Construction

Quinolone drugs have become commonly used antibacterial agents due to their strong and broad-spectrum antibacterial activity, as well as their rapid and complete absorption in humans. As quinolone usage increases, the risk of ADRs increases proportionally. According to the Annual Report of National ADR Monitoring in China, the number of adverse reactions and events from quinolones has continuously been the second-highest among antibacterial drugs. Therefore, the safe administration of quinolones is a serious matter that requires more attention.

After half a century of development, quinolones have evolved from the first generation to the fourth generation. We collected 165 specific drug names of quinolones from the China Pharmaceutical Reference and the National Essential Drugs List in China, and further organized them based on their generation. The electronic package inserts of these drugs were then downloaded from the National Medical Products Administration and Yaozhi drug database in China. Although these package inserts were written in Chinese, the Chinese and English synonyms of the class names defined in the proposed ADR ontology will enable information extraction based on a customized Python program. Finally, an ADR ontology for quinolones was automatically built and evaluated for further analysis.

Ontology Statistics and Visualization

Currently, the specialized ADR ontology covers a total of 2879 classes grouped into nine categories: ADRs and their occurrence, severity, population, and mitigation methods, as well as 165 drugs and their dosages, administration routes, and contraindications (Figure 3A). These classes were deeply divided into subclasses on the basis of biomedical concepts and arrayed hierarchically from most general to most specific in up to 4 levels, with the abovementioned nine categories at the top (Level 1) of the ontology hierarchy (Figure 3B). Furthermore, the obtained ADR ontology also includes eight types and 15,711 nonredundant semantic triples (Figure 4A) extracted from package inserts, where the causalities between ADRs and quinolones (with the type label "has_ADR") account for a large proportion (n=7725, 49.17%). Among the 7725 identified drug-ADR triples, 4043 (52.34%) and 716 (9.27%) explicitly stated the frequency and severity, respectively. The statistics of drug-ADR triples by frequency and severity suggest that most ADRs caused by quinolones (n=4037, 52.26%) occur at a low frequency (<10%), and none of them are life-threatening (Figure 4B).

Figure 5 demonstrates how the proposed ontology organized the ADR information enclosed in the package insert, using the Protégé OWL editor for ontology visualization. Briefly, levofloxacin hydrochloride tablets have been reported to account for different types of ADRs (including dizziness, headache, and insomnia) that have already been defined in our ontology. For each adverse reaction, it is likely that the frequency, severity, and mitigation method are informed by the patient population. Therefore, we used four semantic relation types ("has_occurrence," "has_severity," "has_mitigation method," and "has_population") to link information associated with a particular adverse reaction caused by a specific drug (the lower right section of Figure 5). The "AND" clause is adopted to combine multiple semantic triples into a composite knowledge unit. It is worth mentioning that we extracted ADR information from the explicitly stated content in package inserts; thus, some of the four types linked to a given drug-ADR triple may not be obtained due to a missed or implicit description. To enable knowledge tracing, the package insert citation was recorded as a referenced annotation property (the upper right section of Figure 5). In addition, the Internationalized Resource Identifier (IRI) [32], which is the unique identifier for ontology sharing and reuse around the world, was customized as "ADR+six numbers" within the proposed ADR ontology. For instance, the IRI for the levofloxacin hydrochloride tablets used in the example is ADR000572.

XSL•FO
RenderX

Figure 3. Statistics of the classes in ADRs ontology. ADRs: adverse drug reactions.

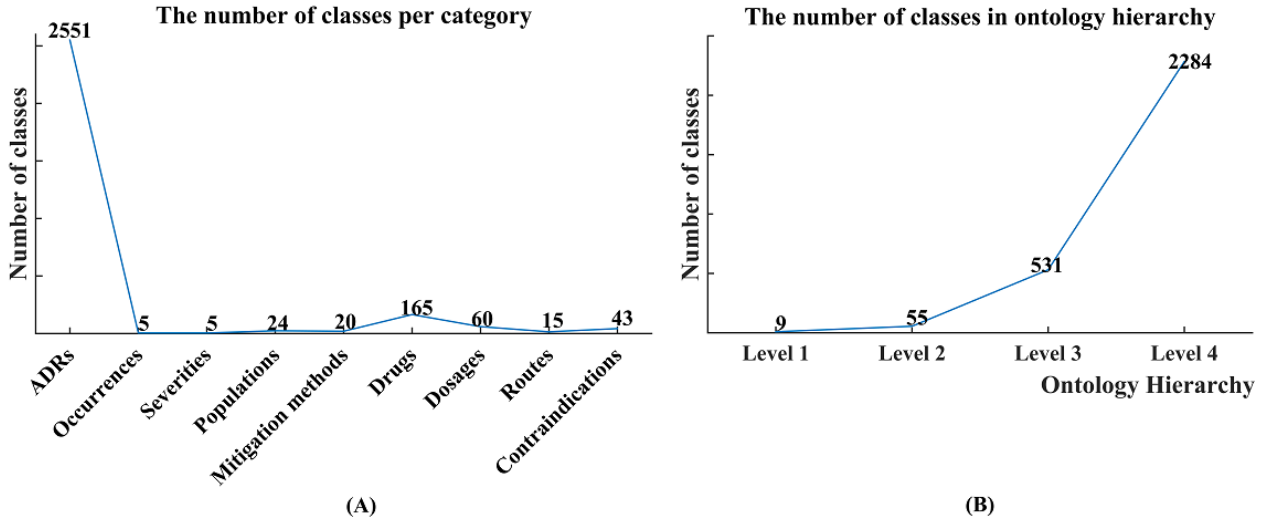


Figure 4. Statistics of the semantic relations in ADRs ontology. ADRs: adverse drug reactions.

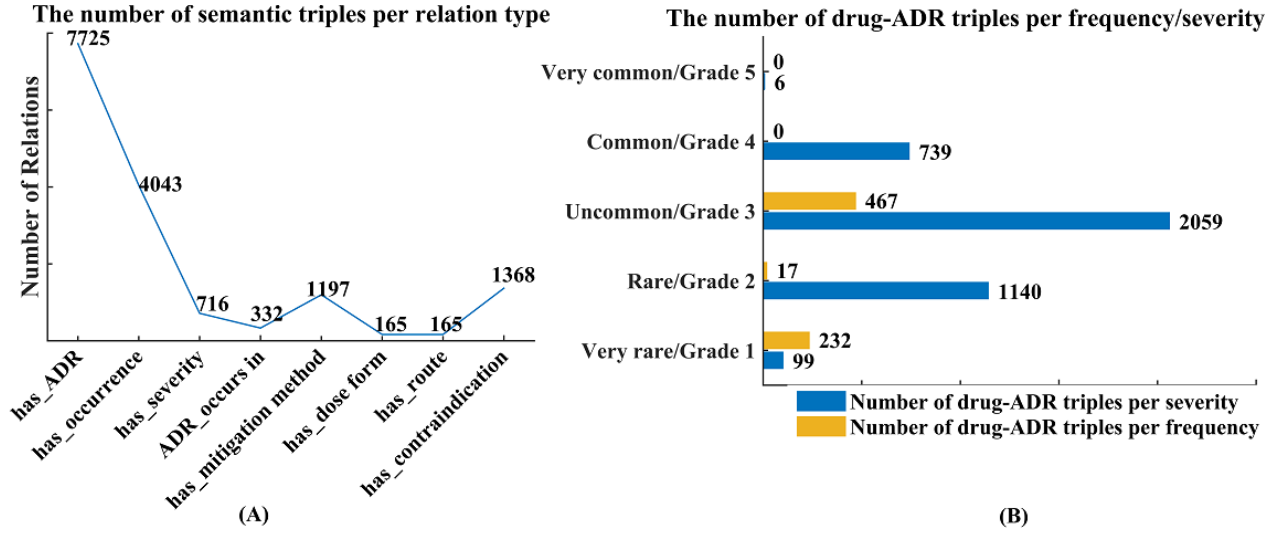
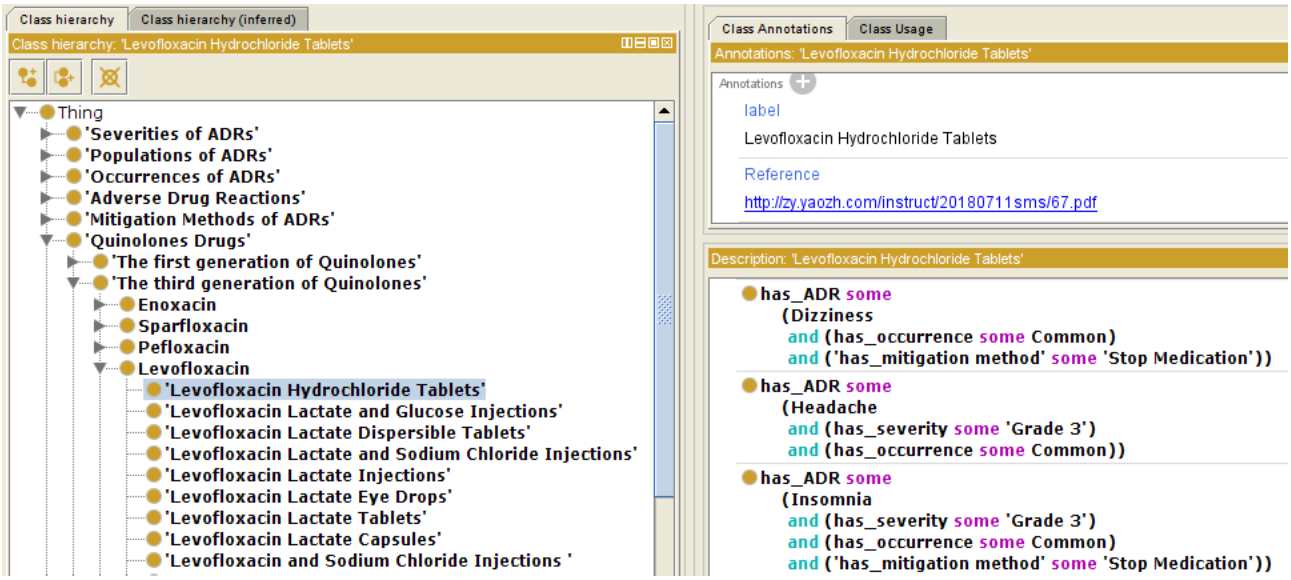


Figure 5. Ontology visualization using levofloxacin hydrochloride tablets as an example. ADRs: adverse drug reactions.



XSL•FO
RenderX

Comparison With Existing Ontologies

We compared our work with four other ontologies (described in the Introduction) in terms of the semantic relations defined in the proposed ADR ontology (Table 1), since they are the essential knowledge unit for the wide utility of an ontology. The proposed ontology represents a complete set of information pertaining to ADRs containing drug-ADR triples and the associated frequency, severity, and mitigation methods. Moreover, the explicit medication instructions (eg, the dosage form, route, and contraindication) extracted from the package insert of a particular drug support physicians in guiding patients in the safe, effective, and rational use of drugs, as an adverse reaction can occur when using the prescribed dosage.

Table 1. Comparison with existing ontologies.

Semantic relations of the proposed ADR[a] ontology	Definition	Adverse Drug Reaction Classification System	Ontology of ADRs	Ontology of Adverse Events	Ontology of Vaccine Adverse Events
has_ADR	The adverse reaction caused by a specific drug.	Yes	Yes	Yes	Yes
has_occurrence	The occurrence frequency of an adverse reaction caused by a specific drug.	Yes	No	Yes	Yes
has_severity	The severity of an adverse reaction caused by a specific drug.	Yes	No	Yes	Yes
ADR_occurs in	The population in which an adverse reaction occurs.	No	No	Yes	Yes
has_mitigation method	Any measure that shortens the duration of an adverse reaction or reduces its severity.	No	No	No	No
has_dose form	The form of a dosage of a specific drug.	No	No	No	No
has_administration route	The prescribed way of administering a drug to a patient.	No	No	No	Yes
has_contraindication	A predefined condition or factor associated with a recipient that makes the use of a specific drug improper or inadvisable.	No	No	No	No

[a]ADR: adverse drug reaction.

Bioinformatic Analysis of the Identified ADRs Caused by Quinolones

The ontology of ADRs caused by quinolones consists of 7725 drug-ADR triples retrieved from 165 package inserts. After duplicate removal, 331 ADRs were identified to be caused by quinolones. Table 2 lists the major ADRs from two dimensions: the quinolone generation and active ingredient. Since topical drugs have fewer ADRs than oral and intravenous ones due to their administration route, we concentrated on the comparison of quinolone drugs that are only administered by the oral and intravenous routes. After bioinformatic analysis, there were three important points that could be summarized as follows: (1) Levofloxacin (third generation) and gatifloxacin (fourth generation) induce a significant number of ADRs (n=139, 43.57% and n=122, 38.24%, respectively), and many drugs are made from these ingredients (n=21, 17.21% and n=20, 16.39%, respectively). (2) The highest numbers of ADRs are caused by ciprofloxacin (n=221, 69.28%) and enoxacin (n=156, 48.90%), respectively. (3) There are relatively few ADRs associated with nalidixic acid (n=21, 6.56%) and pipemidic acid (n=11, 3.45%), which implies that they are comparatively safe in terms of known adverse effects. In future work, we will investigate the ADR differences related to drug dosages.

XSL•FO
RenderX

Table 2. Identified adverse drug reactions of quinolones, excluding topical drugs (N=122).

Quinolone generation and ingredient	Drugs, n (%)	Drug-ADRs[a] (n=7563), n (%)	ADRs (n=319), n (%)	Major ADRs	Typical dosage form(s)	Major routes
First generation						
Nalidixic acid	1 (0.82)	21 (0.28)	21 (6.56)	Nausea, vomiting, diarrhea, abdominal pain	Tablet	Oral
Second generation						
Pipemidic acid	3 (2.46)	33 (0.44)	11 (3.45)	Nausea, eructation, abdominal pain	Granules, capsule, tablet	Oral
Third generation						
Enoxacin	6 (4.92)	335 (4.43)	156 (48.90)	Skin rash, dizziness, abdominal pain	Tablet, capsule, injectable	Oral, intravenous
Sparfloxacin	5 (4.10)	235 (3.11)	47 (14.73)	Headache, dizziness, insomnia, anemia	Tablet, capsule, granules	Oral
Pefloxacin	5 (4.10)	150 (1.98)	35 (10.97)	Anaphylaxis, convulsion, tremor	Tablet, capsule, injectable	Oral, intravenous
Levofloxacin	21 (17.21)	1592 (21.05)	139 (43.57)	Anaphylaxis, insomnia, dizziness	Tablet, capsule, injectable	Oral, intravenous
Tolfloxacin	1 (0.82)	26 (0.34)	26 (8.15)	Fatigue, anorexia, erythematous rash	Capsule	Oral
Fleroxacin	8 (6.56)	210 (2.78)	47 (14.73)	Nausea, vomiting, headache, dizziness	Tablet, capsule, injectable	Oral, intravenous
Ofloxacin	9 (7.38)	303 (4.01)	87 (27.27)	Anaphylaxis, itching, skin rash	Tablet, capsule, injectable	Oral, intravenous
Lomefloxacin	12 (9.84)	474 (6.27)	85 (26.65)	Itching, skin rash, headache, nausea	Tablet, capsule, injectable	Oral, intravenous
Ciprofloxacin	8 (6.56)	684 (9.04)	221(69.28)	Skin rash, itching, diarrhea, hematuria	Tablet, capsule, injectable	Oral, intravenous
Rufloxacin	2 (1.64)	68 (0.90)	34 (10.66)	Skin rash, insomnia, lethargy, convulsion	Tablet, capsule	Oral
Norfloxacin	9 (7.38)	199 (2.63)	35 (10.97)	Itching, skin rash, abdominal pain	Tablet, capsule, injectable	Oral, intravenous
Fourth generation						
Gatifloxacin	20 (16.39)	2270 (30.01)	122 (38.24)	Headache, vision disorder, dysgeusia	Tablet, capsule, injectable	Oral, intravenous
Gemifloxacin	1 (0.82)	106 (1.40)	106 (33.22)	Skin rash, nausea, urticaria, diarrhea	Tablet	Oral
Balofloxacin	2 (1.64)	66 (0.87)	33 (10.34)	Itching, thirst, hypesthesia, headache	Tablet, capsule	Oral
Pazufloxacin	3 (2.46)	180 (2.38)	60 (18.81)	Skin rash, jaundice, myalgia, diarrhea	Injectable	Intravenous
Prulifloxacin	3 (2.46)	189 (2.50)	63 (19.75)	Eructation, dyspnea, hypotension	Tablet, capsule	Oral
Moxifloxacin	3 (2.46)	420 (5.55)	141 (44.20)	Fatigue, constipation, rupture of tendon	Tablet, injectable	Oral, intravenous

[a]ADR: adverse drug reaction.

The 331 ADRs could be further classified into 18 categories that we defined in the ontology (Table 3). Most organs are involved in ADRs. Skin reactions (eg, itching, skin rash) are the most common reactions and are linked to 141 drugs (85.45%), followed by nervous system reactions (eg, headache, hypertonia), which are associated with 127 agents (76.97%). Moreover, the least frequently reported adverse reactions are those of genital organs and the application site. This may be due to the difficulty in detecting genital organ diseases and the minor impact of injection site reactions.

XSL•FO
RenderX

Table 3. Categorized adverse drug reactions of quinolones (N=165).

Category of adverse drug reaction	Associated drugs, n (%)	Possible adverse drug reactions
Skin reactions	141 (85.5)	Itching, skin rash, photosensitive reaction, erythema multiforme, increased sweating
Nervous system reactions	127 (77.0)	Headache, hypertonia, convulsion, coma, paresthesia, vertigo, tremor
Immune reactions and infections	123 (74.6)	Candidiasis, anaphylaxis, angioneurotic edemas, anaphylactic shock, facial edema
Gastrointestinal reactions	122 (73.9)	Nausea, vomiting, abdominal pain, diarrhea, stomatitis, constipation, xerostomia
Generalized reactions	120 (72.7)	Fever, fatigue, syncope, chest pain, shivering, edema, oral edema, discomfort
Mental disorders	119 (72.1)	Sleeplessness, personality disorders, hallucinations, depression, agitation, anxiety
Liver and gallbladder diseases	117 (70.9)	Alanine transaminase (ALT) elevation, jaundice, alkaline phosphatase increased, liver failure, bilirubinemia
Urinary diseases	111 (67.3)	Hematuria, urinary incontinence, dysuria, crystalluria, interstitial nephritis
Musculoskeletal diseases	105 (63.6)	Arthritis, arthralgia, muscle weakness, myalgia, rupture of tendon, bone pain
Hematological diseases	103 (62.4)	Eosinophilia, leukopenia, granulocytopenia, pancytopenia, lymphadenopathy
Vascular, hemorrhagic, and coagulation diseases	102 (61.8)	Elevated international normalized ratio (INR) value, purpura, phlebitis, vasculitis, vasodilatation, flushing, thrombocytosis
Cardiovascular diseases	80 (48.5)	Prolonged QT interval, hypotension, ventricular tachycardia, palpitation, bradycardia
Respiratory system diseases	71 (43.0)	Pulmonary infiltration, bronchial spasm, asthma, laryngeal edema, dyspnea
Metabolic and nutrition diseases	63 (38.2)	Hypoglycemia, electrolytes abnormality, diabetes mellitus, hyperglycemia
Vision diseases	62 (37.6)	Xerophthalmia, eye pain, conjunctivitis, diplopia, photophobia, abnormal vision
Auditory, vestibular, and sensory diseases	62 (37.6)	Tinnitus, hypoacusis, deafness, taste disorders, parosmia, earache, ageusia
Genital organ diseases	48 (29.1)	Vaginitis, epididymitis, orchitis, dysmenorrhea, uterine hemorrhage
Application site reactions	19 (11.5)	Injection site reaction, injection site itching, infusion site reaction, injection site pain

Table 4 lists the top 10 ADRs caused by quinolones and shows that itching and skin rash are listed in nearly 80% of currently licensed quinolone package inserts in China. Although alanine transaminase (ALT) elevation and phlebitis frequently occur after the administration of 3 drugs made of pazufloxacin mesylate, other adverse reactions caused by quinolones are usually infrequent (<10%). All of the reported ADRs have a severity of Grade 1 (mid) to Grade 3 (severe but not immediately life-threatening), and stopping medication is generally recommended as the mitigation method after the detection of an adverse reaction. Additionally, quinolone drugs are administered to patients at various dosages and through different routes (Table 2). Finally, analysis of contraindication data found that the top 5 specific populations are those who are allergic to quinolones, pregnant women, teenagers, infants, and patients with central nervous system diseases.

Table 4. The 10 most commonly reported adverse drug reactions of quinolones (N=165).

Adverse drug reaction	Related drugs, n (%)	Category of adverse drug reaction
Itching	132 (80.0)	Skin diseases
Skin rash	131 (79.4)	Skin diseases
Headache	126 (76.4)	Nervous system diseases
Nausea	119 (72.1)	Gastrointestinal diseases
Abdominal pain	117 (70.9)	Gastrointestinal diseases
Vomiting	116 (70.3)	Gastrointestinal diseases
Diarrhea	114 (69.1)	Gastrointestinal diseases
Insomnia	110 (66.7)	Mental disorders
Fever	104 (63.0)	Generalized diseases
Photosensitive reaction	100 (60.6)	Skin diseases

XSL•FO
RenderX

Discussion

Major Applications of ADR Ontology

The ADR ontology proposed in this study has two major applications. The first is the semantic retrieval system that can use the ADR knowledge to integrate various external sources of information. Since the primary ADR terms and their key codes were reused from WHOART and MedDRA, it should be straightforward to integrate our ontology with these controlled vocabularies, as well as other medical terminologies (eg, Logical Observation Identifiers Names and Codes [LOINC], Systematized Nomenclature of Medicine-Clinical Terms [SNOMED CT]), for semantic knowledge retrieval by providing formally represented ADR information. Another major application is intelligent clinical decision-making support. The proposed ADR ontology provides machine-understandable knowledge, which could be used by artificial intelligence algorithms in biomedicine. To pursue clinical and therapeutic approaches to optimal disease management and rational drug use, it is useful for a physician treating a specific disorder to know all the identified adverse reactions induced by the drugs prescribed for patients with that condition. Manually reading the package inserts to find the ADR information and medication instructions is laborious and time-consuming, as the number of newly approved drugs and reported ADRs increases every year. Conversely, our ontology will aid in the development of an intelligent clinical decision-making system, which would positively affect the drug prescribing patterns of physicians and potentially have a significant socioeconomic impact.

Conclusions

We have shown that the ADR ontology can be used to formally represent and organize the ADR information and medication instructions enclosed in official drug packages. In addition, it can provide machine-understandable knowledge for bioinformatic analysis. By collecting 165 package inserts of quinolone drugs, a specialized ADR ontology was built to classify various identified ADRs caused by quinolones. Systematic analysis of the obtained ontology data improves the deep recognition of drug-specific ADRs, making it possible to intelligently guide safe drug use and benefit human health.

The proposed ADR ontology can be generalized to organize the ADR information from other channels, not just package inserts. Electronic health records, in which the ADR data are frequently reported, will be acquired for the enrichment of the ADR ontology in the near future.

Acknowledgments

This work was funded by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences, grant number 2017-I2M-3-014.

Authors' Contributions

XYL prepared the initial draft of the manuscript and finished building the ontology. XL proposed the ontology model and analyzed the experimental results. HLR was in charge of the project and administered the project. JJG sampled data and investigated the application of ontology for semantic retrieval. All authors provided contributions to the final version of the paper and approved it.

Conflicts of Interest

None declared.

References

1. Rieder M, Ferro A. Adverse drug reactions. Br J Clin Pharmacol 2015 Oct 20;80(4):613-614 [FREE Full text] [doi: 10.1111/bcp.12695] [Medline: 26388499]
2. Hacker M, Bachmann K, Messer W, editors. Pharmacology, Principles and Practice. Burlington, MA, USA: Academic Press; Nov 11, 2010:1-594.
3. Studer R, Benjamins V, Fensel D. Knowledge engineering: Principles and methods. Data & Knowledge Engineering 1998 Mar;25(1-2):161-197. [doi: 10.1016/s0169-023x(97)00056-6]
4. Bello SM, Shimoyama M, Mitraka E, Laulederkind SJF, Smith CL, Eppig JT, et al. Disease Ontology: improving and unifying disease annotations across species. Dis Model Mech 2018 Mar 12;11(3):dmm032839 [FREE Full text] [doi: 10.1242/dmm.032839] [Medline: 29590633]
5. Paul AK, Shill PC. Incorporating gene ontology into fuzzy relational clustering of microarray gene expression data. Biosystems 2018 Jan;163:1-10 [FREE Full text] [doi: 10.1016/j.biosystems.2017.09.017] [Medline: 29113811]
6. Béré WRC, Camara G, Malo S, Lo M, Ouaro S. IDOMEN: An Extension of Infectious Disease Ontology for MENingitis. Stud Health Technol Inform 2019 Aug 21;264:313-317. [doi: 10.3233/SHTI190234] [Medline: 31437936]
7. Robinson P, Mundlos S. The human phenotype ontology. Clin Genet 2010 Jun;77(6):525-534. [doi: 10.1111/j.1399-0004.2010.01436.x] [Medline: 20412080]
8. Natale DA, Arighi CN, Blake JA, Bult CJ, Christie KR, Cowart J, et al. Protein Ontology: a controlled structured network of protein entities. Nucleic Acids Res 2014 Jan 21;42(Database issue):D415-D421 [FREE Full text] [doi: 10.1093/nar/gkt1173] [Medline: 24270789]

XSL•FO
RenderX

9. Margraff F, Bertram D. Adverse drug reaction reporting by patients: an overview of fifty countries. Drug Saf 2014 Jun 19;37(6):409-419. [doi: 10.1007/s40264-014-0162-y] [Medline: 24748428]
10. van Hunsel F, Talsma A, van Puijenbroek E, de Jong-van den Berg L, van Grootheest K. The proportion of patient reports of suspected ADRs to signal detection in the Netherlands: case-control study. Pharmacoepidemiol Drug Saf 2011 Mar 28;20(3):286-291. [doi: 10.1002/pds.2092] [Medline: 21351310]
11. Blenkinsopp A, Wilkie P, Wang M, Routledge PA. Patient reporting of suspected adverse drug reactions: a review of published literature and international experience. Br J Clin Pharmacol 2007 Feb;63(2):148-156 [FREE Full text] [doi: 10.1111/j.1365-2125.2006.02746.x] [Medline: 17274788]
12. van Grootheest K, de Jong-van den Berg L. Patients' role in reporting adverse drug reactions. Expert Opinion on Drug Safety 2005 Feb 23;3(4):363-368. [doi: 10.1517/14740338.3.4.363]
13. Medawar C, Herxheimer A. A comparison of adverse drug reaction reports from professionals and users, relating to risk of dependence and suicidal behaviour with paroxetine. Int J Risk Saf Med 2003;16(1):5-19.
14. Mitchell AS, Henry DA, Sanson-Fisher R, O'Connell DL. Patients as a direct source of information on adverse drug reactions. BMJ 1988 Oct 08;297(6653):891-893 [FREE Full text] [doi: 10.1136/bmj.297.6653.891] [Medline: 3140967]
15. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. J Am Med Inform Assoc 2014 Mar 01;21(2):308-314 [FREE Full text] [doi: 10.1136/amiajnl-2013-001718] [Medline: 23907285]
16. Warrer P, Hansen E, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. Br J Clin Pharmacol 2012 May;73(5):674-684 [FREE Full text] [doi: 10.1111/j.1365-2125.2011.04153.x] [Medline: 22122057]
17. Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, EU-ADR group. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol Drug Saf 2009 Dec;18(12):1176-1184. [doi: 10.1002/pds.1836] [Medline: 19757412]
18. Chen ES, Hripcsak G, Xu H, Markatou M, Friedman C. Automated Acquisition of Disease-Drug Knowledge from Biomedical and Clinical Documents: An Initial Study. Journal of the American Medical Informatics Association 2008 Jan 01;15(1):87-98. [doi: 10.1197/jamia.m2401]
19. Wang C, Lin P, Cheng C, Tai S, Kao Yang Y, Chiang J. Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model. J Med Internet Res 2019 Feb 06;21(2):e11016 [FREE Full text] [doi: 10.2196/11016] [Medline: 30724742]
20. Finkelstein J, Chen Q, Adams H, Friedman C. Automated Summarization of Publications Associated with Adverse Drug Reactions from PubMed. AMIA Jt Summits Transl Sci Proc 2016;2016:68-77 [FREE Full text] [Medline: 27570654]
21. Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, et al. Early Detection of Adverse Drug Reactions in Social Health Networks: A Natural Language Processing Pipeline for Signal Detection. JMIR Public Health Surveill 2019 Jun 03;5(2):e11264 [FREE Full text] [doi: 10.2196/11264] [Medline: 31162134]
22. Zhao M, Yang CC. Drug Repositioning to Accelerate Drug Development Using Social Media Data: Computational Study on Parkinson Disease. J Med Internet Res 2018 Oct 11;20(10):e271 [FREE Full text] [doi: 10.2196/jmir.9646] [Medline: 30309833]
23. Park SH, Hong SH. Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy From Online Patient Medication Reviews: Text Mining of Social Network Data. J Med Internet Res 2018 Oct 24;20(10):e11085 [FREE Full text] [doi: 10.2196/11085] [Medline: 30355555]
24. Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M. Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach. JMIR Public Health Surveill 2018 May 09;4(2):e51 [FREE Full text] [doi: 10.2196/publichealth.8214] [Medline: 29743155]
25. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
26. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
27. Cai MC, Xu Q, Pan YJ, Pan W, Ji N, Li YB, et al. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. Nucleic Acids Res 2015 Jan;43(Database issue):D907-D913 [FREE Full text] [doi: 10.1093/nar/gku1066] [Medline: 25361966]
28. Henegar C, Bousquet C, Lillo-Le Louët A, Degoulet P, Jaulent M. Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. Comput Biol Med 2006 Jul;36(7-8):748-767. [doi: 10.1016/j.compbiomed.2005.04.009] [Medline: 16185681]
29. He Y, Sarntivijai S, Lin Y, Xiang Z, Guo A, Zhang S, et al. OAE: The Ontology of Adverse Events. J Biomed Semantics 2014;5(1):29 [FREE Full text] [doi: 10.1186/2041-1480-5-29] [Medline: 25093068]
30. Marcos E, Zhao B, He Y. The Ontology of Vaccine Adverse Events (OVAE) and its usage in representing and analyzing adverse events associated with US-licensed human vaccines. J Biomed Semantics 2013 Nov 26;4(1):40 [FREE Full text] [doi: 10.1186/2041-1480-4-40] [Medline: 24279920]

XSL•FO
RenderX

31. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies 1995 Nov;43(5-6):907-928. [doi: 10.1006/ijhc.1995.1081]
32. Dürst MJ. Internationalized Resource Identifiers: From Specification to Testing. 2001 Sep 12 Presented at: 19th International Unicode Conference; September 12, 2001; San Jose, CA.

Abbreviations

ADE: adverse drug event
ADR: adverse drug reaction
CIMOS: Council for International Organization of Medical Science
CTCAE: Common Terminology Criteria for Adverse Events
HLT: high level terms
IRI: Internationalized Resource Identifier
IT: included terms
LOINC: Logical Observation Identifiers Names and Codes
MedDRA: Medical Dictionary for Regulatory Activities
NER: named entity recognition
NLP: natural language processing
OWL: Ontology Web Language
PT: preferred terms
RE: relation extraction
SNOMED CT: Systematized Nomenclature of Medicine-Clinical Terms
SOC: system/organ classes
WHO: World Health Organization
WHOART: WHO Adverse Reactions Terminology

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by K Koshechkin; comments to author 09.06.20; revised version received 11.06.20; accepted 14.06.20; published 20.07.20.

Please cite as:
Li X, Lin X, Ren H, Guo J
Ontological Organization and Bioinformatic Analysis of Adverse Drug Reactions From Package Inserts: Development and Usability Study
J Med Internet Res 2020;22(7):e20443
URL: https://www.jmir.org/2020/7/e20443
doi:10.2196/20443
PMID:32706718

©Xiaoying Li, Xin Lin, Huiling Ren, Jinjing Guo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

QardioArm Blood Pressure Monitoring in a Population With Type 2 Diabetes: Validation Study

Victoria Mazoteras-Pardo[1], RN, MSc, PhD; Ricardo Becerro-De-Bengoa-Vallejo[1], RN, MLiS, PhD; Marta Elena Losa-Iglesias[2], RN, MSc, PhD; Eva María Martínez-Jiménez[3], MSc, PhD; César Calvo-Lobo[1], MSc, PhD; Carlos Romero-Morales[4], MSc, PhD; Daniel López-López[5], MSc, PhD; Patricia Palomo-López[6], MSc, PhD

[1]Facultad de Enfermería, Fisioterapia y Podología, Universidad Complutense de Madrid, Madrid, Spain

[2]Faculty of Health Sciences, Universidad Rey Juan Carlos, Alcorcón, Spain

[3]Departamento de Enfermería, Facultad de Fisioterapia y Enfermería, Universidad de Castilla la Mancha, Toledo, Spain

[4]Faculty of Sport Sciences, Universidad Europea de Madrid, Villaviciosa de Odón, Spain

[5]Research, Health and Podiatry Group, Department of Health Sciences, Faculty of Nursing and Podiatry, Universidade da Coruña, Ferrol, Spain

[6]University Center of Plasencia, Universidad de Extremadura, Plasencia, Spain

Corresponding Author:
Marta Elena Losa-Iglesias, RN, MSc, PhD
Faculty of Health Sciences
Universidad Rey Juan Carlos
Avenida de Atenas S/N
Alcorcón
Spain
Phone: 34 91 488 8508
Email: marta.losa@urjc.es

Abstract

Background: Home blood pressure monitoring has many benefits, even more so, in populations prone to high blood pressure, such as persons with diabetes.

Objective: The purpose of this research was to validate the QardioArm mobile device in a sample of individuals with noninsulin-dependent type 2 diabetes in accordance with the guidelines of the second International Protocol of the European Society of Hypertension.

Methods: The sample consisted of 33 patients with type 2 diabetes. To evaluate the validity of QardioArm by comparing its data with that obtained with a digital sphygmomanometer (Omron M3 Intellisense), two nurses collected diastolic blood pressure, systolic blood pressure, and heart rate with both devices.

Results: The analysis indicated that the test device QardioArm met all the validation requirements using a sample population with type 2 diabetes.

Conclusions: This paper reports the first validation of QardioArm in a population of individuals with noninsulin-dependent type 2 diabetes. QardioArm for home monitoring of blood pressure and heart rate met the requirements of the second International Protocol of the European Society of Hypertension.

(J Med Internet Res 2020;22(7):e19781) doi:10.2196/19781

KEYWORDS

blood pressure; hypertension; type 2 diabetes; mobile applications; software validation

Introduction

Ultimately, devices that automatically measure blood pressure have gradually evolved to replace the standard mercury sphygmomanometer, as the risk of mercury toxicity outweighed any potential benefits of its use. Today there are a wide variety of devices capable of measuring blood pressure both in and outside of clinical environments [1-7].

Among the advantages of home blood pressure monitoring are that it helps to detect white coat and masked hypertension; it is highly available at a low price; blood pressure measurements at home are taken in a more natural, relaxed, and domestic

XSL•FO
RenderX

environment than when taken in the doctor's office; patient convenience in blood pressure measurement ensures more commitment; repeated measurements can be easily taken, and it can be used over longer periods to assess the daily variability of blood pressure [8].

But logically, it is a necessary requirement that these devices have been validated by standardized protocols, such as the European Society of Hypertension-International Protocol 2 (ESH-IP2) [9], the British Hypertension Society protocol [10], or the protocol of the Association for the Advancement of Medical Instrumentation [11].

If this validation is essential in the general population, it is even more important in at-risk populations, such as patients with diabetes, where reliability and accurate blood pressure measurement are essential to avoid the disease's progression. Type 2 diabetes is a prevalent disease and represents a very serious social health problem; it has been seen to occur to a greater extent in older adults [12,13]. The main complication of type 2 diabetes is atherosclerotic disease [14,15].

This atherosclerotic disease is the main cause of morbidity and mortality in people with diabetes, and the one that contributes the most to the direct and indirect costs of diabetes [16,17]. In addition, atherosclerosis reduces arterial elasticity, and therefore, causes arterial stiffness. This arterial wall stiffening produces a high risk of mortality [14-18]. Furthermore, the occurrence of stiffened arteries increases with diabetes [19] and hypertension; therefore, it is common that patients with diabetes are also hypertensive [18,20,21]. In fact, it is estimated that more than two-thirds of patients with type 2 diabetes are also hypertensive [22]. It has also been shown that arterial stiffness is more critical in individuals with both hypertension and type 2 diabetes than in those without hypertension [23,24]. This is a major health problem.

For all these reasons, correctly and validly measuring blood pressure in a person with diabetes is crucial. On one hand, blood pressure control has been shown to decrease the risk of diabetes-related complications such as microvascular and microvascular pathology [16,22-25]. On the other hand, the increase in arterial stiffness can affect the accuracy of automatic blood pressure measurements that are essential for diagnosis and administration [26,27]. And, more importantly, measuring blood pressure accurately facilitates the establishment of a threshold above which antihypertensive treatment may be recommended, especially in patients with type 2 diabetes who are very likely to develop complications such as peripheral arterial disease, stroke, heart attack, sudden death from heart failure, or renal pathologies, if not treated [28,29].

In that regard, it has been proven that blood pressure control and the prevention of morbidity and related mortality could be improved in persons with diabetes [3,8,16]. However, knowledge, treatment, and control of hypertension are persistently low worldwide [30,31].

One of the great challenges is to avoid therapeutic inertia, as this would result in an unacceptable burden in terms of human lives, sequelae, and socioeconomic costs [16,32].

Therefore, we consider it highly relevant to evaluate the validity of automatic blood pressure measurements in persons with diabetes because of the high prevalence of hypertension in this specific population and its significant morbidity and mortality rate.

The main goal of this study was to test the validity of the measurements of the QardioArm blood pressure monitoring device in individuals with noninsulin-dependent type 2 diabetes persons, in accordance with ESH-IP2 [9]. We hypothesized that the QardioArm for home blood pressure monitoring measurements of blood pressure and heart rate in individuals with noninsulin-dependent type 2 diabetes would be equivalent to those from a gold standard device within the requirements of ESH-IP2.

Methods

Study Design

We performed an observational concordance study to validate the QardioArm device for measurement of heart rate, diastolic blood pressure, and systolic blood pressure, in individuals with noninsulin-dependent type 2 diabetes according to ESH-IP2 [9] and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) criteria [33]. The study was conducted between September 2019 to January 2020.

Ethics

This research study received ethical approval (number 173/2019). This study respected the Helsinki Declaration [34]. All participants signed an informed consent form prior to being included in the study.

Devices

Omron M3 Intellisense

The Omron M3 Intellisense (Omron Healthcare Co Ltd) was used as the gold standard instead of a mercury sphygmomanometer, in this study, because it has been validated in the general population according to ESH-IP2 [35].

The Omron M3 Intellisense device records heart rate in the range of 40 to 180 bpm and brachial blood pressure in the range of 0 to 299 mmHg using the oscillometric method; systolic blood pressure, diastolic blood pressure, and heart rate are shown on screen. The inflation system does not require pressure presetting or reinflation because of its technology, and deflation is automatically released by a pressure valve. It weighs approximately 340 g without batteries. The standard cuff fits arm circumferences ranging from 22 to 32 cm; it is also available in a large cuff to fit arm circumferences from 32 to 42 cm.

QardioArm

The QardioArm (Atten Electronic Co) is an automatic home blood pressure measurement monitoring device. QardioArm is a blood pressure measurement system intended to assess heart rate, diastolic blood pressure, and systolic blood pressure in an adult population, for pulse rates in the range of 40 to 200 bpm and blood pressure in the range of 40 to 250 mmHg.

XSL•FO
RenderX

This device utilizes an inflatable cuff that is wrapped around the upper arm, with circumferences ranging from 22 to 37 cm. The weight of unit is 310 g with batteries and its dimensions, when closed, are 140 mm×68 mm×38 mm. There is a freely downloadable app on the company's website, Google Play, or the Apple App Store. A smart device with Android 4.4 (KitKat) or later, or Bluetooth 4.0 and iOS 7.0 or later, is required, being compatible with iPhone, iPod, iPad, Apple Watch, tablets, and smartphones. QardioArm also provides graphics to facilitate visual data interpretation on screen. This app may be configured by warnings and reminders, also measurements and progress may be shared in real time with clinicians and other users.

Patients and Recruitment

All patients were recruited from Fresenius Medical facilities in Plasencia-Extremadura, Spain. The inclusion criteria were age greater than 25 years, gender (at least 10 men and 10 women), and recruitment blood pressure requirements according to the guidelines [9]: 33 participants diagnosed with noninsulin-dependent type 2 diabetes were included.

The exclusion criteria were having an arrhythmia or using insulin as treatment at the time of the study.

Study Protocol

Two nurses with experience in blood pressure measurement performed all assessments. The assessment room was a comfortable temperature without noise, distractions, or any factors that could have influenced the measurements [9].

Birth date, gender, weight, and height of each participant were recorded, and the arm circumference was measured in order to ensure an adequate cuff size. Body mass index was calculated using the Quetelet index.

The same assessment room was used for all participants. Each participant was seated in the assessment room, and the measurements were assessed after a period of 10 to 15 minutes of rest. Measurements by both units were taken on the left arm supported at the heart level, and a total of 9 consecutive measurements (systolic blood pressure, diastolic blood pressure and heart rate) were carried out alternating the Omron M3 Intellisense and the QardioArm in the following order: blood pressure A—entry blood pressure using the Omron M3 as standard device; blood pressure B—entry blood pressure using the Omron M3 as test device; blood pressure 1 by Omron M3; blood pressure 2 by QardioArm; blood pressure 3 by Omron M3; blood pressure 4 by QardioArm; blood pressure 5 by Omron M3; blood pressure 6 by QardioArm; and blood pressure 7 by Omron M3.

During measurement, the patient remained quiet and calm, sitting placing the back straight maintaining the feet over the floor in parallel position without crossing their legs as well as resting the arm over a flat surface, with the hand palm upwards and the elbow in a slightly flexed position in order to place his fist at the height of its heart. The interval between one measurement and the next was 30 or 60 seconds [9].

Data Analysis

Results are described as mean and standard deviation with range. Sociodemographic variables were examined using the Kolmogorov–Smirnov test to assess normality, and data were considered normally distribution if $P>.05$. Two-tailed independent *t* tests or Mann–Whitney *U* tests were used for parametric or nonparametric variables, respectively.

Device accuracy, following the ESH-IP2 [9], is based on a comparison between the device and reference measurements. Part 1 of the protocol refers to the number of differences within the required ranges for individual measurements (99 measurements), and part 2 of the protocol refers to the number of differences within the required ranges for each individual (n=33).

For each patient, systolic blood pressure measurements obtained from the QardioArm device (blood pressure measurements 2, 4, and 6) were compared with the systolic blood pressure measurements from the Omron device (blood pressure measurements 1, 3, and 5, respectively, or with blood pressure measurements 3, 5, and 7, respectively); the comparisons which were favorable to the device were utilized. The same procedure was followed with diastolic blood pressure and heart rate measurements.

Differences were separately classified for diastolic blood pressure and systolic blood pressure, depending on whether the difference was within 5, 10, or 15 mmHg [9] as well as for heart rate depending on whether the difference was within 3, 5, or 8 bpm.

Results were analyzed and detailed according to ESH-IP2 requirements in order to determine if the device passed the validation protocol. Accuracy was determined by the number differences in these ranges for both individual measurements (part 1) and individuals (part 2). To pass, a device must meet the minimum pass requirements.

Furthermore, Bland-Altman plots were utilized to quantify agreement between such systolic blood pressure, diastolic blood pressure, and heart rate by constructing limits of agreement (a graphical method in which the differences between both devices are used to compare two measurements of the same variable).

All analyses were carried out with SPSS statistical software (version 19.0; IBM Corp). In all analyses, the threshold for statistical significance was $P=.05$ with a 95% confidence interval.

Results

Participants

All the sociodemographic variables showed a normal distribution ($P>.05$, Table 1). A sample of 37 participants diagnosed with noninsulin-dependent type 2 diabetes were recruited; 4 were excluded due to device failure (n=2), arrhythmias (n=1), and cuff size unavailability (n=1), therefore, there were 33 participants (17 men and 16 women) who met the ESH-IP2 inclusion criteria. Table 1 shows participant demographic characteristics.

XSL•FO
RenderX

Table 1. Demographic characteristics and descriptive data of the participants.

Participant characteristics (N=33)	Value, mean (SD)	Value, range[a]	*P* value
Age (years)	65.9 (9.8)	39.0-85.0	.73
Men (n=17)	66.4 (12.3)	39.0-85.0	
Women (n=16)	72.6 (6.7)	65.0-84.0	
Weight (kg)	74.8 (12.6)	48.0-101.0	.42
Men (n=17)	76.5 (12.5)	48.0-100.0	
Women (n=16)	72.9 (12.9)	57.0-101.0	
Height (cm)	164.9 (7.3)	147.0-180.0	.17
Men (n=17)	166.6 (5.3)	160.0-180.0	
Women (n=16)	163.1 (8.8)	147.0-175.0	
BMI (kg/m^2)	27.6 (5.2)	18.1-43.7	.90
Men (n=17)	27.5 (3.7)	18.1-34.1	
Women (n=16)	27.7 (6.5)	20.3-43.7	
Arm circumference (mm)	286.4 (32.7)	220.0-360.0	.19
Men (n=17)	279.1 (26.8)	220.0-320.0	
Women (n=16)	294.1 (37.3)	243.0-360.0	

[a]minimum to maximum

Blood Pressure Outcome Measurements

Parts 1 and 2 of the validation according to the ESH-IP2 for the QardioArm blood pressure are presented in Table 2; the number of differences between the device and reference for systolic blood pressure and diastolic blood pressure classified within 5, 10, or 15 mmHg are detailed.

Mean differences between the QardioArm and Omron M3 were 3.37 (SD 3.19) mmHg for diastolic blood pressure and 3.15 (SD 4.67) mmHg for systolic blood pressure.

A total of 87/99 differences (89%) for systolic blood pressure and a total of 88/99 differences (89%) for diastolic blood pressure showed an absolute difference within 5 mmHg (ESH-IP2 criteria: at least 65 for diastolic blood pressure and at least 73 for systolic blood pressure). Furthermore, a total of 95/99 comparisons (96%) for systolic blood pressure and a total of 96/99 (97%) for diastolic blood pressure showed an absolute difference within 10 mmHg (ESH-IP2 criteria: at least 81 for diastolic blood pressure and at least 87 for systolic blood pressure). A total of 98/99 differences (99%) for systolic blood pressure and a total of 99/99 differences (100%) for diastolic blood pressure exhibited an absolute difference within 15 mmHg (ESH-IP2 criteria: at least 93 for diastolic blood pressure and at least 96 for systolic blood pressure). Part 1 of the validation was successfully completed since 2 or more of the 3 absolute difference ranges (within 5, 10, or 15 mmHg) for systolic blood pressure met the minimum requirements, and 3 out of 3 absolute difference ranges for diastolic blood pressure met the minimum requirements.

A total of 32/33 patients (97%) for systolic blood pressure and a total of 29/33 patients (88%) for diastolic blood pressure showed a minimum of 2 out of 3 differences within 5 mmHg (ESH-IP2 criteria: at least 24 out of 33 patients for systolic blood pressure and at least 24 out of 33 patients for diastolic blood pressure). Nevertheless, 1/33 patients (3%) had 3 differences outside 5 mmHg for systolic blood pressure, though 0/33 patients (0%) had 3 differences outside 5 mmHg for diastolic blood pressure (ESH-IP2 criteria: a maximum of 3 patients for diastolic blood pressure and systolic blood pressure according to the ESH-IP2 criteria. Part 2 of the device validation was successfully completed; therefore, part 3 of the QardioArm device validation was passed, since parts 1 and 2 were both passed for diastolic blood pressure and systolic blood pressure.

XSL•FO
RenderX

Table 2. Number of measurement differences in each range.

Measurement type and range	ESH-IP2 requirement, n		Achieved (N=99), n	Difference (mmHg), mean (SD)
	For 2 of 3 ranges	For all 3 ranges		
Part 1				
Systolic blood pressure				3.37 (3.19)
≤5 mmHg	73	65	87	
≤10 mmHg	87	81	95	
≤15 mmHg	96	93	98	
Diastolic blood pressure				3.15 (SD 4.67)
≤5 mmHg	73	65	88	
≤10 mmHg	87	81	96	
≤15 mmHg	96	93	99	
Heart rate				1.65 (2.91)
≤3 bpm	73	65	91	
≤5 bpm	87	81	96	
≤8 bpm	96	93	99	

Heart Rate Outcome Measurements

Parts 1 and 2 of the validation according to the ESH-IP2 for heart rate are also shown in Table 2 and Table 3; the number of differences between QardioArm and the standard device, Omron M3, within 3, 5, and 8 bpm are detailed. The mean difference between the tested device and standard was 1.65 (SD 2.91) bpm.

A total of 91/99 differences (92%) showed an absolute difference within 3 bpm, a total of 96/99 comparisons (97%) showed an absolute difference within 5 bpm, and a total of 99/99 differences (100%) showed an absolute difference within 8 bpm. Thus, part 1 of the device validation was successfully completed for the heart rate.

Table 3. Within-participant measurement differences.

Measurement type and range	2 of 3 measurements		0 of 3 measurements	
	ESH-IP2 requirement, n	Achieved (N=33), n	ESH-IP2 requirement, n	Achieved (N=33), n
Part 2				
Systolic blood pressure				
≤5 mmHg	≥24	32	≤3	1
Diastolic blood pressure				
≤5 mmHg	≥24	29	≤3	0
Heart rate				
≤3 bpm	≥24	31	≤3	1

A total of 31/33 participants (94%) had a minimum of 2 of 3 comparisons within 3 bpm difference for heart rate. Nevertheless, a total of 1/33 participants (3%) had 3 differences greater than 3 bpm; therefore, part 2 of the device validation was successfully completed for heart rate, and consequently, part 3 of the QardioArm device validation was passed.

Indeed, the QardioArm device met the validation criteria of the ESH-IP2 for heart rate, systolic blood pressure, and diastolic blood pressure for this sample of 33 individuals with noninsulin-dependent type 2 diabetes.

The Bland-Altman plots show the differences between QardioArm and Omron M3 measurements for systolic blood pressure (Figure 1), diastolic blood pressure (Figure 2), and heart rate (Figure 3).

XSL•FO
RenderX

Figure 1. Bland-Altman plot showing differences between QardioArm and Omron M3 measurements for systolic blood pressure.

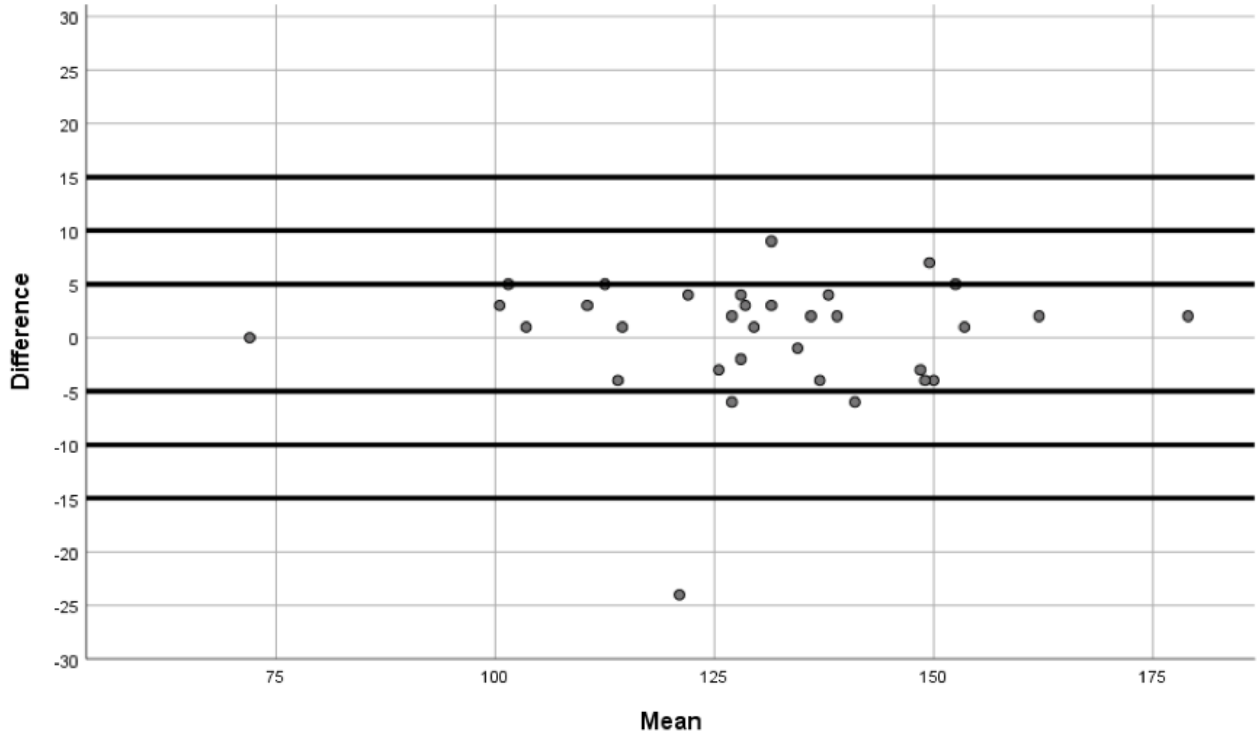
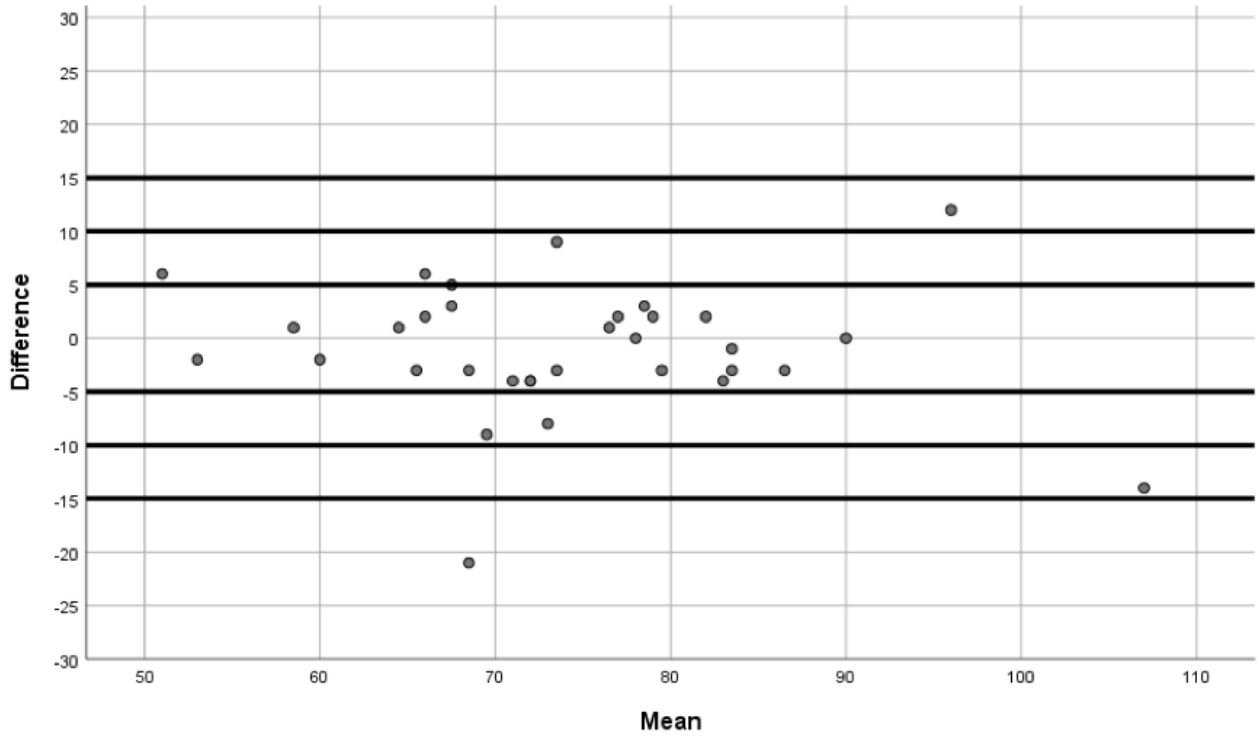
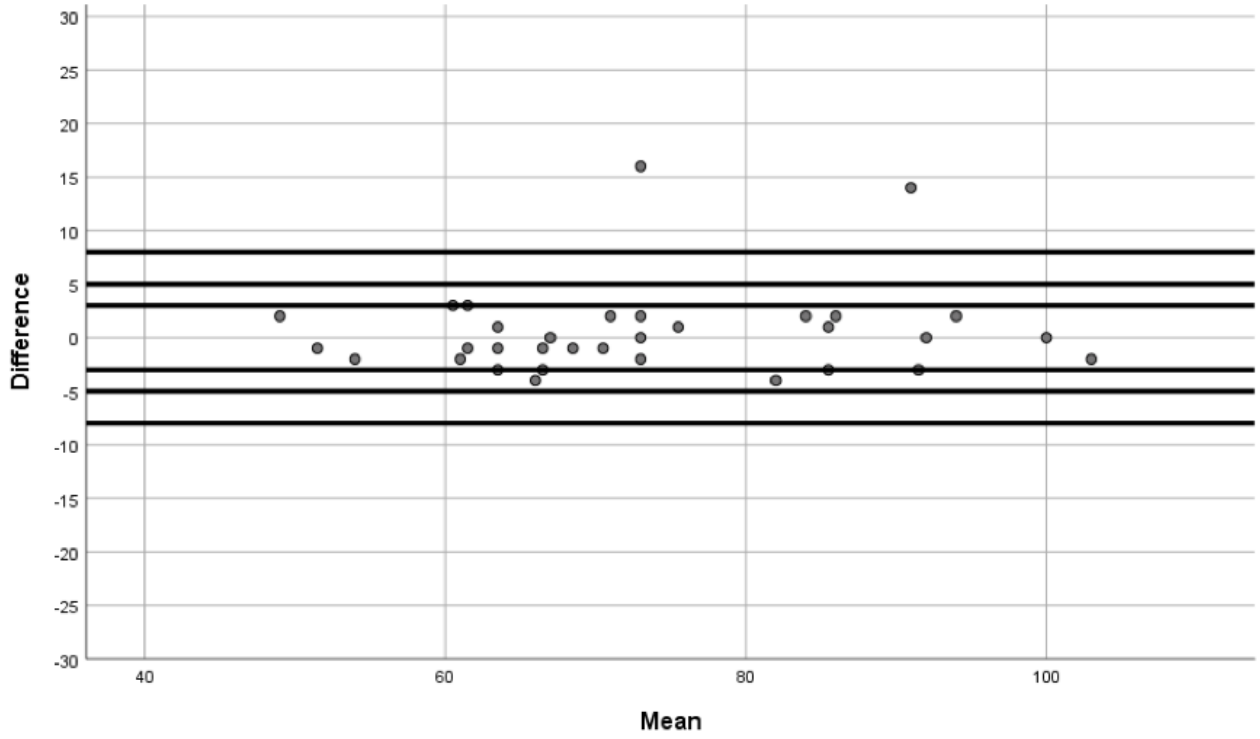


Figure 2. Bland-Altman plot showing differences between QardioArm and Omron M3 measurements for diastolic blood pressure.



XSL•FO
RenderX

Figure 3. Bland-Altman plot showing differences between QardioArm and Omron M3 measurements for heart rate.



Discussion

As far as we know, this is the first study to validate an upper arm oscillometric wireless device connected to an app to measure heart rate and blood pressure in a population with noninsulin-dependent type 2 diabetes. QardioArm was validated in 2017 [36,37] in the general population. Furthermore, QardioArm has also been validated in a population of obese individuals in 2018 [38], and recently, in a population of individuals with chronic kidney disease [39].

The results of this study have shown that QardioArm device passed both parts of the validation [9] in patients diagnosed with noninsulin-dependent type 2 diabetes as hypothesized. This validation has been carried out following the ESH-IP2 guidelines, although a validated noninvasive oscillometric upper arm device was used as a reference instead a mercury sphygmomanometer [3-6]. Our findings, however, should not be extrapolated to other populations with other diseases.

We have already seen the enormous advantages of home blood pressure monitoring, especially in specific populations prone to hypertension, such as persons with diabetes. Chronic diseases such as diabetes are one of the most common pathologies and have significant physical, psychological, social, and economic impacts [3,8,16,17,19-25,28,29,32]. mHealth and apps that assess blood pressure are progressively gaining a fundamental role in the management of hypertension with the potential to improve the quality of managed care by offering additional advantages, especially when it comes to blood pressure self-measurement. Advantages could include assisting in lowering blood pressure [3,40,41]; enabling individuals to have 24-hour access to detailed and personalized blood pressure information [40,42]; improving adherence to treatment [3,43]; providing information to facilitate diagnostic and therapeutic decisions [40,41,43,44]; improving patient knowledge of blood pressure, lifestyle, and risk factors that may accompany hypertension [40,42,43]; more effectively preventing cardiovascular complications [43]; or health cost savings by avoiding unnecessary hypertensive treatment and reducing the number of visits to the clinic [3,42].

Nevertheless, we found very few studies analyzing the validity of automatic blood pressure monitors in a population with noninsulin-dependent type 2 diabetes, [37,45-48], and furthermore, most of these validations did not follow specific protocols such as ESH-IP2 [9-11].

Only one study by Chahine et al [37] was found that followed the same recommendations as those of the ESH-IP2 [9]. They validated Omron M6 IT Comfort using a mercury sphygmomanometer as the standard. If we compare our results with their results [37], the number of differences for each category (5, 10, and 15 mmHg) for the systolic blood pressure and the diastolic blood pressure were similar in the two validations, and parts 1 and 2 were passed in both. QardioArm achieved better results in part 1 of the protocol, especially for diastolic blood pressure with a number of higher differences in the 3 categories. Within part 2, QardioArm and Omron M6 differed in 3 individuals for the first criterion (2 of 3 comparisons within 5 mmHg difference) for both systolic and diastolic pressure, in favor of QardioArm. The results of the second criterion (3 differences outside 5 mmHg) were almost identical for both devices. On the other hand, the data obtained for heart rate cannot be compared because Chahine and company [37] did not assess this variable.

Another important fact is that measurements of people with type 2 diabetes involve stiffer arteries; however, in our study

XSL•FO
RenderX

this was not assessed, since it was not the main aim of this research.

Finally, for future studies, consecutive sampling bias should be considered; a simple random sampling process might be more appropriate.

Moreover, we consider that further validations are needed for devices that measure blood pressure in patients with noninsulin-dependent type 2 diabetes. In fact, a future line of research could be to specifically investigate the validation of this QardioArm device in patients with type 2 diabetes who have microarteriopathy as diabetic retinopathy, patients with type 1 diabetes as well as other populations with specific diseases such as pregnant women.

QardioArm for home blood pressure monitoring showed validated measures of blood pressure and heart rate in individuals with noninsulin-dependent type 2 diabetes meeting the requirements of ESH-IP2. The findings of the study may be relevant since it is the first validation showing that a device linked to an app to measure blood pressure and heart rate met the requirements of the ESH-IP2 in noninsulin-dependent type 2 diabetes population.

Conflicts of Interest

None declared.

References

1. O Brien E, Asmar R, Beilin L, Imai Y, Mancia G, Mengden T, et al. Practice guidelines of the European Society of Hypertension for clinic, ambulatory and self blood pressure measurement. Journal of Hypertension 2005;23(4):697-701. [doi: 10.1097/01.hjh.0000163132.84890.c4]
2. Stergiou GS, Parati G, Asmar R, O'Brien E. Requirements for professional office blood pressure monitors. Journal of Hypertension 2012;30(3):537-542. [doi: 10.1097/hjh.0b013e32834fcfa5]
3. Mancia G, Fagard R, Narkiewicz K, Redón J, Zanchetti A, Böhm M, Task Force Members. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). J Hypertens 2013 Jul;31(7):1281-1357. [doi: 10.1097/01.hjh.0000431740.32696.cc] [Medline: 23817082]
4. Asayama K, Ohkubo T, Hoshide S, Kario K, Ohya Y, Rakugi H. From mercury sphygmomanometer to electric device on blood pressure measurement: correspondence of Minamata Convention on Mercury. Hypertens Res Internet 2016;39(4):179-182 [FREE Full text] [doi: 10.1038/hr.2015.158]
5. Duncombe SL, Voss C, Harris KC. Oscillometric and auscultatory blood pressure measurement methods in children. Journal of Hypertension 2017;35(2):213-224. [doi: 10.1097/hjh.0000000000001178]
6. Shahbabu B. Which is more accurate in measuring the blood pressure? a digital or an aneroid sphygmomanometer. JCDR 2016;10(3):LC11-LC14. [doi: 10.7860/jcdr/2016/14351.7458]
7. Cohen JB, Padwal RS, Gutkin M, Green BB, Bloch MJ, Germino FW, et al. History and justification of a national blood pressure measurement validated device listing. Hypertension 2019 Feb;73(2):258-264. [doi: 10.1161/hypertensionaha.118.11990]
8. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 Practice Guidelines for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. Journal of Hypertension 2018;36(12):2284-2309. [doi: 10.1097/hjh.0000000000001961]
9. &NA;. European Society of Hypertension International Protocol revision 2010 forthe validation of blood pressure measuring devices in adults: Erratum. Blood Pressure Monitoring 2010;15(3):171-172. [doi: 10.1097/mbp.0b013e32833b7c0c]
10. O?Brien E, Petrie J, Littler W. The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. J Hypertens 1993;11(Suppl 2):S43-s46.
11. Association FTAOMI. 2013. ANSI/AAMI/ISO 1 2013:1137.
12. Quinn L. Type 2 diabetespidemiology, pathophysiology, and diagnosis. Nurs Clin North Am. ? 2001;36:92.
13. Wilkinson E, Waqar M, Sinclair A, Randhawa G. Meeting the challenge of diabetes in ageing and diverse populations: a review of the literature from the UK. Journal of Diabetes Research 2016;2016:1-15. [doi: 10.1155/2016/8030627]
14. Kochkina MS, Zateĭshchikov DA, Sidorenko BA. [Measurement of arterial stiffness and its clinical value]. Kardiologiia 2005;45(1):63-71. [Medline: 15699941]
15. Cardoso C, Salles G. Aortic stiffness as a surrogate endpoint to micro- and macrovascular complications in patients with type 2 diabetes. Int J Mol Sci 2016 Dec 06;17(12):2044. [doi: 10.3390/ijms17122044]
16. Standards OMCID2. ADA Clinical Practics Recommendations. Diabetes Care 2018;41:S1-153.
17. Arrieta F, Rubio-Terrés C, Rubio-Rodríguez D, Magaña A, Piñera M, Iglesias P, et al. Estimation of the economic and health impact of complications of type 2 diabetes mellitus in the autonomous community of Madrid (Spain). Endocrinología y Nutrición (English Edition) 2014 Apr;61(4):193-201. [doi: 10.1016/j.endoen.2014.03.009]
18. Stehouwer CDA, Henry RMA, Ferreira I. Arterial stiffness in diabetes and the metabolic syndrome: a pathway to cardiovascular disease. Diabetologia 2008 Feb 1;51(4):527-539. [doi: 10.1007/s00125-007-0918-3]

XSL•FO
RenderX

19. Lehmann E, Gosling R, Sonksen P. Arterial wall compliance in diabetes. Diabet Med. ? 1992;9:19. [doi: 10.1111/j.1464-5491.1992.tb01746.x]
20. Lim HS, Lip GYH. Arterial stiffness in diabetes and hypertension. J Hum Hypertens 2004 Feb 26;18(7):467-468. [doi: 10.1038/sj.jhh.1001693]
21. Bruno RM, Penno G, Daniele G, Pucci L, Lucchesi D, Stea F, et al. Type 2 diabetes mellitus worsens arterial stiffness in hypertensive patients through endothelial dysfunction. Diabetologia 2012 Mar 13;55(6):1847-1855. [doi: 10.1007/s00125-012-2517-1]
22. Ferrannini E, Cushman WC. Diabetes and hypertension: the bad companions. The Lancet 2012 Aug;380(9841):601-610. [doi: 10.1016/s0140-6736(12)60987-8]
23. Salanitro AH, Roumie CL. Blood pressure management in patients with diabetes. Clinical Diabetes 2010 Jul 22;28(3):107-114. [doi: 10.2337/diaclin.28.3.107]
24. Chen Y, Huang Y, Li X, Xu M, Bi Y, Zhang Y, et al. Association of arterial stiffness with HbA1c in 1,000 type 2 diabetic patients with or without hypertension. Endocr 2009 Jul 14;36(2):262-267. [doi: 10.1007/s12020-009-9221-z]
25. Liang X, Zhong H, Xiao L. The effect of community hypertension management on blood pressure control and its determinants in southwest China. Int Health 2020 May 1:2020. [doi: 10.1093/inthealth/ihaa002]
26. van Popele NM, Bos WJW, de Beer NAM, van der Kuip DAM, Hofman A, Grobbee DE, et al. Arterial stiffness as underlying mechanism of disagreement between an oscillometric blood pressure monitor and a sphygmomanometer. Hypertension 2000 Oct;36(4):484-488. [doi: 10.1161/01.hyp.36.4.484]
27. Hamilton P, Lockhart C, Quinn C, McVeigh G. Arterial stiffness: clinical relevance, measurement and treatment. Clin Sci 2007;113(4):157-170. [doi: 10.1042/cs20070080]
28. Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical update: cardiovascular disease in diabetes mellitus. Circulation 2016 Jun 14;133(24):2459-2502. [doi: 10.1161/circulationaha.116.022194]
29. Paneni F, Lüscher TF. Cardiovascular protection in the treatment of type 2 diabetes: a review of clinical trial results across drug classes. The American Journal of Medicine 2017 Jun;130(6):S18-S29. [doi: 10.1016/j.amjmed.2017.04.008]
30. Natarajan N, Putnam W, Van Aarsen K, Beverley Lawson K, Burge F. Adherence to antihypertensive medications among family practice patients with diabetes mellitus and hypertension. Can Fam Physician 2013 Feb;59(2):e93-e100 [FREE Full text] [Medline: 23418264]
31. Abughosh SM, Wang X, Serna O, Henges C, Masilamani S, James Essien E, et al. A pharmacist telephone intervention to identify adherence barriers and improve adherence among nonadherent patients with comorbid hypertension and diabetes in a medicare advantage plan. JMCP 2016 Jan;22(1):63-73. [doi: 10.18553/jmcp.2016.22.1.63]
32. González-Gómez S, Meléndez-Gomez MA, López-Jaramillo P. Fixed-dose combination therapy to improve hypertension treatment and control in Latin America. Archivos de Cardiología de México 2018 Apr;88(2):129-135. [doi: 10.1016/j.acmx.2017.06.001]
33. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. International Journal of Surgery 2014 Dec;12(12):1500-1524. [doi: 10.1016/j.ijsu.2014.07.014]
34. Asociación MM. 2013 Oct. Declaración de Helsinki de la Amm - principios éticos para las investigaciones médicas en seres humanos. URL: https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principioseticos-para-las-investigaciones-medicas-en-seres-humanos/ [accessed 2020-03-04]
35. Asmar R, Khabouth J, Topouchian J, El Feghali R, Mattar J. Validation of three automatic devices for self-measurement of blood pressure according to the International Protocol: the Omron M3 Intellisense (HEM-7051-E), the Omron M2 Compact (HEM 7102-E), and the Omron R3-I Plus (HEM 6022-E). Blood Pressure Monitoring 2010;15(1):49-54. [doi: 10.1097/mbp.0b013e3283354b11]
36. Mazoteras Pardo V, Losa Iglesias ME, López Chicharro J, Becerro de Bengoa Vallejo R. The QardioArm app in the assessment of blood pressure and heart rate: reliability and validity study. JMIR Mhealth Uhealth 2017 Dec 15;5(12):e198. [doi: 10.2196/mhealth.8458]
37. Chahine MN, Topouchian J, Zelveian P, Hakobyan Z, Melkonyan A, Azaki A, et al. Validation of BP devices QardioArm® in the general population and Omron M6 Comfort® in type II diabetic patients according to the European Society of Hypertension International Protocol (ESH-IP). MDER 2017 Dec;Volume 11:11-20. [doi: 10.2147/mder.s142126]
38. Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, López-López D, Palomo-López P, Rodríguez-Sanz D, et al. The QardioArm blood pressure app for self-measurement in an obese population: validation study using the European Society of Hypertension International Protocol Revision 2010. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e11632. [doi: 10.2196/11632]
39. Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, López-López D, Rodríguez-Sanz D, Casado-Hernández I, et al. QardioArm upper arm blood pressure monitor against Omron m3 upper arm blood pressure monitor in patients with chronic kidney disease: a validation study according to the European Society of Hypertension International Protocol Revision 2010. J Med Internet Res 2019 Dec 2;21(12):e14686. [doi: 10.2196/14686]

XSL•FO
RenderX

40. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, et al. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. Journal of Hypertension 2016;34(4):600-612. [doi: 10.1097/hjh.0000000000000859]
41. Omboni S, Caserini M, Coronetti C. Telemedicine and M-Health in Hypertension Management: Technologies, Applications and Clinical Evidence. High Blood Press Cardiovasc Prev 2016 Sep;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
42. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: Patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
43. Ciemins EL, Arora A, Coombs NC, Holloway B, Mullette EJ, Garland R, et al. Improving Blood Pressure Control Using Smart Technology. Telemedicine and e-Health 2018 Mar;24(3):222-228. [doi: 10.1089/tmj.2017.0028]
44. Ruiz-Hurtado G, Ruilope LM. New strategy to control blood pressure: interactive mobile phone support. J Clin Hypertens 2015 Oct 14;18(2):109-110. [doi: 10.1111/jch.12683]
45. Ruiz-Hurtado G, Ruilope LM. New strategy to control blood pressure: interactive mobile phone support. J Clin Hypertens (Greenwich) 2016 Feb;18(2):109-110 [FREE Full text] [doi: 10.1111/jch.12683] [Medline: 26462732]
46. van Ittersum FJ, Wijering RM, Lambert J, Donker AJ, Stehouwer CD. Determinants of the limits of agreement between the sphygmomanometer and the SpaceLabs 90207 device for blood pressure measurement in healthy volunteers and insulin-dependent diabetic patients. Journal of Hypertension 1998;16(8):1125-1130. [doi: 10.1097/00004872-199816080-00007]
47. Friis Lauszus F, Rosgaard A, Lousen T, Winther Rasmussen O, Martin Klebe T, Klebe JG. Precision, consistency, and reproducibility of blood pressure in diabetic and non-diabetic pregnancy: the appraisal of repeated measurements. Acta Obstet Gynecol Scand 2007 Jan;86(9):1063-1070. [doi: 10.1080/00016340701472773]
48. Theilade S, Joergensen C, Persson F, Lajer M, Rossing P. Ambulatory tonometric blood pressure measurements in patients with diabetes. Diabetes Technology & Therapeutics 2012 Jun;14(6):453-456. [doi: 10.1089/dia.2012.0006]

Abbreviations

ESH-IP2: second International Protocol of the European Society of Hypertension

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by E Navarro Flores, MDP Marques León, AM Jiménez Cebrian; comments to author 12.06.20; revised version received 14.06.20; accepted 14.06.20; published 24.07.20.

Please cite as:
Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, Martínez-Jiménez EM, Calvo-Lobo C, Romero-Morales C, López-López D, Palomo-López P
QardioArm Blood Pressure Monitoring in a Population With Type 2 Diabetes: Validation Study
J Med Internet Res 2020;22(7):e19781
URL: http://www.jmir.org/2020/7/e19781/
doi:10.2196/19781
PMID:32706672

©Victoria Mazoteras-Pardo, Ricardo Becerro-De-Bengoa-Vallejo, Marta Elena Losa-Iglesias, Eva María Martínez-Jiménez, César Calvo-Lobo, Carlos Romero-Morales, Daniel López-López, Patricia Palomo-López. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding the Acceptability of Subdermal Implants as a Possible New HIV Prevention Method: Multi-Stage Mixed Methods Study

Christine Tagliaferri Rael[1], PhD; Cody Lentz[1], BS; Alex Carballo-Diéguez[1], PhD; Rebecca Giguere[1], MPH; Curtis Dolezal[1], PhD; Daniel Feller[2], MS; Richard T D'Aquila[3], MD; Thomas J Hope[4,5], PhD

[1]HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute, Columbia University, New York, NY, United States

[2]Department of Biomedical Informatics, Columbia University, New York, NY, United States

[3]Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

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

[5]McCormick School of Engineering and Applied Sciences, Northwestern University, Chicago, IL, United States

Corresponding Author:
Christine Tagliaferri Rael, PhD
HIV Center for Clinical and Behavioral Studies
New York State Psychiatric Institute
Columbia University
1051 Riverside Dr
New York, NY, 10032
United States
Phone: 1 6467746967
Email: cr2857@cumc.columbia.edu

Abstract

Background: A long-acting implant for HIV pre-exposure prophylaxis (PrEP) is in development in the Sustained Long-Action Prevention Against HIV (SLAP-HIV) trial. This could provide an alternative to oral PrEP.

Objective: Our mixed methods study aimed to understand (1) users' experiences with a similar subdermal implant for contraception and (2) factors influencing the likelihood that gay and bisexual men (GBM) would use a proposed PrEP implant.

Methods: Work was completed in 4 stages. In stage 1, we conducted a scientific literature review on existing subdermal implants, focusing on users' experiences with implant devices. In stage 2, we reviewed videos on YouTube, focusing on the experiences of current or former contraceptive implant users (as these implants are similar to those in development in SLAP-HIV). In stage 3, individuals who indicated use of a subdermal implant for contraception in the last 5 years were recruited via a web-based questionnaire. Eligible participants (n=12 individuals who liked implants a lot and n=12 individuals who disliked implants a lot) completed in-depth phone interviews (IDIs) about their experiences. In stage 4, results from IDIs were used to develop a web-based survey for HIV-negative GBM to rate their likelihood of using a PrEP implant on a scale (1=very unlikely and 5=very likely) based on likely device characteristics and implant concerns identified in the IDIs.

Results: In the scientific literature review (stage 1), concerns about contraceptive implants that could apply to the PrEP implants in development included potential side effects (eg, headache), anticipated high cost of the device, misconceptions about PrEP implants (eg, specific contraindications), and difficulty accessing PrEP implants. In the stage 2 YouTube review, individuals who had used contraceptive implants reported mild side effects related to their device. In stage 3, implant users reported that devices were comfortable, unintrusive, and presented only minor discomfort (eg, bruising) before or after insertion and removal. They mainly reported removing or disliking the device due to contraceptive-related side effects (eg, prolonged menstruation). Participants in the stage 4 quantitative survey (N=304) were mainly gay (204/238, 85.7%), white (125/238, 52.5%), cisgender men (231/238, 97.1%), and 42.0% (73/174) of them were on oral PrEP. Not having to take a daily pill increased the likelihood of using PrEP implants (mean 4.13). Requiring >1 device to achieve 1 year of protection (mean range 1.79-2.94) mildly discouraged PrEP implant use. Participants did not mind moderate bruising, a small scar, tenderness, or bleeding after insertion or removal, and an implant with a size slightly larger than a matchstick (mean ratings 3.18-3.69).

XSL•FO
RenderX

Conclusions: PrEP implants are promising among GBM. Implant features and insertion or removal-related concerns do not seem to discourage potential users. To ensure acceptability, PrEP implants should require the fewest possible implants for the greatest protection duration.

(J Med Internet Res 2020;22(7):e16904) doi:10.2196/16904

KEYWORDS

PrEP implant; YouTube; acceptability; long-acting PrEP; systemic PrEP; human-centered design; HIV prevention; removable implant; long-acting HIV prevention; gay and bisexual men

Introduction

Subdermal implants represent a promising, long-acting biomedical HIV prevention innovation in the development pipeline. A total of two trials are currently investigating subdermal implants for HIV pre-exposure prophylaxis (PrEP): (1) the Sustained Long-Acting Prevention against HIV (SLAP-HIV) trial (preclinical phases) [1,2] and (2) Centre for the AIDS Programme of Research in South Africa-018 (phase I or II) [3]. Little is known about whether such HIV prevention devices would be acceptable or feasible for end users [4]. This is a concern given that uptake of existing PrEP products (eg, oral PrEP) is suboptimal in some key populations who are disproportionately affected by the HIV epidemic. For example, in a 2017 cohort study of 995 gay and bisexual men (GBM) across the United States, only 4% of participants reported currently using oral PrEP [5]. The reasons for this limited uptake are continually emerging [6-9]. Involving potential consumers in the initial stages of product development could help mitigate later issues with method use. Specifically, early consumer involvement could help researchers understand end users' anticipated challenges with uptake and adherence, allowing them to address and potentially limit these problems in the design of clinical tools [10].

SLAP-HIV is an interdisciplinary research project that aims to identify, develop, and test a long-acting PrEP implant. A proposed version of this product could be inserted under the skin of a user's upper arm. This device is conceptualized to slowly dispense a single HIV prevention medication (such as tenofovir alafenamide or cabotegravir) over a period (eg, 2-12 months). Similar devices are currently used by cisgender women (ie, individuals who were assigned female at birth and identify as women) to dispense contraceptive medication (eg, Nexplanon) [11]. Thus, it is possible to learn from the experiences of current and past users of an implant similar to the one in development to understand their likes and dislikes about this mode of drug delivery.

This study was undertaken as part of SLAP-HIV. We aimed to identify barriers and facilitators anticipated by GBM who might use PrEP implants in the future. Although the end goal of this work focused on the opinions of GBM, we focused on cisgender women during the initial stages, as the contraceptive implant is the product most similar to the one in development by SLAP-HIV.

Methods

This work was completed in 4 stages, which are summarized below.

Stage 1: Scientific Literature Review

We conducted a review of the scientific literature on publicly available (eg, Google Scholar and PubMed) and university-specific (eg, Columbia University libraries) databases to identify barriers and facilitators of contraceptive implant use among cisgender women. We focused our review on contraceptive implants, as this device is most similar to the one in development by SLAP-HIV. Furthermore, there is extremely limited data on cisgender men's experiences with implants to deliver gender-specific medications (eg, the Testopel testosterone implant); these implants also operate differently than those under study in SLAP-HIV (eg, Testopel is a nonremoveable biodegradable pellet implant) [12]. We limited our search to studies published in the last 10 years (January 1998 to July 2017). Search terms included "contraceptive implant barriers," "contraceptive implant removal," "contraceptive implant facilitators," "contraceptive implant experiences," and "contraceptive implant continuation." We limited the search to studies published about US populations and manuscripts on arm-implant contraceptives (eg, studies focused only on intrauterine devices [IUDs] were excluded).

The primary author (CR) and a research assistant (CL) abstracted relevant studies and independently reviewed their full content to determine their applicability to the topic under study (eg, device-related barriers and facilitators to contraceptive implants). CR and CL met in person to discuss the results of the independent reviews, resolve a disagreement about study relevance in concert, and identify studies to move forward with. After the studies to be included in the final review were identified, CR tallied the specific mention of device-related and contraceptive medication−related barriers and facilitators in a word document. There was very little mention in the scientific literature about the device-related effects of contraceptive implants. Thus, stage 2 (summarized below) intended to learn more about users' device-related subdermal implant experiences, as this will be important for our target population, GBM (stage 4).

Stage 2: Review of YouTube Videos

A total of 2 members of the research team identified a collection of potentially relevant YouTube videos through a manual search. Search terms included lay and commercial names of contraceptives: "Nexplanon," "Nexplanon insertion," "Jadelle," "Jadelle implant," "Nexplanon birth control," "Norplant,"

XSL•FO
RenderX

"Implanon," "Implanon birth control review," and "Arm bar birth control." We excluded sponsored testimonials, insertion or removal footage that included no assessment of the patient's experience, clinical videos (ie, procedural videos on how to insert or remove implants), promotional or educational videos (ie, videos explaining what contraceptive implants are or how they work), and advertisements for any product. Then, spoken word transcripts of potentially relevant videos were automatically generated using YouTube's state-of-the-art speech recognition algorithms [13].

To identify biomedical terms in YouTube transcripts, we used a publicly available natural language processing system called National Center for Biomedical Ontology Annotator, maintained by the National Library of Medicine [14]. The system recognizes sets of words and phrases using string matching and regular expressions that operate directly on text without preprocessing, such as spelling correction, part-of-speech tagging, and word sense disambiguation [15]. To identify the various side effects associated with subdermal implants, we utilized Systemized Nomenclature of Medicine (SNOMED) [16], a comprehensive terminology that provided term-to-concept mapping for a vast array of conditions, symptoms, and clinical findings.

We restricted our search to several types of concepts or *semantic types*. We identified relevant semantic types by using SNOMED to annotate a string of words typically associated with unpleasant experiences with drug delivery implants. We then reported the frequency of the specific terms describing these unpleasant experiences (eg, contusion and pain). Terms were paired to ensure clarity in the themes to which they were intended to refer (eg, bleeding, menstrual spotting). We then quantified the frequency (eg, number of videos) with which they appeared.

Stage 3: In-Depth Interviews With Contraceptive Implant Users

In stage 3, we recruited potential participants using paid Facebook advertisements and posts on contraceptive implant-themed Facebook pages (eg, *Nexplanon* and *Implanon*) and other female-centered Facebook groups (eg, *Girls Love Travel*), with permission from page moderators. We exclusively sampled cisgender women in this stage because we wanted to ensure that the implant participants had used a design that was consistent with those of the PrEP implants in the development pipeline. Subdermal implants commonly available to men (eg, Testopel testosterone pellets, which are a nonremovable pellet implant, placed in the gluteal region) have a much different design [12,17]. This sampling approach had 2 purposes as it would allow (1) PrEP implant developers to incorporate feedback from past or current users of a similar implant into their device design and (2) the research team to be more precise about the questions asked to potential future PrEP implant users in the stage 4 web-based survey.

Participants were invited to click on a link that took them to a brief screening questionnaire to assess eligibility for telephone interviews about their experiences with implants. Eligible participants were at least 18 years old, identified as cisgender women, and had used an implant for contraception in the last 5 years. To sample a wide range of users' experiences, we selected those who reported that they *liked very much* or *disliked very much* their implant on a 4-point Likert scale (1=disliked very much, 2=disliked a little, 3=liked a little, and 4=liked very much). In all, 24 participants were selected for interviews on a *first come first serve* basis, 12 who *liked very much* and 12 who *disliked very much* their implants and could be reached using the telephone numbers they provided.

Telephone interviews were audio-recorded and guided by findings from the YouTube analysis and scientific literature review. Specifically, we focused on the in-depth phone interview (IDI) guide on themes from the scientific literature and those YouTube users referenced frequently (identified using SNOMED). We categorized these themes into 6 broad categories: (1) general questions about participants' experiences with subdermal implants, (2) decision making around the use of an implant, (3) acceptability of the implant, (4) the implantation process, (5) the removal process, and (6) recommendations for researchers working on a future PrEP implant. Participants were also asked about their reasons for removing the contraceptive implant. Upon completing the interviews, participants were compensated with US $50 Amazon gift cards.

Next, audio-recordings of the interviews were transcribed and cleaned for accuracy. To analyze the data, a codebook was developed jointly by the principal investigator of the project and a research assistant. It incorporated categories and themes from the interview guide and included definitions, inclusion and exclusion criteria, and examples of passages for inclusion. To validate and finalize the codebook, 2 researchers coded 2 transcripts independently, met to reconcile discrepant codes, and modified the codebook as needed to ensure that all themes of interest were being accurately captured. Researchers, then, proceeded to code the remainder of the transcripts in pairs, meeting after every 11 transcripts (half of the remaining data) were coded to reconcile discrepant codes. The finalized codes were analyzed using content analysis [18] to identify prevalent themes, illustrative examples of each theme, and knowledge gaps.

Stage 4: Web-Based Surveys With Gay and Bisexual Men

Findings from the IDIs among cisgender women in stage 3 informed the development of a web-based quantitative assessment. This assessment took approximately 3 min to complete and aimed to determine how likely members of key HIV risk demographics (ie, GBM) would be to use an implant for HIV prevention. Specifically, we published text-based pop-up advertisements (known as broadcasts) on Grindr, a dating and *hookup* app for GBM. Logistically, these advertisements popped up the first time any Grindr user within a specified location opened the app; users who clicked on the advertisement were automatically rerouted to the quantitative assessment. For the purposes of this study, we published 2 broadcast advertisements (per city) in New York City, Baltimore, and Chicago. The quantitative assessment was open to anyone who wished to take it; eligible participants reported identifying as a gay man who was at least 18 years old. We focused our recruitment on GBM (rather than cisgender heterosexual cisgender women) because this population is at comparatively

XSL•FO
RenderX

higher risk for HIV exposure and is one of the planned target populations for future PrEP implant studies. To reduce the odds of receiving multiple entries from the same person, participants did not receive compensation for their time, and surveys that had identical Internet Protocol addresses were eliminated. Additionally, for the purposes of the analyses presented in this paper, only HIV-negative GBM who lived in the United States were included.

As the PrEP implant is still hypothetical, we posed a variety of device configurations to participants. Specifically, the quantitative assessment measured how willing participants would be to use a hypothetical HIV prevention implant given a number of different factors: varying lengths of effectiveness (eg, 12 months, 6-8 months, and 2-3 months), number of implantable devices required, potential physical side effects (eg, bruising and scarring) resulting from the implantation or removal processes, implant size, and the convenience of using an implant compared with taking a daily HIV prevention pill (PrEP). All responses were recorded using a 5-point Likert scale, ranging from 1 (*I am very unlikely to use it)* to 5 (*I am very likely to use it)*. The value 3 on this scale represented a neutral response. Questions were presented in a different random order for each participant to minimize potential bias. Basic demographic, sexual risk, and injection drug use information were also collected using the web-based survey (Multimedia Appendix 1).

To give participants an idea of what a PrEP implant would entail, we embedded photos of the proposed implant next to a dime to help respondents estimate its size. Additionally, we embedded graphics interchange formats (GIFs: a lossless image file format that presents time-lapse video imagery) of the implant insertion and removal processes so that participants could gain an understanding of how they worked.

Results

Stage 1: Scientific Literature Review

In the scientific literature, all studies that were identified by the research team focused on cisgender women. Barriers to contraceptive implant uptake and sustained use most frequently included some that do not apply to an HIV prevention implant (eg, irregular or unpredictable menstrual bleeding [19-21]) and others that do apply, such as potential nonmenstrual side effects (eg, headache) [20,21]; perceived or actual high cost of implants [22-29]; patient or provider misconceptions or lack of information about implants (eg, believing that implants are contraindicated for specific patient populations) [21,23,26-32]; and challenges with access to implants [25,27]. Few studies discuss facilitators of the use of contraceptive implants as a birth control method and instead mostly focus on facilitators of implementation of programs designed to encourage implant use [26,33]. However, the scant body of existing evidence on implant facilitators shows that these include low implant cost, same-day insertion [27], and the inherent privacy of implants (eg, you can use this contraceptive method with low risk of others finding out) [29]. We were unable to identify scientific studies focused on user preferences regarding specific implant attributes (eg, size) or device-related effects (eg, bruising and local pain). Consequently, we sought additional sources of information.

Stage 2: Self-Reported Side Effects Associated With Contraceptive Implants

Stage 2 summarizes our systematic review of YouTube videos. In July 2017, 272 videos about users' experiences with contraceptive implants (that did not meet the exclusion criteria specified in the Methods section) had been uploaded to YouTube. Table 1 reports the frequency of videos in which the most often cited terms related to undesirable implant effects appear. These terms include the following: (1) pain, stinging, burning, or soreness and tenderness unrelated to menstruation (51 videos); (2) menstrual bleeding or spotting (33 videos); (3) contusion (32 videos); (4) scarring (25 videos); and (5) bleeding unrelated to menstruation (24 videos). We do not report demographic data as terms originated from videos posted on YouTube, and in the sample of videos accessed for the purposes of this study, posters rarely reported their demographic characteristics.

Table 1. Most frequently reported terms related to undesired implanted medication delivery device effects and the number of videos in which they are mentioned (stage 1; N=272 videos).

Term	Number of videos in which the term is referenced, n (%)
Pain, stinging, burning, or soreness and tenderness unrelated to menstrual bleeding or spotting	51 (18.8)
Menstrual bleeding or spotting	33 (12.1)
Contusion	32 (11.8)
Scarring	25 (9.2)
Bleeding unrelated to menstruation	24 (8.8)

Stage 3: Users' Reported Experiences With Contraceptive Implanted Medication Delivery Devices

Participants in IDIs were cisgender women who were aged, on average, 24 years and had used an implant for contraception in the last 5 years. Coders identified several themes within the topics previously recognized using SNOMED and covered in the IDI guide. Topics included the following: (1) specific attributes of implant products that attracted participants to these devices; (2) concerns about the implant; (3) feelings about the insertion process; (4) participants' observed likes and dislikes

XSL•FO
RenderX

about the implant, once it was inserted; and (5) recommendations for future implant designs and placement.

Specific Attributes of Implant Products That Attracted Participants

A majority of participants (n=20) reported that they were motivated to use contraceptive implants because they were long lasting and did not require additional attention or maintenance once inserted:

> *...I was on the pill for a long time and then I tried the shot and I didn't like either of those options. So I kind of wanted something a little bit more permanent that I didn't really have to worry about, which is why I thought it was a good fit for me...Like I said, it's kind of more permanent, which I liked. You don't really have to worry about it...* [Participant 52: age 26 years, liked implant very much]

> *I wanted something that I didn't have to worry about...Like trying to remember to take a pill every day and, you know, this, you know, you don't notice it's there. You know, I mean, you can feel it if you touch it, but you don't have to worry about anything. It's just there...You don't have to worry about taking a pill at the same time every day, or with the IUD, with it falling out, or something. I've heard stories of that. This you know is there. It's not going to come out, unless you get it taken out.* [Participant 46: age 36 years, liked implant very much]

Almost half of the participants (n=10) reported that they chose implants because they believed that the side effect profile was favorable to other contraceptive medications on the market and even had some positive effects (eg, reduction in or stoppage of menstrual bleeding):

> *Well, it just seems less, like there'd be less side effects. Like, I have friends and stuff who've had the IUD and they say it causes really bad period cramps and I've never really heard of any bad stories with the Nexplanon, so that's why I opted to go with it rather than the IUD.* [Participant 96: age 25 years, liked implant very much]

> *I liked the Nexplanon because there was a really high chance that I would not have a cycle, or I'd have a very light cycle because previous to that, I was having very heavy cycles, like painful, you know, just really couldn't even go sit in class all day, I was cramping so much, stuff like that. So, I was hoping I'd be in that percentage of like, lightening up a little bit, or not having a cycle at all...* [Participant 119: age 20 years, disliked implant very much]

Concerns About the Subdermal Implant

Although many participants felt that the side effects of contraceptive implants were less problematic than those of other contraceptives, they still expressed some concerns (n=12):

> *And then some of the cons that I had heard were weight gain, depression, mood swings, stuff like that. But at the time -- you hear different things about different birth controls and everybody's body's different so even though I had heard those few bad things I was still willing to give it a try.* [Participant 37: age 20 years, disliked implant very much]

> *…I anticipated highly irregular periods. I also anticipated looking for any sort of hormonal imbalances or emotional imbalances based on hormone usage. Nexplanon is my first birth control, so since I didn't have any prior experience I was just wary of hormonal medication at all...* [Participant 138: age 23 years, liked implant very much]

However, given that PrEP implants do not contain hormonal medications, these concerns would not apply.

Feelings About the Insertion Process

Many participants (n=15) felt that the insertion process was quick and straightforward:

> *it's very straightforward. It's really like a two-process insertion, which I thought was great, rather than them trying to feel around where to put it. They insert the little machine and click it in and that's it. So it's very simple...* [Participant 56: age 26 years, liked implant very much]

> *...I mean, it was pretty easy for them to...get it in there with the -- I guess they use a needle after they numb you, and they make the incision -- I guess it's a needle, I don't really -- it's a big one, but it's pretty easy to put in there...I guess when they made that incision, and then used the little insert - that was pretty fast. Like, it -- when they put it in it goes really fast. It's just waiting to be numb that took the longest.* [Participant 260: age 20 years, disliked implant very much]

Participants reported few dislikes about the insertion process, though a minority (n=5) felt intimidated by the size of the implant or applicator:

> *Yeah, like I mentioned, whenever you first see it -- and I'm not a big needle person, anyways, but whenever you first see it, it looks scary. The needle that they insert it with is pretty wide.* [Participant 37: age 20 years, disliked implant very much]

After the device was inserted, the majority of participants (n=21) experienced localized pain related to device insertion. This pain was mostly mild and included bruising, soreness, and bleeding:

> *Immediately, the area was like kind of a little bit sore, but like I said, it just like bruised a little bit and then it was fine...It was kind of bruised like for a couple of days, but after that it was fine.* [Participant 52: age 26 years, liked implant very much]

> *My arm bled for a little while after they got it in, so I had to keep the wrap on for three days and it was a really smooth transition putting it in...They recommended to keep it on for three days, but it was really, really tight, so it was a little rough to keep the wrap on for three days, and showering and moving around with a tight, brown wrap on your arm. So I*

XSL•FO
RenderX

did end up taking it off, I think, after two. But by then, it was fine. [Participant 16: age 24 years, disliked implant very much]

Participants' Likes and Dislikes About the Implant After It Was Inserted

After the implant was inserted, participants continued to view the long-lasting properties of the device and its little need for user attention as a benefit (n=13):

I liked that I didn't have to take something every day, that it was just there...but pretty much I didn't have to remember to take anything and I knew I wasn't going to get pregnant because it had worked for so many other girls before. [Participant 64: age 27 years, liked implant very much]

...Overall, I feel it's been more beneficial to my life than the pill was. Because, once again, convenience. I don't have to forget anything. I've got the little card in my wallet that says, "OK, this is when you need to get it removed." It's in my arm, it's there, good to go... [Participant 98: age 19 years, liked implant very much]

Additionally, almost half of the participants (n=10) liked the implant's effectiveness in preventing pregnancy:

...I like the fact that I know that I could have unprotected sex with this, and I won't get pregnant. And I don't have to use condoms, because it's 100 percent active, like effective...I know I'm not going to get pregnant because I don't need that, especially after I just had a baby not too long ago... [Participant 235: age 30 years, disliked implant very much]

Participants liked the implant's small size (n=13):

It's small...So, if it would have been something a bigger size, I probably wouldn't have put it on, because I -- to have this thing stay in my arm, it would have been probably maybe uncomfortable, but the fact that it's small, it -- I can't feel it. It's not heavy. [Participant 86: age 28 years, liked implant very much]

...it was smaller. I wouldn't have to worry about anybody really noticing, or being able to associate anything with having birth control... It just made me feel more comfortable that it was smaller, and that I wouldn't have to worry about if it was large or thick, restraining any activity, or anything like that, that it might bend or break or anything. [Participant 201: age 20 years, disliked implant very much]

Furthermore, participants liked that they could not feel the device while they carried out their normal daily activities (n=14):

I think, also the fact that you could tell like it was flexible, my job that I had, I would have to pull out of boxes that were kind of almost about like chest height, and I – there was one time I bent in, and I felt it hit against my arm, but because of it being flexible, it just kind of went with it instead of it had been more sturdy or like, more structured, I felt like it could have possibly broken in that situation just because I had forgotten about it and didn't even think about reaching in like that. [Participant 68: age 24 years, disliked implant very much]

Although participants liked the small size of the device and that it was relatively unnoticeable during normal activities, some participants (n=8) also liked that the device was detectable if they searched for it. They found this reassuring, as they could ensure that the device was still in the correct spot and had not migrated:

You can't -- with an IUD, you can't check to see if it has moved, really. I mean, you kind of can, but not really. With this implant, you can feel where it is. If it feels like it's starting to, like, imbed further, I could go into the OB and have them check to make sure it hasn't moved. [Participant 67: age 23 years, liked implant very much]

...it doesn't bother me. I can poke it, and I can see it, but other than that it doesn't bother me. It's just there...It doesn't weird me out or anything. It's normal. I got used to it. It was kind of weird at first. But as long as I know it's there, and it's not like going further into my tissue, then it makes me a lot more comfortable knowing it's actually still there. Because I've heard the horror stories of having to get it taken out and having to have an actual surgery to get it out versus it just being pulled out. [Participant 211: age 24 years, disliked implant very much]

Some participants (n=8) reported that the implant felt uncomfortable when bumped:

...If I bump into anything on that arm where the implant is, it's uncomfortable, but it doesn't swell or anything like that...When I bump it it's just kind of a dull pain, like almost like a bruise feels. It's kind of weird to explain because it's internal pain, but it's -- I don't know, it's kind of weird. It's like when you bruise your foot and you keep bumping it, it feels like that for a short while. [Participant 98: age 19 years, liked implant very much]

It's a thin rod. If it was any bigger, I think they would probably bother people because if I put pressure on it -- you know, like, sometimes you turn around or something in the car -- not while you're driving, obviously, but I'll turn around in the passenger seat to check on my kids. And if you push your arm where the implant is up against something, it will start to hurt. It'll start to irritate me, so I try not to put pressure on the implant. [Participant 67: age 23 years, liked implant very much]

Recommendations for Future PrEP Implant Designs and Placement

Participants had suggestions for researchers regarding the design of the PrEP implant device. Specifically, some participants wanted the device to be designed in a way that would facilitate easier removal (n=7), be small in size, constructed from a flexible material (n=15), and be placed in an area of the arm that would not interfere with the device (n=9):

XSL•FO
RenderX

...Maybe make it easier to remove if there's any way to do that. I think that's the only thing that's really deterring people is that they're scared it's going to get stuck...I've seen pictures where they've cut people a couple of times, and they can't seem to get it out. So if there's any way to make sure that it's going to be easy to come out with this design you guys are making -- I don't know if there's a way to do that, but...I wish they could remove it like that with the needle instead of cutting your arm. [Participant 67: age 23 years, liked implant very much]

> *It's like a slim rod that's really lightweight and plastic so like you don't feel it, and it just comes -- so, it goes in easily, but yeah, if it was shorter, it probably would have been a little more appealing...I don't really have an opinion about the size. I mean, it was I think a couple inches long, which is kind of big, and I probably would have liked it to be a little smaller, but it still wasn't too bad.* [Participant 66: age 21 years, liked implant very much]

> *I don't feel like it hit on anything; I don't feel like it would hit against like the underwire of my bra because of where it was placed. And so, it was very convenient because it was just out of the way...I'd say keep the flexibleness of that, but just kind of, make sure it's like as in-invasive to somebody's day-to-day life as possible. So, placed in a spot that isn't going to affect people that have different lifestyles, or maybe even have different options based on people's lifestyles and what they do, that that could be a possibility too.* [Participant 68: age 24 years, disliked implant very much]

Stage 4: Gay and Bisexual Men's Likelihood of Using a PrEP Implant

Participants (N=304) in the stage 4 web-based survey were aged, on average, 38 years and predominantly self-identified as gay or homosexual (204/238, 85.7%) cisgender men (231/238, 97.1%); about half (125/238, 52.5%) of them identified as white and a quarter (58/238, 24.4%) identified as Latino. Over two-thirds of the sample were college graduates (89/238, 37.4%) or had attended graduate school (77/238, 32.4%) and reported ever (101/174, 43.4%) or currently (73/174, 30.7%) using oral PrEP. Condomless sex with an HIV-positive or unknown status partner (104/126, 43.7%) was the most frequently reported HIV risk behavior in the past year (Table 2).

XSL•FO
RenderX

Table 2. Demographic information (stage 4; N=238).

Sample characteristics	Value, mean (SD)	Value, range	Value, n (valid %)[a]
Age (years; N=238)	36.44 (11.33)	18-71	N/A[b]
Race/ethnicity (N=238)			
White	N/A	N/A	125 (52.5)
Latino	N/A	N/A	58 (24.4)
African American	N/A	N/A	27 (11.3)
Asian or Pacific Islander	N/A	N/A	19 (8.0)
Other	N/A	N/A	9 (3.7)
Education (N=238)			
College graduate	N/A	N/A	89 (37.4)
Graduate school	N/A	N/A	77 (32.4)
Some college	N/A	N/A	40 (16.8)
Trade/Technical/Vocational	N/A	N/A	12 (5.0)
High school graduate/general education diploma	N/A	N/A	19 (8.0)
Less than high school	N/A	N/A	1 (0.4)
Gender identity (N=238)			
Man	N/A	N/A	231 (97.1)
Transgender, gender queer, nonbinary	N/A	N/A	7 (3.0)
Sexual orientation (N=237)			
Gay or homosexual	N/A	N/A	204 (86.1)
Bisexual	N/A	N/A	28 (11.8)
Straight or heterosexual	N/A	N/A	2 (0.8)
Other	N/A	N/A	3 (1.3)
Pre-exposure prophylaxis (N=174)			
Ever used	N/A	N/A	101 (58.0)
Currently using	N/A	N/A	73 (42.0)
HIV risk behaviors in past year (N=126)			
Condomless sex with HIV-positive or unknown status partner	N/A	N/A	104 (82.5)
Traded sex for money or other goods or services	N/A	N/A	16 (12.7)
Injected drugs	N/A	N/A	6 (4.8)

[a]Valid percent is the percent when missing data are excluded from the calculation.

[b]N/A: not applicable.

Table 3 shows the participants' preferences related to specific PrEP implant characteristics. Of these characteristics, not having to take a daily pill increased the likelihood of PrEP implant use (mean 4.13). Requiring more than 1 device to achieve 1 year of protection (mean participant rating range:1.79-2.94) mildly discouraged implant use. Participants did not mind the following: moderate bruising, a small scar, tenderness, or bleeding after insertion or removal, or having an implant slightly longer or fatter than a matchstick.

XSL•FO
RenderX

Table 3. Participants' likelihood of using pre-exposure prophylaxis implants based on their specific potential attributes (stage 4; N=304).

Implant characteristic	Likelihood of use[a]
Duration of protection, mean (SD)	
Protection lasts up to 12 months	3.6 (1.5)
Protection lasts up to 6-8 months	3.3 (1.5)
Protection lasts up to 2-3 months	2.7 (1.6)
Number of implants needed, mean (SD)	
2 implants needed	2.9 (1.6)
3 implants needed	2.2 (1.4)
4 implants needed	1.8 (1.3)
Potential application site related−side effects, mean (SD)	
Moderate bruising	3.7 (1.4)
Small scar	3.5 (1.5)
Implant can be felt	3.5 (1.5)
Some pain	3.6 (1.4)
Bleeding	3.6 (1.4)
Tenderness	3.5 (1.4)
Other PrEP[b] implant device–related characteristics, mean (SD)	
No daily oral PrEP	4.1 (1.3)
Trocar is used to insert implant	3.7 (1.4)
Must go to the doctor's office for removal	3.9 (1.4)
Implant longer than a matchstick	3.5 (1.4)
Implant fatter than a matchstick	3.2 (1.4)

[a]Values presented are descriptive statistics only; value of 3 on the Likert scale indicates a neutral rating.

[b]PrEP: pre-exposure prophylaxis.

Discussion

The scientific literature review conducted in stage 1 did not yield data on specific implant *device-related* barriers or facilitators. However, the existing literature did show that many of the problems associated with *general* contraceptive implant use are similar to those observed in relation to oral PrEP use. For example, several studies identified patient or provider misconceptions or lack of information about implants as a barrier to their uptake. Specifically, patients who have providers who harbor misconceptions about PrEP (eg, PrEP is only moderately effective) are less likely to obtain a prescription for this medication from their provider [34]. Additionally, concerns about contraceptive implants' side effect profiles and potential high cost deterred use; these same barriers also affected oral PrEP use [35]. Finally, access to contraceptive implants or oral PrEP was a challenge to use for each of these products [36]. All these concerns are likely to translate to PrEP implants. Suggestions from the contraceptive and oral PrEP studies on how to overcome these issues may be helpful. For instance, concerted efforts (eg, trainings) to ensure that providers have accurate information about PrEP [37,38] implants (including information on insertion and removal techniques) and effective tools to identify PrEP implant candidates could be helpful. Same-day access to PrEP implant insertion could also boost uptake [39]; offering PrEP implants in emergency department settings could also boost implant uptake.

Potential device side effects were a concern, as expressed by participants in stages 2, 3, and 4. Specifically, in stage 2, YouTube videos identified a number of potential side effects related to contraceptive implant use that could potentially be concerns for PrEP implant users, including pain, stinging, burning, soreness or tenderness, contusion, scarring, and bleeding. Participants in stage 3 also reported these effects after contraceptive implants were inserted or removed, although this was typically mild. Results from stage 4 showed that the possibility of experiencing these effects (eg, moderate bruising, small scar, some pain, bleeding, and tenderness) did not impact GBM participants' likelihood of using PrEP implants. Although these outcomes may not present a barrier to use for these men, this is not the case for all potential users. For example, other studies have shown that some transgender women were concerned that mild scarring and other skin-related reactions related to potential future PrEP implants could become serious [40,41]. Thus, it may be necessary to engage novel strategies to communicate actual risk [41], as some populations may overestimate this. Icon arrays, which communicate risk using graphical representations of icons symbolizing people affected

XSL•FO
RenderX

by a given outcome (eg, circles), may be one such strategy [42]. Icon arrays are shown to be better at communicating risk for low-risk procedures and may be especially effective for populations that have variability in numeracy skills [43].

Stage 2 YouTube videos reported menstrual complaints related to contraceptive implants. Stage 3 interviews also revealed that contraceptive implants had a negative effect on menstruation for some users. Specifically, over half of the participants endorsed this idea, complaining of prolonged or unpredictable cycles. It is highly unlikely that the menstrual or hormonal issues identified in stages 2 and 3 would be problematic for PrEP implant users, as these effects are associated with the use of hormonal medications, not antiretrovirals. However, given that this is such a pervasive issue among contraceptive implant users, it is paramount to ensure that potential PrEP implant users do not attribute these effects to implants in general. Rather, we must clarify that implants are a method to distribute a variety of medications. They are not a medication themselves. Potential PrEP implant users must be made aware that hormonal medications used in contraceptive implants, and antiretroviral medications used in PrEP implants will have completely different effects from one another. Communicating this will be particularly important for cisgender and transgender women users, as they may have greater familiarity with contraceptive implants and may be more concerned about how PrEP implants would interact with female hormones. Stage 4 participants were not asked about this theme because (1) this concern would not apply to male users and (2) medications contained in PrEP implants are not anticipated to interfere with male hormones.

Although negative side effects can deter implant use, other side effects may be seen as positive. Nearly half of the participants in stage 3 revealed that they chose the contraceptive implant because they believed that it had a more favorable side effect profile than other available options. Although specific side effects of the antiretroviral drug that would be delivered via implant are still under study (eg, these side effects may be different, and data show that these drugs are well tolerated [44,45]), it is critical to consider their side effects parallel to the HIV prevention effects of these medications. Users' experiences with oral PrEP show that this is the case; fear or experience with unfavorable short- (eg, gastrointestinal effects) or long-term side effects (eg, renal toxicity) are factors for drug discontinuation in 9.2%-33.0% of GBM who stop taking PrEP [46-48]. implants that distribute PrEP medications that have side effect profiles perceived as too intolerable may have limited uptake and continued use.

Conversely, if they seem to have fewer or beneficial side effects due to the delivery method, this could be viewed as positive. For instance, some stage 3 participants reported using contraceptive implants because they have the potential to have the beneficial side effect of lessened menstruation. Although it is highly unlikely that PrEP implants would have a similar effect, it is important to note the appeal of beneficial side effects. This is particularly important for those beneficial side effects that produce a notable enhancement in the life or appearance of the user, since these enhancements could potentially facilitate improved uptake or sustained use. Beneficial side effects have been noted in other medications, including finasteride (hair regrowth in men experiencing male pattern baldness [49] and initially used to treat benign prostatic hyperplasia [50]) or oral contraceptives (facial acne reduction and initially used to prevent pregnancy [51]). If and when beneficial side effects are identified in drugs contained in PrEP implants, this could play a role in future messaging and uptake campaigns.

The effectiveness and duration of products as well as the amount of maintenance they would require were important to participants in stages 2-4. Specifically, participants in stage 3 liked that the contraceptive implant was effective at preventing pregnancy and lasted for a prolonged period, during which little maintenance was required on their behalf (eg, they did not have to perform a routine activity, such as taking a daily pill). Participants in stage 4 shared this sentiment, and not having to take a daily pill increased the likelihood of using the PrEP implant. Additionally, in stage 4, protection against HIV lasting 6 months or more modestly increased potential users' likelihood of implant use; protection lasting 2-3 months modestly decreased the likelihood of use. Given this, it is critical in development phases to prioritize the amount of time that the implant will last and the amount of maintenance it requires.

The size and number of devices were also important. For example, in stage 3, participants found the small size and unintrusiveness (eg, most participants could not feel the implant during daily activities) of the contraceptive implant to be positive attributes. In stage 4, participants found the potential implant size (slightly longer and fatter than a matchstick) neither motivated nor dissuaded the likelihood of use. This may mean that the current size is acceptable. However, reducing the size of the implant could have a positive impact on potential users' perceptions of it. Related to this, stage 4 participants also reported that as the number of implants necessary to provide protection against HIV increases, the likelihood of use of this prevention strategy decreases. Taken together, this means that implant size and the number of implants necessary to achieve HIV protection are also critical factors for potential user uptake. The trajectory of contraceptive implant development shows that although it may not be possible to devise a small, 1-device implant in the first product iteration, it is possible to reduce implant size and number over time [11,52]. Thus, developers of PrEP implants should aim to reduce the size and number of devices, even if this is not immediately possible.

Participants in stage 3 (all cisgender women) reported that the insertion and removal procedures for contraceptive implants were quick and straightforward, with minimal pain. There was no mention of discomfort related to having this procedure done in a health care provider's office. However, a minority of participants reported feeling intimidated by the size of the trocar used to insert the contraceptive implant. This was not the case in stage 4. Stage 4 participants (all GBM) did not find the potential trocar planned for use in clinical research to implant the PrEP implant to be a barrier to uptake. It is important to note that no participants in stage 4 had ever undergone this procedure, although stage 3 participants had. This could have contributed to differences in participants' level of comfort with trocar insertion.

XSL•FO
RenderX

Furthermore, having the implant t removed by a health care provider appeared to mildly improve the likelihood of use among stage 4 participants. Thus, it appears that users of contraceptive implants and potential users of PrEP implants are mostly comfortable with the proposed model for device application and removal. However, although we can only speculate, it may be that this level of comfort is based on the proposed length of time that the implant could be used. That is, potential users may be less likely to undergo these insertion and removal processes if the implant does not last for very long. This again highlights the importance of the device duration.

Limitations

In stage 2, we may have missed some references to specific participant experiences if participants used descriptive words not contained in SNOMED's database. The interviews we conducted as part of stage 3 were partially meant to address this potential gap by more fully fleshing out users' experiences with contraceptive implants. As is the case with all research on acceptability and feasibility of hypothetical clinical products, it is difficult to estimate the gap between participants' perceived likelihood of use and actual use of implants for PrEP. However, data from past and current users of similar implants for contraception in stages 2 and 3 help to mediate this issue.

Additionally, because we were concerned with ensuring that users' experiences with existing implants were as similar as possible to potential future PrEP implants, we sampled only women in stage 3. In stage 4, we targeted only GBM (as they have a higher risk for HIV exposure compared with heterosexual cisgender women) with web-based surveys on their preferences for PrEP implants. Given the lack of formative work specific to GBM, there may be a gap in the preferences and perceptions of this implant product between these 2 groups. Future studies should conduct qualitative interviews with GBM on their preferences around PrEP implant products to limit this shortcoming. Despite these limitations, given the dearth of research on this topic, this work makes an important and unique contribution to the literature on implants for PrEP.

Conclusions

In all stages of research, participants reported experiencing (stages 2 and 3) or anticipating (stage 4) implant-related side effects. These effects were related to the device itself (bleeding, bruising, and pain after insertion) or, among contraceptive implant users, medication-related side effects (menstrual changes, weight gain, and acne). Although it is highly unlikely that PrEP implant users will experience similar hormonal side effects to contraceptive implant users, it is critical to point out the difference between these 2 products to potential PrEP users. Icon arrays could be a useful way to communicate serious risks related to implant devices. Future studies should monitor for the presence of beneficial side effects related to implant medication, as this has been evidenced with other medications and may improve uptake and sustained use. Furthermore, small implants need only 1 device, last for at least six months, are unintrusive, and require little to no user effort to optimize uptake. Current and potential future users appear comfortable with the currently used procedures for implant insertion and removal. Finally, future work should include more formative work specific to GBM, as this will allow researchers to address their concerns with greater precision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based survey.
[PDF File (Adobe PDF File), 837 KB - jmir_v22i7e16904_app1.pdf]

References

1. Hope TJ, Kiser P. Next Generation Drugs and Delivery for PrEP: Sustained Long-Acting Protection from HIV (SLAP-HIV). In: Proceedings of the 9th IAS Conference on HIV Science. 2017 Presented at: IAS'17; July 23-26, 2017; Paris, France.
2. Hope TJ, Kiser PF. Feinberg Labs: Northwestern University. 2017. Sustained Long-Acting Protection From HIV (SLAP HIV) Program URL: https://labs.feinberg.northwestern.edu/slap-hiv/ [accessed 2020-04-13]
3. Karim SS, Gengiah TN, Karim QA. Caprisa. 2018. A Phase I/II Trial to Assess the Safety, Acceptability, Tolerability and Pharmacokinetics of a Sustained-Release Tenofovir Alafenamide Sub-Dermal Implant for HIV Prevention in Women URL: https://www.caprisa.org/DBFile/Files/29a84f2d-d267-41c9-8e98-981a4cb9769d/CAPRISA%20018_Study%20protocol%20V1.1_1June18_FINAL.pdf [accessed 2020-04-13]
4. Calder BJ, Schieffer RJ, Tfaily EB, D'Aquila R, Greene GJ, Carballo-Diéguez A, et al. Qualitative consumer research on acceptance of long-acting pre-exposure prophylaxis products among men having sex with men and medical practitioners in the United States. AIDS Res Hum Retroviruses 2018 Oct;34(10):849-856 [FREE Full text] [doi: 10.1089/AID.2018.0214] [Medline: 30229684]
5. Parsons JT, Rendina HJ, Lassiter JM, Whitfield TH, Starks TJ, Grov C. Uptake of HIV pre-exposure prophylaxis (PrEP) in a national cohort of gay and bisexual men in the United States. J Acquir Immune Defic Syndr 2017 Mar 1;74(3):285-292 [FREE Full text] [doi: 10.1097/QAI.0000000000001251] [Medline: 28187084]

XSL•FO
RenderX

6. Ayala G, Makofane K, Santos G, Beck J, Do T, Hebert P, et al. Access to basic HIV-related services and PrEP acceptability among men who have sex with men worldwide: barriers, facilitators, and implications for combination prevention. J Sex Transm Dis 2013;2013:953123 [FREE Full text] [doi: 10.1155/2013/953123] [Medline: 26316968]
7. Bauermeister J, Meanley S, Pingel E, Soler J, Harper G. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct;11(7):520-527 [FREE Full text] [doi: 10.2174/1570162x12666140129100411] [Medline: 24476355]
8. Gallagher T, Link L, Ramos M, Bottger E, Aberg J, Daskalakis D. Self-perception of HIV risk and candidacy for pre-exposure prophylaxis among men who have sex with men testing for HIV at commercial sex venues in New York City. LGBT Health 2014 Sep;1(3):218-224. [doi: 10.1089/lgbt.2013.0046] [Medline: 26789715]
9. Grov C, Whitfield TH, Rendina HJ, Ventuneac A, Parsons JT. Willingness to take PrEP and potential for risk compensation among highly sexually active gay and bisexual men. AIDS Behav 2015 Dec;19(12):2234-2244 [FREE Full text] [doi: 10.1007/s10461-015-1030-1] [Medline: 25735243]
10. Tolley EE, Morrow KM, Owen DH. Designing a multipurpose technology for acceptability and adherence. Antiviral Res 2013 Dec;100(Suppl):S54-S59 [FREE Full text] [doi: 10.1016/j.antiviral.2013.09.029] [Medline: 24188706]
11. Merck. 2018. FDA-Approved Patient Labeling: Nexplanon (Etonogestrel Implant) URL: https://www.merck.com/product/usa/pi_circulars/n/nexplanon/nexplanon_ppi.pdf [accessed 2020-03-31]
12. Cavender RK, Fairall M. Subcutaneous testosterone pellet implant (testopel) therapy for men with testosterone deficiency syndrome: a single-site retrospective safety analysis. J Sex Med 2009 Nov;6(11):3177-3192. [doi: 10.1111/j.1743-6109.2009.01513.x] [Medline: 19796052]
13. YouTube. 2019. URL: https://www.youtube.com [accessed 2020-03-31]
14. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. Nucleic Acids Res 2011 Jul;39:W541-W545 [FREE Full text] [doi: 10.1093/nar/gkr469] [Medline: 21672956]
15. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story M, NCBO Team. The national center for biomedical ontology. J Am Med Inform Assoc 2012;19(2):190-195 [FREE Full text] [doi: 10.1136/amiajnl-2011-000523] [Medline: 22081220]
16. Shahpori R, Doig C. Systematized nomenclature of medicine-clinical terms direction and its implications on critical care. J Crit Care 2010 Jun;25(2):364.e1-364.e9. [doi: 10.1016/j.jcrc.2009.08.008] [Medline: 19836194]
17. RxList. 2020. TESTOPEL (Testosterone) Pellets URL: https://www.rxlist.com/testopel-drug.htm [accessed 2020-04-13]
18. Patton MQ. Qualitative Research & Evaluation Methods. Third Edition. Thousand Oaks, CA: Sage Publications; 2002.
19. Mansour D, Bahamondes L, Critchley H, Darney P, Fraser IS. The management of unacceptable bleeding patterns in etonogestrel-releasing contraceptive implant users. Contraception 2011 Mar;83(3):202-210. [doi: 10.1016/j.contraception.2010.08.001] [Medline: 21310280]
20. Blumenthal PD, Gemzell-Danielsson K, Marintcheva-Petrova M. Tolerability and clinical safety of implanon. Eur J Contracept Reprod Health Care 2008 Jun;13(Suppl 1):29-36. [doi: 10.1080/13625180801960012] [Medline: 18330815]
21. Russo JA, Miller E, Gold MA. Myths and misconceptions about long-acting reversible contraception (LARC). J Adolesc Health 2013 Apr;52(4 Suppl):S14-S21. [doi: 10.1016/j.jadohealth.2013.02.003] [Medline: 23535052]
22. Eisenberg D, McNicholas C, Peipert JF. Cost as a barrier to long-acting reversible contraceptive (LARC) use in adolescents. J Adolesc Health 2013 Apr;52(4 Suppl):S59-S63. [doi: 10.1016/j.jadohealth.2013.01.012] [Medline: 23535059]
23. Kumar N, Brown JD. Access barriers to long-acting reversible contraceptives for adolescents. J Adolesc Health 2016 Sep;59(3):248-253. [doi: 10.1016/j.jadohealth.2016.03.039] [Medline: 27247239]
24. Zerden ML, Tang JH, Stuart GS, Norton DR, Verbiest SB, Brody S. Barriers to receiving long-acting reversible contraception in the postpartum period. Womens Health Issues 2015;25(6):616-621. [doi: 10.1016/j.whi.2015.06.004] [Medline: 26212318]
25. Kavanaugh ML, Jones RK, Finer LB. Perceived and insurance-related barriers to the provision of contraceptive services in US abortion care settings. Womens Health Issues 2011;21(3 Suppl):S26-S31. [doi: 10.1016/j.whi.2011.01.009] [Medline: 21530835]
26. Politi MC, Estlund A, Milne A, Buckel CM, Peipert JF, Madden T. Barriers and facilitators to implementing a patient-centered model of contraceptive provision in community health centers. Contracept Reprod Med 2016;1:21 [FREE Full text] [doi: 10.1186/s40834-016-0032-3] [Medline: 29201410]
27. Satterwhite CL, Ramaswamy M. Let's talk about sex (again): advancing the conversation around long-acting reversible contraception for teenagers. Womens Health (Lond) 2015 Nov;11(6):841-850. [doi: 10.2217/whe.15.66] [Medline: 26626398]
28. Pritt NM, Norris AH, Berlan ED. Barriers and facilitators to adolescents' use of long-acting reversible contraceptives. J Pediatr Adolesc Gynecol 2017 Mar;30(1):18-22. [doi: 10.1016/j.jpag.2016.07.002] [Medline: 27477904]
29. Miller MK, Randell KA, Barral R, Sherman AK, Miller E. Factors associated with interest in same-day contraception initiation among females in the pediatric emergency department. J Adolesc Health 2016 Mar;58(2):154-159 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.10.016] [Medline: 26802990]
30. Madden T, Secura GM, Nease RF, Politi MC, Peipert JF. The role of contraceptive attributes in women's contraceptive decision making. Am J Obstet Gynecol 2015 Jul;213(1):46.e1-46.e6 [FREE Full text] [doi: 10.1016/j.ajog.2015.01.051] [Medline: 25644443]

XSL•FO
RenderX

31. Murphy MK, Stoffel C, Nolan M, Haider S. Interdependent barriers to providing adolescents with long-acting reversible contraception: qualitative insights from providers. J Pediatr Adolesc Gynecol 2016 Oct;29(5):436-442 [FREE Full text] [doi: 10.1016/j.jpag.2016.01.125] [Medline: 26851537]
32. Hall KS, Ela E, Zochowski MK, Caldwell A, Moniz M, McAndrew L, et al. 'I don't know enough to feel comfortable using them:' women's knowledge of and perceived barriers to long-acting reversible contraceptives on a college campus. Contraception 2016 Jun;93(6):556-564 [FREE Full text] [doi: 10.1016/j.contraception.2016.02.007] [Medline: 26879627]
33. Hallum-Montes R, Middleton D, Schlanger K, Romero L. Barriers and facilitators to health center implementation of evidence-based clinical practices in adolescent reproductive health services. J Adolesc Health 2016 Mar;58(3):276-283 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.11.002] [Medline: 26903427]
34. Castel AD, Feaster DJ, Tang W, Willis S, Jordan H, Villamizar K, et al. Understanding HIV care provider attitudes regarding intentions to prescribe PrEP. J Acquir Immune Defic Syndr 2015 Dec 15;70(5):520-528 [FREE Full text] [doi: 10.1097/QAI.0000000000000780] [Medline: 26247895]
35. Bekker LG, Hughes J, Amico R, Roux S, Hendrix C, Anderson PL, et al. HPTN 067/ADAPT Cape Town: A Comparison of Daily and Nondaily PrEP Dosing in African Women. In: Proceedings of the Conference on Retroviruses And Opportunistic Infections. 2015 Presented at: CROI'15; February 23-26, 2015; Seattle, Washington URL: http://www.croiconference.org/sessions/hptn-067adapt-cape-town-comparison-daily-and-nondaily-prep-dosing-african-women
36. Elopre L, Kudroff K, Westfall AO, Overton ET, Mugavero MJ. Brief report: the right people, right places, and right practices: disparities in PrED access among African American men, women, and MSM in the deep south. J Acquir Immune Defic Syndr 2017 Jan 1;74(1):56-59 [FREE Full text] [doi: 10.1097/QAI.0000000000001165] [Medline: 27552156]
37. Bacon O, Gonzalez R, Andrew E, Potter MB, Iñiguez JR, Cohen SE, et al. Brief report: informing strategies to build PrEP capacity among San Francisco bay area clinicians. J Acquir Immune Defic Syndr 2017 Mar 1;74(2):175-179 [FREE Full text] [doi: 10.1097/QAI.0000000000001182] [Medline: 27654813]
38. Krakower DS, Mayer KH. The role of healthcare providers in the roll out of preexposure prophylaxis. Curr Opin HIV AIDS 2016 Jan;11(1):41-48 [FREE Full text] [doi: 10.1097/COH.0000000000000206] [Medline: 26417953]
39. Biggs MA, Arons A, Turner R, Brindis CD. Same-day LARC insertion attitudes and practices. Contraception 2013 Nov;88(5):629-635. [doi: 10.1016/j.contraception.2013.05.012] [Medline: 23809277]
40. Rael C, Bockting W, MacCrate C, Martinez M, Mellman W, Valente P, et al. Barriers and Facilitators to Potential Future Long-acting Biomedical HIV Prevention Strategies for Transgender Women. In: Proceedings of the International AIDS Society Meeting. 2017 Presented at: IAS'17; July 23-26, 2017; Paris, France.
41. Rael CT, Martinez M, Giguere R, Bockting W, MacCrate C, Mellman W, et al. Transgender women's concerns and preferences on potential future long-acting biomedical HIV prevention strategies: the case of injections and implanted medication delivery devices (IMDDS). AIDS Behav 2019 Oct 25:- epub ahead of print. [doi: 10.1007/s10461-019-02703-5] [Medline: 31654172]
42. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc 2006;13(6):608-618 [FREE Full text] [doi: 10.1197/jamia.M2115] [Medline: 16929039]
43. Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. Am J Public Health 2009 Dec;99(12):2196-2202. [doi: 10.2105/AJPH.2009.160234] [Medline: 19833983]
44. Muoio D. Healio: Medical News, Journals, and Free CME. 2016. FDA Approves Descovy for Treatment of HIV URL: https://www.healio.com/infectious-disease/hiv-aids/news/online/%7B13f47f79-1d1b-449a-9833-30206c327ca2%7D/fda-approves-descovy-for-treatment-of-hiv [accessed 2020-04-13]
45. Rai MA, Pannek S, Fichtenbaum CJ. Emerging reverse transcriptase inhibitors for HIV-1 infection. Expert Opin Emerg Drugs 2018 Jun;23(2):149-157 [FREE Full text] [doi: 10.1080/14728214.2018.1474202] [Medline: 29737220]
46. Whitfield TH, John SA, Rendina HJ, Grov C, Parsons JT. Why I quit pre-exposure prophylaxis (PrEP)? A mixed-method study exploring reasons for prep discontinuation and potential re-initiation among gay and bisexual men. AIDS Behav 2018 Nov;22(11):3566-3575 [FREE Full text] [doi: 10.1007/s10461-018-2045-1] [Medline: 29404756]
47. Morgan E, Ryan DT, Newcomb ME, Mustanski B. High rate of discontinuation may diminish PrEP coverage among young men who have sex with men. AIDS Behav 2018 Nov;22(11):3645-3648 [FREE Full text] [doi: 10.1007/s10461-018-2125-2] [Medline: 29728950]
48. Holloway IW, Dougherty R, Gildner J, Beougher SC, Pulsipher C, Montoya JA, et al. Brief report: PrEP uptake, adherence, and discontinuation among California YMSM using geosocial networking applications. J Acquir Immune Defic Syndr 2017 Jan 1;74(1):15-20 [FREE Full text] [doi: 10.1097/QAI.0000000000001164] [Medline: 27552158]
49. Leyden J, Dunlap F, Miller B, Winters P, Lebwohl M, Hecker D, et al. Finasteride in the treatment of men with frontal male pattern hair loss. J Am Acad Dermatol 1999 Jun;40(6):930-937. [doi: 10.1016/s0190-9622(99)70081-2]
50. Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003 Jul 17;349(3):215-224. [doi: 10.1056/NEJMoa030660] [Medline: 12824459]
51. Williams HC, Dellavalle RP, Garner S. Acne vulgaris. Lancet 2012 Jan 28;379(9813):361-372. [doi: 10.1016/S0140-6736(11)60321-8] [Medline: 21880356]

XSL•FO
RenderX

52. Satoskar R, Rege N, Bhandarkar S. Pharmacology and Pharmacotherapeutics. Twenty-Fourth Edition. New Delhi, India: Elsevier Health Sciences; 2015.

Abbreviations

GBM: gay and bisexual men
IUD: intrauterine device
IDI: in-depth phone interview
PrEP: pre-exposure prophylaxis
SLAP-HIV: Sustained Long-Action Prevention Against HIV
SNOMED: Systemized Nomenclature of Medicine

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by F Gorman, T Torres, B Castonguay; comments to author 14.02.20; revised version received 27.02.20; accepted 21.03.20; published 27.07.20.

Please cite as:
Rael CT, Lentz C, Carballo-Diéguez A, Giguere R, Dolezal C, Feller D, D'Aquila RT, Hope TJ
Understanding the Acceptability of Subdermal Implants as a Possible New HIV Prevention Method: Multi-Stage Mixed Methods Study
J Med Internet Res 2020;22(7):e16904
URL: https://www.jmir.org/2020/7/e16904
doi:10.2196/16904
PMID:32348277

©Christine Tagliaferri Rael, Cody Lentz, Alex Carballo-Diéguez, Rebecca Giguere, Curtis Dolezal, Daniel Feller, Richard T D'Aquila, Thomas J Hope. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open MS BioScreen): Human-Centered Design and Development

Erica Schleimer[1], BSc; Jennifer Pearce[2], MPA; Andrew Barnecut[1], BA; William Rowles[1], BA; Antoine Lizee[1], MSc; Arno Klein[3], PhD; Valerie J Block[1], DPTSc; Adam Santaniello[1], BSc; Adam Renschen[1], BA; Refujia Gomez[1], BSc; Anisha Keshavan[1], PhD; Jeffrey M Gelfand[1], MD, MAS; Roland G Henry[1], PhD; Stephen L Hauser[1], MD; Riley Bove[1], MD, MMSc

[1]Department of Neurology, UCSF Weill Institute for Neurosciences, San Francisco, CA, United States

[2]Plain Language Health, San Francisco, CA, United States

[3]Child Mind Institute, New York, NY, United States

Corresponding Author:
Riley Bove, MD, MMSc
Department of Neurology
UCSF Weill Institute for Neurosciences
675 Nelson Rising Lane
San Francisco, CA, 94158
United States
Phone: 1 14155952795
Email: Riley.bove@ucsf.edu

Abstract

Background: Patients with multiple sclerosis (MS) face several challenges in accessing clinical tools to help them monitor, understand, and make meaningful decisions about their disease course. The University of California San Francisco MS BioScreen is a web-based precision medicine tool initially designed to be clinician facing. We aimed to design a second, openly available tool, Open MS BioScreen, that would be accessible, understandable, and actionable by people with MS.

Objective: This study aimed to describe the human-centered design and development approach (inspiration, ideation, and implementation) for creating the Open MS BioScreen platform.

Methods: We planned an iterative and cyclical development process that included stakeholder engagement and iterative feedback from users. Stakeholders included patients with MS along with their caregivers and family members, MS experts, generalist clinicians, industry representatives, and advocacy experts. Users consisted of anyone who wants to track MS measurements over time and access openly available tools for people with MS. Phase I (inspiration) consisted of empathizing with users and defining the problem. We sought to understand the main challenges faced by patients and clinicians and what they would want to see in a web-based app. In phase II (ideation), our multidisciplinary team discussed approaches to capture, display, and make sense of user data. Then, we prototyped a series of mock-ups to solicit feedback from clinicians and people with MS. In phase III (implementation), we incorporated all concepts to test and iterate a minimally viable product. We then gathered feedback through an agile development process. The design and development were cyclical—many times throughout the process, we went back to the drawing board.

Results: This human-centered approach generated an openly available, web-based app through which patients with MS, their clinicians, and their caregivers can access the site and create an account. Users can enter information about their MS (basic level as well as more advanced concepts), visualize their data longitudinally, access a series of algorithms designed to empower them to make decisions about their treatments, and enter data from wearable devices to encourage realistic goal setting about their ambulatory activity. Agile development will allow us to continue to incorporate precision medicine tools, as these are validated in the clinical research arena.

Conclusions: After engaging intended users into the iterative human-centered design of the Open MS BioScreen, we will now monitor the adaptation and dissemination of the tool as we expand its functionality and reach. The insights generated from this

XSL•FO
RenderX

approach can be applied to the development of a number of self-tracking, self-management, and user engagement tools for patients with chronic conditions.

(J Med Internet Res 2020;22(7):e15605) doi:10.2196/15605

KEYWORDS

human-centered design; mobile phone; personal health record; participatory medicine; visualization in eHealth; human factors

Introduction

Background

Delivering actionable clinical tools into the hands of patients and clinicians represents a major unmet need for the delivery of precision medicine for complex diseases such as multiple sclerosis (MS). MS is a chronic, inflammatory, and neurodegenerative disease characterized by onset typically during young adulthood, a protracted and heterogeneous course, variable impairments across multiple functional domains, and variable response to medications [1]. Digital tools help to track this heterogeneity and inform decision making [2] These include digital tools that create research cohorts and registries (eg, iConquerMS [3] and The North American Research Committee on Multiple Sclerosis [4]) and help patients connect with one another and collect and share information (eg, PatientsLikeMe [5], SmartPatients [6], and Facebook groups [7]). In health care settings without a single-payer system or a national electronic health record (EHR), there is a further need for a data collection system to allow patients to collect essential information about their MS course as they navigate multiple providers and EHRs over time. Within this landscape, we aimed to create a freely available platform through which patients can enter basic MS measurements, visualize their course, and access actionable research tools curated by clinicians to inform their decision making.

Open MS BioScreen

In 2013, the University of California, San Francisco (UCSF) Multiple Sclerosis Group announced the *UCSF Multiple Sclerosis BioScreen* [8]. The MS BioScreen is *a data infrastructure platform* that gathers all relevant MS data from different sources, including clinical, imaging, and biomarker information; visually represents the disease course of an individual with MS from a front-end interface; and frames this course within the context of a large cohort of patients treated according to contemporary standards. The goal is to inform more precise clinical decisions and by providing clear information and decision aids, to empower patients to participate more actively in their clinical care. This work received key support from the Patient-Centered Outcomes Research Institute and the Conrad N. Hilton Foundation. However, this first prototype was developed as a clinician-facing, tablet-based tool and hence was not accessible beyond the limited group of patients followed at UCSF. Due to the limitations of tablet storage and connectivity, the MS BioScreen was not used after the pilot phase.

To expand the reach of our precision medicine solution, we aimed to develop the Open MS BioScreen [9]: a publicly available web-based app, free of commercial interest, that allows users within or beyond highly specialized academic care settings the opportunity to enter data on their condition; obtain a richly contextualized, digestible, and actionable predictive output; and participate in a shared decision-making process. We included *open* in the title as a key concept because the platform is freely available for anyone in any location to create a profile and enter their MS measurements, without requiring a fee or any user validation process.

Here, we describe the human-centered design approach that we applied throughout the development process, an approach that focuses on the usability and needs of those the tool is meant to serve in addition to the specific theoretical framework that was applied as an underpinning to the platform content.

Methods

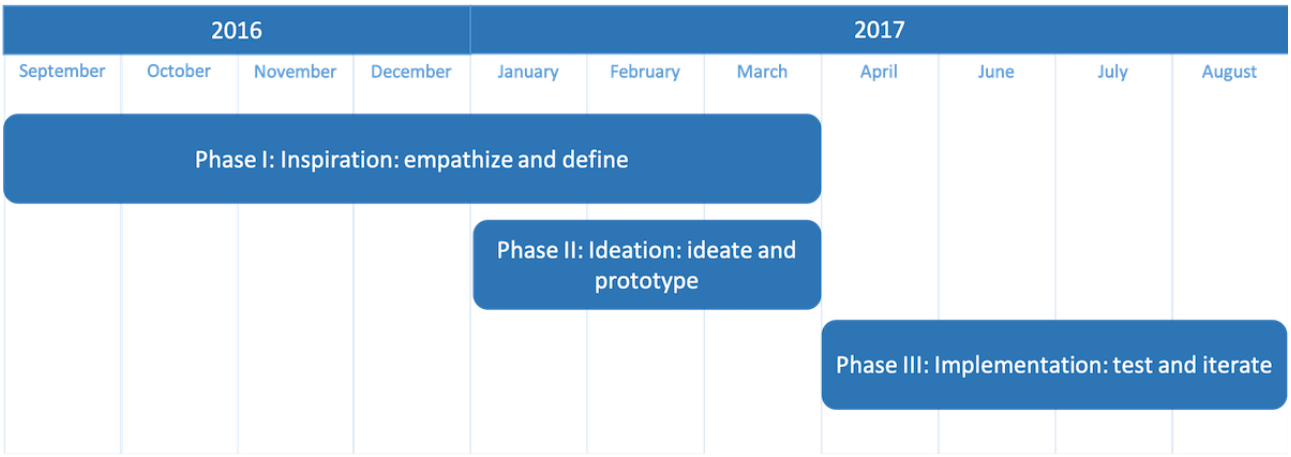
Human-Centered Design Approach

App development of the Open MS BioScreen began in March 2017 and is ongoing. We applied the concept of human-centered design [10,11], an approach that focuses on the usability and needs of those the tool is meant to serve. We included a variety of key stakeholders: individual patients, patient support groups, clinicians, technology consultants, and institutional groups, and we integrated their feedback continuously throughout the development of the app. In the context of chronic disease, the application of human-centered design represents an opportunity to address key underlying provider and patient gaps that obstruct improved health outcomes. As previously described [11], to bridge the knowing and doing gap between evidence-based provider recommendations and patient implementation, solutions must wield individualized and actionable goals that can be addressed within real-world limitations. Without involving key stakeholders in the development, promising technology-leveraged management platforms have difficulty achieving clinical improvement (eg, type 2 diabetes [12] and cardiac rehabilitation [13]). In contrast, tools incorporating human-centered design demonstrate positive effects for specified outcomes across diverse populations [14], including caregivers of patients with Alzheimer disease [15], patients managing chronic obstructive pulmonary disease [16], and in pain management [17].

As outlined in Figure 1, we applied the three phases of human-centered design [11]: (1) inspiration (empathize and define), (2) ideation (ideate and prototype), and (3) implementation (test and iterate).

XSL•FO
RenderX

Figure 1. The human-centered design process for Open MS BioScreen.



Team and Stakeholders

The *Open MS Bioscreen* team *at* UCSF includes a clinical lead (RB), developer (ES), back-end data consultants (AS and AR), technical consultant (AL), radiology consultant (RH), compliance and logistical assistant (WR), designer (AB), students, and research assistants. This group leveraged the expertise of the broader group of UCSF-based MS researchers (the UCSF Expression, Proteomics, Imaging, Clinical [EPIC] longitudinal study [18,19]) as well as team experience with other tools developed through the UCSF MS BioScreen project, namely, Sutter Health MS Share [8] and UCSF BRIDGE [20].

Health Literacy Expert

A health literacy consultant with expertise in chronic conditions, including MS, was engaged to provide individual feedback on solution design and patient interview protocols as well as guide content creation using input gleaned from patient focus groups.

Key Stakeholders

Patients With Multiple Sclerosis

Through convenience sampling, we identified 50 adults with MS who participated either in individual interviews with patients identified by participating clinicians or approached by the study team in the clinic waiting room and at an MS fundraising walk (n=15, often accompanied with a partner, friend, or caregiver) and focus groups (3 groups comprising 6, 8, and 21 people; 2 community-based [National Multiple Sclerosis Society, NMSS, support groups] and 1 UCSF-based). An additional 24 participated in the validation of patient-reported MS duration and treatment and provided individual feedback.

Clinicians

Through convenience sampling, we identified 6 UCSF-based MS consultants, as well as 2 external neurologists (1 MS and 1 general), who were included through phases I and II.

Advocacy Group

Representatives from the NMSS, a primary organization for MS advocacy and research in the United States, were consulted in phase 1 (N=1) and phase 3 (N=3) to advise on the overall goals of the Open MS BioScreen, clarity and relevance of the content to MS patients, and any ethical or other considerations anticipated.

Industry Representatives

This convenience sample of representatives who had approached the MS BioScreen team during the development phases included representatives from 2 biotech companies seeking to develop patient-facing technical solutions for the delivery of MS precision medicine (1 pharmaceutical and 1 biotechnological) as well as 3 experts in the delivery of advanced magnetic resonance imaging (MRI) solutions to patients. The role of input from these representatives was to better delineate the value proposition of an openly accessible, freely available platform within the context of current digital health efforts; however, the platform was not designed to collect data of commercial relevance.

Theoretical Underpinning

Although the design process was iterative and human-centered, we identified a priori several key underpinnings of the platform, namely, Open MS BioScreen should comprise 5 features: (I) profile, (II) measures, (III) patient-reported expanded disability status scale (prEDSS), (IV) course, and (V) decision aids, as outlined in Figure 2. Finally, the user can represent a patient, a patient's proxy (friend, spouse, partner, child, and caregiver), or a clinician [21].

Feature I allows a user to create a profile about themselves. The profile contains required information: month and year of birth and age of onset of MS. In addition, users can add optional information: gender, MS type, city, country, education, employment, race, smoking history, and spinal tap results.

Feature II is used for patient-reported basic MS measures. Users can enter in their standard disability scores (expanded disability status scale, EDSS) [22] results and date of results. Relapse dates can be reported and stored. Additionally, users can enter their MS medication types and dates.

Feature III focuses on 1 specific measure, the prEDSS questionnaire. The prEDSS feature allows a user to answer a series of questions to generate a prEDSS score and visualization and can be used when patients do not have access to a clinician's objective EDSS score.

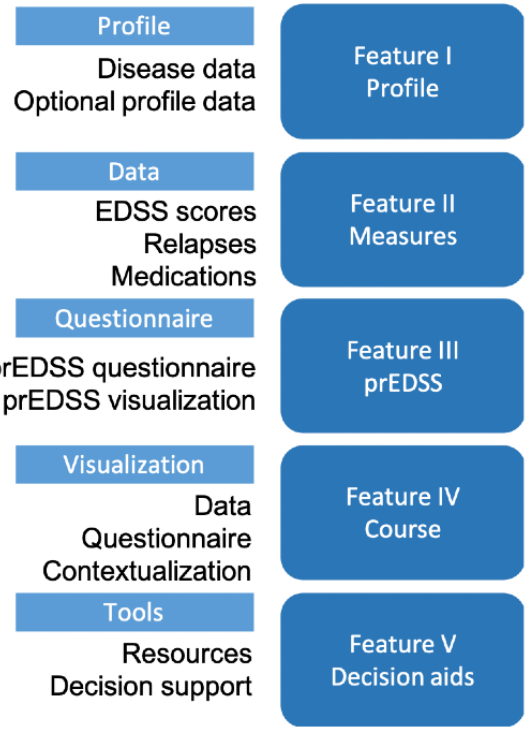
Feature IV is a data visualization component that brings in data from features I-III. The patient course is plotted in a single interactive view, allowing the user to see all their MS

XSL•FO
RenderX

measurements plotted along with their prEDSS scores. Users can optionally display their course in context with our contextualization cohort to see how their MS course compares.

Feature V is a set of many evolving features. Here, we take relevant and new MS research findings and translate them into interactive and actionable features that serve as decision aids regarding if and when to start MS therapies and how to balance the risks and benefits of treatment approaches. We also link it to other tools that could be useful.

Figure 2. Open MS BioScreen features. EDSS: expanded disability status scale; prEDSS: patient-reported expanded disability status scale.



Phase I—Inspiration: Empathize and Define

The goal of this phase was to define the problem we were trying to solve. To accomplish this, between September 2016 and March 2017, one-on-one in-depth in-person interviews were conducted with clinicians, MS patients (and their friends, families, and caregivers), and industry and advocacy experts. Interview transcripts and questionnaire responses were then reviewed and parsed into key *insights* (Tables 1 and 2).

We gathered information from clinicians, patients, partners and through market research.

1. Clinician interviews: We interviewed 1 MS expert (2 hours) and shadowed 5 others in the clinic (one 4-5 hour session each) to observe the interaction and document the information provided. As part of the observation, we talked with the patients to develop an understanding of them, their MS, and how they stay informed. Following the clinical observations, all interviews focused on the tools available to clinicians to monitor their patients (eg, what types of patient-centered data were missing from the clinical EHR?), decision support tools available to them in routine clinical encounters (eg, contextualization or predictive algorithms readily available and would they be useful), and unmet needs (eg, what types of visualizations would help them communicate risk with patients?).
2. Patient interviews: In total, 12 individual patients were interviewed using semistructured questionnaires, developed with input from the health literacy expert. Interviews lasted 30-60 min and asked patients general questions about their MS and their use of tools to understand and track their condition outside of clinical encounters (Multimedia Appendices 1 and 2).
3. Interviews with technology and advocacy partners: These general interviews lasted 30-120 min and were related to the availability of other tools on the market, relative strengths, and unmet needs.
4. Comparative analysis of existing technological solutions: Here, our team researched other technical solutions available to patients with MS and their clinicians to track their own course and access relevant algorithms and visualizations that empowered them to participate in their own care. To accomplish this, we used the Google search engine, the iOS App Store, and the Google Play Store and used the search terms “multiple sclerosis”
or
“MS” and the terms “symptom tracker,” “monitor,” “tracker,” and “algorithm.” From these results, rather than reporting an exhaustive list, we selected tools intended to highlight desired features, possible redundancies/overlap, as well as technological or other pitfalls of existing tools.

XSL•FO
RenderX

Table 1. Patient interview insights.

Insight	Patient input
Clarity	• Does not want to be inundated with information • Wants to reclaim control of their disease—too little clarity, not enough power in their hands • Currently feels like there are no clear answers given, feels alone in trying to understand • Wants information that is easy to understand and translate between centers • Can be medically illiterate at times—feels they have a lot of misconceptions
Data collection and visualization	• Tracks in everything in a notebook—filled with other information • Wants to see scores
Personalized	• Tracking their own progress and treatment efficacy, balanced with their own preferences and needs
Trajectory	• Wants to know if they are getting better or worse • Wants to know 5-year outlook • Constantly playing mind games to stay positive
Comparison	• Would like to see other patient data without communicating with them • Wants to meet other people with the same progression • Wants to gain more perspective about other people with MS[a]
Treatment	• Wants to know that the treatments they are getting are useful or the best for their disease
Resources	• Wants more information about specific areas of interest—genetics, therapies, and education about MS • Wants lots of educated opinions
Time with specialist	• Prefers time spent with doctor to time spent researching on their own • Does not know what to contact their neurologist about • Will not go to the doctor until it is unbearable • Wants a printout from every visit

[a]MS: multiple sclerosis.

XSL•FO
RenderX

Table 2. Comparative analysis of existing technological solutions: illustrative examples.

Tool evaluated	Strengths	Limitations	How strengths and limitations shaped our phase II process
UCSF[a] MS[b] BioScreen [8]	• Integrates clinical, MRI, and biomarker data in 1 coherent view	• iOS app • Not real time; depends on data extracted from study registry or EMR[c]	• Enabled patient-entered data • Web-based platform preferable
Data living in Advancing Patient-Centered Excellence (UCSF Epic-based EMR)	• Clinician-validated • Common EMR system allows for aggregation of patient data across multiple sites and institutions	• As of 2019, no existing algorithms to extract all key data available from a patient's entire MS history (in the United States) • Difficult for the layperson to identify salient metrics	• Build a patient-facing app
MyMSandMe [27]	• Patient engagement resource with an active forum • Medication diary	• No longer supported on iOS or Android platforms	• Web-based platform preferable
myMS [28]	• Connect to 23andMe genetics • Contextualization of personal data • Ability to self-report metrics • Ability to view MRIs[d] in-app • Tasks and questionnaires available in-app	• iOS app • Significant involvement with a private company (23andMe)	• Genetics not yet actionable • Web-based platform preferable • Fewer metrics, more relevant to patient clinical picture • Enable patient-determined EDSS[e]
Specific activity or symptom trackers (eg, Fitbit)	• Ubiquitous • Passive data collection and friendly user interface encourage personalized goal setting and engagement	• Limited lifecycle • Cost • Do not integrate with other clinical data	• Agility—allow inclusion of data derived from these trackers without needing to integrate with any 1 device or vendor
PatientsLikeMe [5]	• Free to use • Widely used	• Heavily funded by pharmaceutical industry • No specific clinical decision support aides	• Remain clinically focused to aid key decision-making points in MS history
Swedish MS registry [29]	• Presents clinically meaningful data in a user-friendly interface • Clinicians/providers contribute validated data to individual profiles	• Not available in the United States	• Longitudinal visualizations
Floodlight [30]	• Validated metrics of MS-related function • Smartphone app	• Data sent to pharmaceutical company • Not available for all smartphone platforms	• In the future, enable integration with wearables • Enable manual entry of MS functional composite metrics
Aby [31]	• Connect with MS experts • Guided PT[f] programs • Apple Health integration	• Data sent to pharmaceutical company	• Links to various sources of clinical expertise
BeCare Link [32]	• Tracks symptoms over time by recording a host of unsupervised tests (25' walk, TUG[g] test, vibratory sensitivity, etc) • Reports proprietary EDSS and MSFC[h]	• Testing can take several hours • No clinical or clinician-validated data from the EMR	• Enable patient-determined EDSS and relapses

[a]UCSF: University of California, San Francisco.

XSL•FO
RenderX

[b]MS: multiple sclerosis.

[c]EMR: electronic medical record.

[d]MRI: magnetic resonance imaging.

[e]EDSS: expanded disability status scale.

[f]PT: physical therapy.

[g]TUG: Timed Up & Go.

[h]MSFC: multiple sclerosis functional composite.

Phase II—Ideation: Ideate and Prototype

At the completion of phase I, in January 2017, and as detailed below in the *Results* section, we defined the problem that we wanted to solve as: "People with MS are faced with a lot of variables, numbers, and changes, and it is hard to keep track of everything." In phase II, our goal was to develop mock-ups. Therefore, we began to come up with solutions to this problem: how best to allow users to enter and visualize all their data and incorporate research findings and additional resources together in a cohesive and actionable way, using tools that are sustainable, low cost, and would not require payment on the part of patients. We met weekly to discuss how best to collect and visualize patient data. Phase II meetings lasted for an hour and focused on determining the user interface for data collection and visualization. Methods for idea generation included brainstorming options, whiteboarding, and searching for ways other people visualized similar problems. Phase II meetings started in January 2017 and lasted until March 2017.

The most promising ideas were moved to the prototype, using tools including Balsamiq [23] for creating wireframes and Invision [24] for creating click-through mock-ups. These wireframes and mock-ups were then shown, in a second round of interviews, to clinicians and patients with MS for further feedback. At this stage, the health literacy expert was re-engaged to specifically explore the pros and cons of various components of the prototype (eg, color range, explanation of disability, and degree of precision regarding predictive models).

These were shown both to individual patients after their clinical appointment (n=4) as well as to an MS support group (n=11) and patients participating in an MS fundraising walk (n=6). Demographic information was not always collected, but the cohort included both ambulatory and nonambulatory patients, patients seen both within our academic institution and from the broader Northern California community of people living with MS, and also explicitly included patients who self-identified as racial and ethnic minorities. We used the same initial questions as in Textboxes 1 and 2 as well as additional questions from Table 1. After each set of interviews, we incorporated patient feedback into the mock-ups.

Textbox 1. Clinician interview insights to include in the app.

Include in the app

- How patients compare with those of similar demographics
- Potential treatments
- Data as a function of time
- Better metrics of progression
- Simple visualizations
- Tool to promote communication
- Clarity about the design process

Textbox 2. Clinician interview insights to exclude from the app.

Exclude from the app

- Information above a high-school level
- Too many numbers or laboratory test data
- Physician feelings and analyses
- Default contextualization (patients should have an option to view how they are doing compared with others)
- Unfiltered scores that patients can misunderstand
- Overly specific information, such as specific risk scores, given the imprecision in data used to generate algorithms

XSL•FO
RenderX

Phase III—Implementation: Test and Iterate

Our goal in phase III was to develop a minimally viable product (MVP). Therefore, once our phase II mock-ups were refined based on user input; we went through an agile development process to build Open MS BioScreen. We met weekly to review progress of the app and determine how to keep in line with feedback from patients and providers. Phase III meetings were conducted from March 2017 to July 2017. Meetings lasted for 1 hour and focused on app demos and feedback. Open MS BioScreen was built in Ruby [25], a programming language, using a Ruby on Rails development framework, jQuery, and d3.js for the front end. In this phase, the tool itself was submitted for approval by the UCSF institutional review board (IRB), an Information Technology security risk assessment, and privacy legal and risk review. Once we had a final product, we again brought the app to an MS support group, to 15 individual patients, and to advocacy group consultants, to gather and incorporate feedback into the app.

To evaluate the accuracy of basic data entered by participants, in pilot testing, we asked 24 patients who were long-time participants in the UCSF EPIC study [18,19] to enroll in Open MS BioScreen and create a basic profile. After signing a written informed consent for an IRB-approved research protocol, participants enrolled and entered their MS duration, data, and treatments. We then compared these entries to study clinician-entered data available in the EPIC study.

Ethics Approvals

Funding for this project was provided by the Conrad N. Hilton Foundation. All procedures performed in studies involving human participants were in accordance with the ethical standards of the UCSF IRB and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Electronic, implied informed consent was obtained from all individual participants included in the study under UCSF IRB no. 17-22732.

Results

Phase I—Inspiration: Empathize and Define

Clinician Interviews

Clinicians provided a range of responses regarding key features that would be provided in a tool that engaged patients with MS, allowed them to track their condition in a manner useful to them, share what they track with clinicians, and use visualizations to become more informed participants in their care process (Textboxes 1 and 2). To summarize clinicians' insights, patient data had to be presented in an understandable format, without excessive details or over-reliance on specific risk ratios or numbers, given the limitations in the data used to generate risk, and the need for visualizations to be inherently meaningful to patients *without* clinician interpretation or input.

Patient Interviews

During this phase, 12 patients were interviewed. To summarize their responses, they articulated a need for clarity, simple tracking and display of key MS data, ability to personalize how they viewed their MS course and what factors went into the contextualization algorithm, and positive emphasis on their trajectory and algorithms. For example, if the contextualization algorithm showed that their MS was worsening at a faster rate than others, then a pop-up might encourage them to return to their physical therapist or discuss their optimal MS management with their clinician. These are detailed in Table 1.

Comparative Analysis of Existing Patient-Facing Tools

The comparative analysis approach allowed us to identify a number of tools with rich information (illustrative examples given in Table 2) and to use their relative strengths and limitations to inform our own subsequent ideation process. Some of these tools evolved since our original phase I in 2016 and have since been summarized elsewhere [2,26].

On the basis of this analysis, we were able to define several key features for our own ideal platform, namely: (1) web-based and platform-agnostic, that is, not dependent on the life cycle of any other individual technology piece; (2) limited imprint from the pharmaceutical industry; (3) reliable over time without requiring daily input; and (4) clinician-derived informational modules.

Phase II—Ideation: Ideate and Prototype

Mock-Ups

The priorities identified in the empathize and define phase were used to develop an initial series of mock-ups and click-through presentations of the app (Figure 3). The longitudinal view of the patient course (middle panel, left image) was designed based on the longitudinal view of the original tablet-based MS BioScreen, whereas other features were new. This prototype was then presented to users for feedback on the design and usability of the Open MS BioScreen.

The upper panel demonstrates various iterations of the landing page, which were designed to allow the user to navigate the site with clarity. The middle panel demonstrates a series of iterations of the longitudinal disease contextualization, which was designed to allow the patient to visualize and interpret their clinical course while minimizing the potential for distress. The bottom panel demonstrates iterations of the patient's global neurological functioning, which was designed to allow patients to enter information about the various functional domains affected by MS (eg, vision, cognition, and walking) and their severity (mild, moderate, or severe), and to visualize these globally. We initially anticipated that patients would want to see their functional domains represented as a sort of avatar, but patients reported that they preferred to have their functional representation alongside, rather than showing their symptoms on a figure. The questions comprising the tool were derived from the prEDSS initially developed by Goodin et al [33], expanded to enable greater sensitivity to changes in cognition and upper limb function over time, and iteratively tested and refined through a series of validation studies (reported separately, manuscript under review). Here, as part of our design work, we specifically iterated visualization of the final prEDSS score but not the questions or scoring.

XSL•FO
RenderX

Figure 3. Phase II: Ideation. Evolution of Open MS BioScreen mock-ups. Top panel: landing page. Middle panel: longitudinal disease course visualization and contextualization tool. Bottom panel: assessment and representation of global neurological function.



Patient Interviews

With the prototypes, we interviewed another 21 patients. Targeted interviews lasted approximately 15 min each and asked patients to reflect on the clarity of the information, impact (educational or emotional) of the information and visualizations provided, ease of navigation and of data entry when appropriate, and missing aspects of the tool. We recorded all feedback, and then, we developed codes for the responses using a grounded theory approach [34]. The authors (RB and JP) read through all transcripts and identified open codes based on line-by-line analysis of emerging themes. The authors then discussed relationships among open codes and created selective codes, or themes, of larger concepts.

Patient insights were primarily clustered around 6 themes:

1. Information: not knowing which information is important
2. Education: feeling like their doctors are not telling them the whole story
3. Data: wanting more tracking options and the ability to see their own data
4. Emotion: expressions of fear and discretion
5. Clarity: wanting clearer labels and explanations of the tool
6. Perception: wanting the tool to present an action-oriented rather than a pessimistic view of MS progression

Phase III—Implementation: Test and Iterate

Once phase II was complete, app development of the MVP occurred between March 2017 and September 2017, incorporating insights from phases I and II (Table 3) and turning these into a web-based prototype (Figure 4). We then conducted an additional round of interviews with our 4 stakeholder groups: patients, clinicians, advocacy, and industry representatives. We compiled all the interviews into actionable insights that we incorporated into the MVP (Table 4). This MVP was launched in February 2018 [9].

XSL•FO
RenderX

Table 3. Insights derived from 21 patients who viewed mock-ups.

Insight	Examples	Quotes
Information: Patients do not know which information is important (18/21, 86%)	• Sharing information across specialties • Understanding their course	• “If I talk to any other doctor, they don’t know about my specific clinical trial.” • “We don’t know which level we're at—so it's important that now that we know how to treat it, now that we have the chance to take better drugs, it's very lucky.”
Education: Patients feel like their doctors are not telling them the whole story (12/21, 57%)	• Knowing when their diagnosis changed • Understanding the landscape	• “I didn't know I was secondary progressive until I decided to read all of the informational packets they give you at the appointment. I said ‘Wait, when did that happen? Nobody told me!’“ • “The judgements about who is getting better and who is getting worse, and which drugs are working—because there's no way to truly compare.”
Data: Patients want more tracking options and ability to see their own data (19/21, 91%)	• Tracking things that are relevant • Comparing their data with other relevant data	• “I want a tool that will benefit real people” • “There's so many aspects—travel, activity, medicines. I want to see that.” • “I'm not curious about other people—usually the information is not relevant because it's so so different for everyone. But this [support group] is a great group, I just wish we were larger.”
Emotion: Patients express fear and discretion (9/21, 43%)	• Wanting to not progress or hear about progression	• “I don’t want to know if it’s getting worse.” • “Sometimes I don't want to know—because I grew up with my father, watching his stages.” • “Loved to read, prayed that her vision wouldn't go even if she would be in a wheelchair ”and the lord granted my wish, I'm in a wheelchair, I can read, but I don't read, really anymore. I lost the concentration of what I was reading—I can't remember it.”
Clarity: Patients want clearer labels and explanations of the tool (6/21, 29%)	• Understanding the parts of the graph • Clarifying the content and labels of the graph	• “How is this graph made?” • “What is this plotting?” • “What do these terms mean?” • “It's unclear because there's a chance I'm in one part of the graph, and a chance I'm in another.” • “Lack of experience, makes it hard to know where to go.”
Perception: Patients want the tool to present a positive view of multiple sclerosis progression (13/21, 62%)	• Tracking full course • Understanding treatments and options	• “I am living with it, I am not suffering.” • “Then they say ‘so and so died from MS,’ and it's like 'Yeah, it's not the MS that kills us, it's the different things that the MS is affecting, but that's only if we sit on our butt and don't do anything.'” • “’You'll never be normal’ But I persevered, and it was important to be able to track all my shots, my own things to look out for—I'll know what the things are that I need to do, so I don't fall off the wagon.” • “We don't know which level we're at—so it's important that now that we know how to treat it, now that we have the chance to take better drugs, it's very lucky.” • “I haven't been seeing anything progressing—but getting better.”

XSL•FO
RenderX

Figure 4. Phase III: Implementation screenshots of Open MS BioScreen's initial minimally viable product. Top panel: landing page and initial data entry tools. Middle panel: the longitudinal multiple sclerosis course visualization tool, without a contextualization tool. Bottom panel: the patient-reported expanded disability status scale tool, whose development occurred as part of Open MS BioScreen and whose formal clinical validation is reported separately (manuscript under review).



XSL•FO
RenderX

Table 4. Examples of actions taken during phase III (implementation) based on patient insights generated in phase 2.

Insights	Actions	Page
Information, Emotion, Clarity	• Better explanation of our tool, its origins, and its cost	• Tagline on first page
Information, Education, Clarity	• Better explanation of how the tool works • Bring down perceived barrier for cognitively impaired patients	• Chevron design on the front page
Information, Data, Clarity	• Easy entry into the tool • Clearer navigation	• *Get Started* button
Information, Education, Clarity, Perception	• Explanation of how to use the tool • Provide guidance to the capabilities of the tool	• Tutorial page
Information, Education, Emotion, Perception	• Clear disclaimer	• Edited headers on the Disclaimer page
Information, Education, Clarity	• Frequently asked questions and glossary • Definitions and explanations page	• Questions and Answers page
Information, Education, Data, Clarity, Perception	• Information about how this tool uses patient data, and where it came from • Better explanation of our mission • Better awareness campaign	• About Us
Information, Education, Data	• Explanation of what is being tracked	• Contextualization introduction page
Information, Education, Data, Clarity	• Simple data entry • Preset options so patients will not have to think of what fields would be useful	• Enter your information page
Information, Education, Data, Emotion, Clarity, Perception	• More intuitive charts • Icons that reflect the words they represent • Clearer marking for parts that are important to bring to clinic • Explain what it means to “progress”	• Clearer design, iterated by patient interactions
Information, Education, Data, Clarity, Perception	• Clarity about the difference between your data and the contextualized data of others • Clarity about how the curves are generated • Careful wording about “similar patients” • Filters for demographics	• Labels at the top of the screen • More explanation about the cohort being compared
Education, Clarity, Emotion	• Education about complex topics—EDSS[a], HIPAA[b] algorithms, and cohorts	• Educational modules
Information, Education, Data, Emotion, Clarity, Perception	• Clear labels and instructions for the patient-reported EDSS section	• Clearer design, iterated by patient interactions
Information, Education, Data, Emotion, Clarity, Perception	• Clearer language • Blue-on-white visualization of symptom severity • Print-friendly interface	• All pages

[a]EDSS: expanded disability status scale.

[b]HIPAA: Health Insurance Portability and Accountability Act.

XSL•FO
RenderX

Comparison of Basic Clinical Data Collected Through Open MS BioScreen Against Study Clinician–Entered Data

In our pilot testing of 24 patients who were EPIC study participants (median EDSS 2, IQR1-4), we compared patient- and study clinician–entered responses for 3 basic clinical data: MS duration, MS type, and MS disease-modifying therapy (DMT). We found that the mean difference between the patient-reported and study-recorded year of MS onset was −0.29 years (SD 3.16). The MS type included 16 relapsing-remitting (RR), 4 secondary progressive (SP), 2 primary progressive, 2 clinically isolated syndrome (CIS); 83% (20/24) agreed on MS type, with EPIC patient-reported mismatches as follows: SP-RR, SP-RR, RR-SP, and CIS-RR. There was also 83% (20/24) concordance on the most recent DMT; in all residual cases, the patient indicated a newer approved DMT relative to the self-injectable recorded in the database.

Focus on Accessibility

We prioritized accessibility in a number of ways. We included *open* in the title as a key concept because the platform is freely available for anyone in any location to create a profile and enter their MS measurements, without requiring a fee or any user validation process. Throughout the design process, we worked with end users to make Open MS BioScreen easy to use for different technical skill levels. Open MS BioScreen is currently an English language app, but users could use a web-based translation to view the site in another language, and patients with cognitive or physical limitations can request assistance from a proxy user.

Added Functionalities

In addition to incorporating user feedback into the MVP, we also built Open MS BioScreen in such a way that we could easily edit existing functionality or add new features. We focused on three initial types of tools, as illustrated in Figure 5.

As an example of a tool that allows patients to record and track their wearable device data in an effort to address the known limitations of existing tools such as the EDSS [36], we built a tool that allows users to manually enter their step count (as reported by any commercially available wearable device) and to compare their count with disability level- and walking speed–matched patients from a research cohort of 100 UCSF patients with MS followed for over 1 year [13,14,35]. Aside from the data type check, there is no data validation done on the Open MS BioScreen side.

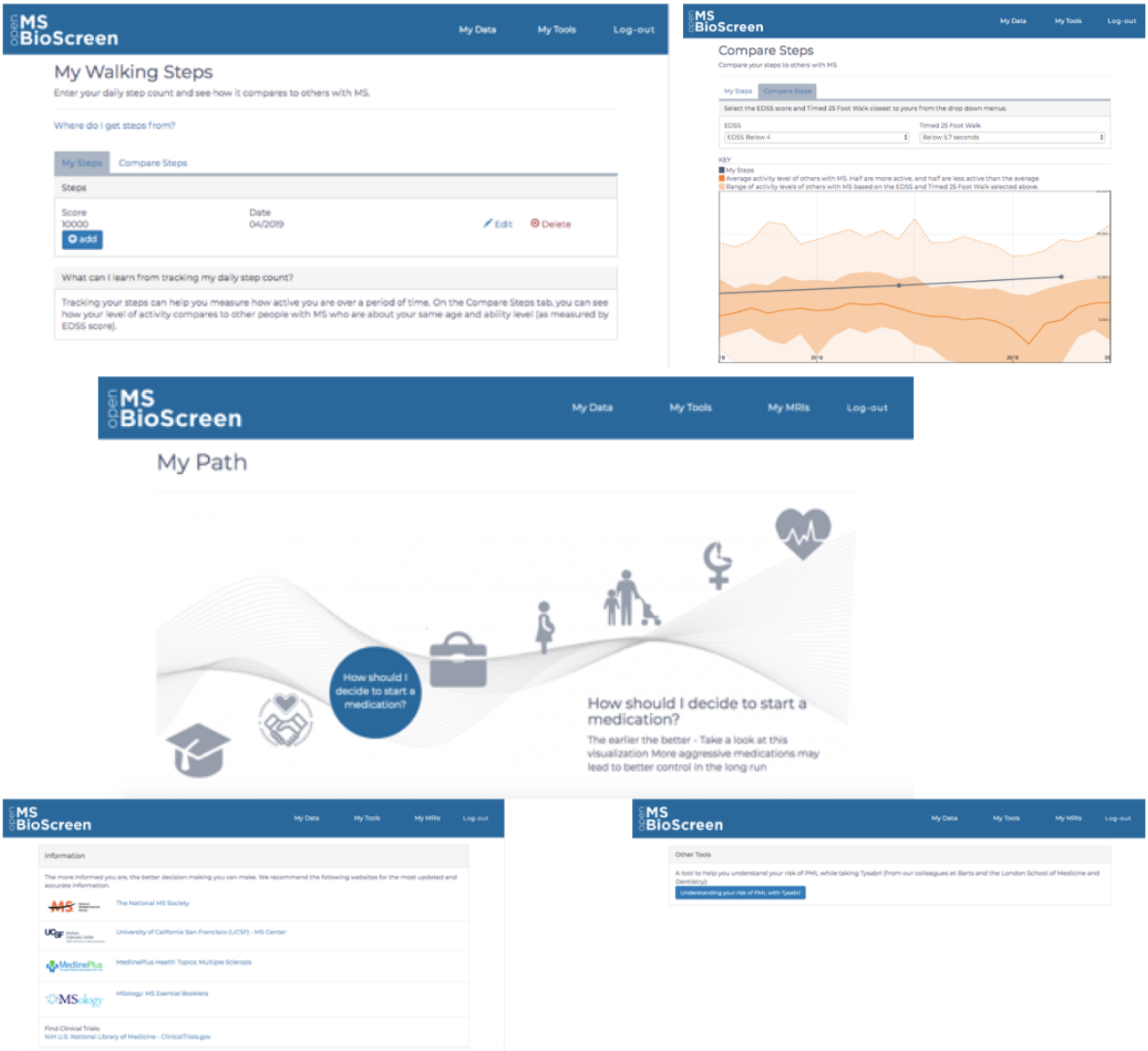
As an example of a tool designed to provide personalized guidance to patients with MS at various points over their life course, we built My MS Paths, a visualization narrating specific decisions and challenges that patients may encounter.

As an example of a clinical decision support tool developed and validated by other investigators but felt to provide significant value to patients with MS, we created a series of links on Open MS BioScreen to sites that allowed patients to calculate their risk, for example, of developing a fatal brain infection, progressive multifocal leukoencephalopathy, while on natalizumab treatment [37].

Again, we worked with end users and our development team to design and build these features. Due to our flexible platform and agile development process, we were able to go live with features as soon as the research was published; this allows patients to have access to actionable research immediately. At the time of manuscript submission, additional features are in development.

XSL•FO
RenderX

Figure 5. Additional features developed after the Open MS BioScreen live data. The top panel depicts the step count comparison tool that allows patients to enter their daily step count and compare this with others from a cohort of patients with multiple sclerosis (MS) at their phase of the disease [35]. The middle panel depicts My MS Paths, a visualization narrating specific decisions and challenges that patients may face over their lifetime living with MS. The bottom panel shows links to algorithms and clinical decision support tools developed by other investigators that provided complementary information and for which links were provided on Open MS BioScreen.



Discussion

Strengths and Limitations

We detail a human-centered approach undertaken to engage stakeholders in the iterative development of an on-demand tool designed to expand patient access to clinically actionable research insights generated by MS researchers and academics. A major limitation of the original MS BioScreen tablet-based app was that it did not take into account many of these principles, and as a result, it was not accessible, agile, or designed with user (patient and clinician) experience in mind. Here, we prioritized the patient and clinician's experience, carefully assessing for similar tools that might be available commercially or not-for-profit, and returned to stakeholders at each step of the development process to ensure that the end result was understandable and actionable. We were able to achieve the 4 main features articulated during our ideation process, features that to our knowledge were not all present in any existing tool reviewed: (1) web-based and platform-agnostic, that is, not dependent on the life cycle of any other individual technology piece; (2) limited imprint from the pharmaceutical industry; (3) reliable over time without requiring daily input; and (4) clinician-derived informational modules.

The current platform provides a number of benefits to patient users. It allows patients to enter and track regular changes in their MS course, including important treatment changes, relapses, and changes in clinical function. As these data are available live in Open MS BioScreen, as opposed to a health system's electronic medical record, patients can easily maintain continuity over time, across providers and health systems. This is beneficial in health care settings such as the United States, where, in contrast to countries such as Sweden that benefit from a central, integrated MS registry [29], patients must navigate a

XSL•FO
RenderX

number of platforms and systems—with the potential for data loss, data duplication, and burdensome navigation of EHRs to synthesize the relevant MS-related information as patients transition between specialists and care settings. Of note, within UCSF, we have built a platform, BRIDGE, that launches directly from the EHR (bridge.ucsf.edu). Open MS BioScreen also allows patients to access narrative visualizations through which they can explore current and next phases in their clinical trajectory with MS, such as how to select an initial DMT, how to interpret and respond to a diagnosis of clinical progression, and how to approach upcoming life transitions (eg, pregnancy, menopause, or healthy aging). It also allows them to develop personalized action plans based on other patients living with MS (rather than the general population), taking into account realistic goal setting for their daily activity levels. The next phase of development is to measure and monitor the uptake of the tool in response to planned dissemination efforts and to continue to fine tune features based on user input.

The current platform also prioritizes certain clinician and researcher needs. First, Clinicians meeting a patient with long-standing MS must often spend valuable encounter time reviewing copious historical and administrative records to extract key MS-related clinical information. Second, Neurologists who are not MS experts also benefit from benchmarking a given patient's current function against the UCSF MS research cohort, a form of *virtual cohort* to inform expectations about patient course. Third, patients who are more informed about treatment decisions, such as weighing risks and benefits of therapies or what to expect at various life stages, are better able to participate in the clinical decision-making process. Over time, the research community will benefit from these data (which patients explicitly consent to be shared in a deidentified manner), to monitor trends in patient-related function, information-seeking, adaptation, and interpretation of data from wearable technology, and eventually MRI analytics. One obvious limitation, for both clinicians and researchers, is that there is no formal validation process for user-entered data, although our pilot testing for 24 participants suggests reasonable agreement between patients and clinicians regarding key MS metrics.

Conclusions

A major recommendation from patients, industry, and advocacy stakeholders was the need to incorporate a greater number of patient-generated data, such as ambulatory activity and MRIs. Unfortunately, the field of commercially available wearable devices is rapidly evolving in terms of both device engineering as well as regulations surrounding devices and data use. Given this landscape, rather than invest substantially in developing Application Programming Interfaces with selected vendors and industry partners (who may or may not retain market share) to synchronize data from wearables, we focused instead on a simpler, more agile MVP. We, therefore, designed a simple, device-agnostic solution that enables patients with MS to enter data about their daily step count obtained from any number of commercial activity trackers and to obtain contextualizations of their ambulatory activity according to their own age and ability level. Our initial platform will allow for agile and responsive development and flexible adaptation of clinically actionable algorithms and tools developed by other research groups. This will be instrumental in ensuring that Open MS BioScreen allows patients to monitor their own course and make informed decisions based on current advances in clinical research.

As with any digital tool, there are a number of sustainability concerns that will be navigated, for which we have created a dissemination and sustainability roadmap. Dissemination and adaptation of digital tools represent a significant next phase in Open MS BioScreen development. Planned analysis of this dissemination plan includes a 2-year analysis of Google Analytics user activity data to track the dissemination and adaptation of this platform by patients living outside the UCSF research space. We will also include direct user feedback through the platform itself as well as through an additional cycle of interviews with users after the completion of the next modules (MRI data). These planned sources of feedback will allow us to expand within the inevitable biases of any specific academic, clinical, or geographic setting. Although we took care throughout our design process to interview a range of patients and clinicians both within and outside the UCSF MS group, there are likely residual biases that will be addressed through this planned feedback process. Financially, beyond the current philanthropy (Conrad N. Hilton Foundation) funding cycle, the dissemination results will allow us to determine if the platform is best maintained within an academic setting or would benefit from the adaptation by a nonprofit partner.

To date, there have been limited descriptions of human-centered design in MS care [2,38-42]. The insights gained from our extensive human-centered design process described here can inform approaches for developing a number of tools that enable tracking, communicating, and eventually improving function in patients with chronic conditions characterized by variable course, functional impairments, and a strong focus on quality of life.

Acknowledgments

This project was generously supported by the Conrad N. Hilton Foundation. The authors would like to thank the patients and UCSF MS and Neuroinflammation Center clinicians who participated in research interviews as well as participants in various Northern California MS support groups and fundraising activities, the entire UCSF EPIC research team, and the 2016-2017 UCSF-University of California, Berkeley Masters in Translational Medicine students who contributed to the Open MS BioScreen design as part of their Capstone Project: Christine Joseph, Zian Liu, Stephanie Nemec, and Arash Nourian.

XSL•FO
RenderX

Conflicts of Interest

JG reports consulting fees from Biogen and Alexion, research support (to UCSF) from Genentech, service contract support (to UCSF) from MedDay, honoraria for editorial work from Dynamed Plus, and personal compensation for medical-legal consulting. RH reports research grants from Roche and MedDay and consulting fees from Roche, Novartis, and Sanofi. SH serves on the board of trustees for Neurona and on scientific advisory boards for Alector, Annexon, Bionure, Molecular Stethoscope, and Symbiotix and has received travel reimbursement and writing assistance from F Hoffmann-La Roche Ltd for CD20-related meetings and presentations. RB has received research support from the National Multiple Sclerosis Society, the Conrad N. Hilton Foundation, the California Initiative to Advance Precision Medicine, the Sherak Foundation, and Akili Interactive. RB has also received personal compensation for consulting from Novartis, Sanofi Genzyme, Roche Genentech, and Pear Therapeutics. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Patient interview questions used in phase I.
[DOCX File , 14 KB - jmir_v22i7e15605_app1.docx]

Multimedia Appendix 2
Patient interview questions used in phase II.
[DOCX File , 13 KB - jmir_v22i7e15605_app2.docx]

References

1. Hauser SL, Goodin D. Multiple sclerosis and other demyelinating diseases. In: Jameson JL, Fauci A, Hauser S, Longo D, Kasper D, Loscalzo J, editors. Harrison's Principles of Internal Medicine. New York, USA: McGraw-Hill Education; 2012.
2. Giunti G, Fernández EG, Zubiete ED, Romero OR. Supply and demand in mhealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
3. McBurney R, Zhao Y, Loud S, Balasubramanian R, Schmidt H, Kolaczkowski L. Initial Characterization of Participants in the iConquerMS Network. In: Proceedings of the Americas Committee for Treatment and Research in Multiple Sclerosis. 2017 Presented at: ACTRIMS'17; February 23-25, 2017; Orlando, FL URL: https://actrims.confex.com/actrims/2017/meetingapp.cgi/Paper/1914
4. Vollmer TL, Ni W, Stanton S, Hadjimichael O. The NARCOMS patient registry: a resource for investigators. Int J MS Care 1999 Sep;1(1):28-34. [doi: 10.7224/1537-2073-1.1.28]
5. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
6. Hill E. Smart patients. Lancet Oncol 2014 Feb;15(2):140-141 [FREE Full text] [doi: 10.1016/S1470-2045(14)70044-0]
7. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar;20(1):50-58. [doi: 10.1177/1460458213476969] [Medline: 24550564]
8. Gourraud P, Henry RG, Cree BA, Crane JC, Lizee A, Olson MP, et al. Precision medicine in chronic disease management: the multiple sclerosis BioScreen. Ann Neurol 2014 Nov;76(5):633-642 [FREE Full text] [doi: 10.1002/ana.24282] [Medline: 25263997]
9. Open MS BioScreen. Welcome to Open MS BioScreen URL: https://openmsbioscreen.ucsf.edu [accessed 2019-07-25]
10. IDEO. 2015. Design Kit: The Human-Centered Design Toolkit URL: https://www.ideo.com/post/design-kit [accessed 2019-07-24]
11. Matheson GO, Pacione C, Shultz RK, Klügl M. Leveraging human-centered design in chronic disease prevention. Am J Prev Med 2015 Apr;48(4):472-479. [doi: 10.1016/j.amepre.2014.10.014] [Medline: 25700655]
12. Yu CH, Parsons JA, Mamdani M, Lebovic G, Hall S, Newton D, et al. A web-based intervention to support self-management of patients with type 2 diabetes mellitus: effect on self-efficacy, self-care and diabetes distress. BMC Med Inform Decis Mak 2014 Dec 14;14:117 [FREE Full text] [doi: 10.1186/s12911-014-0117-3] [Medline: 25495847]
13. Karmali KN, Davies P, Taylor F, Beswick A, Martin N, Ebrahim S. Promoting patient uptake and adherence in cardiac rehabilitation. Cochrane Database Syst Rev 2014 Jun 25(6):CD007131. [doi: 10.1002/14651858.CD007131.pub3] [Medline: 24963623]
14. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
15. Cristancho-Lacroix V, Moulin F, Wrobel J, Batrancourt B, Plichart M, de Rotrou J, et al. A web-based program for informal caregivers of persons with Alzheimer's disease: an iterative user-centered design. JMIR Res Protoc 2014 Sep 15;3(3):e46 [FREE Full text] [doi: 10.2196/resprot.3607] [Medline: 25263541]
16. Velardo C, Shah SA, Gibson O, Clifford G, Heneghan C, Rutter H, EDGE COPD Team. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19 [FREE Full text] [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]

XSL•FO
RenderX

17. Trail-Mahan T, Heisler S, Katica M. Quality improvement project to improve patient satisfaction with pain management: using human-centered design. J Nurs Care Qual 2016;31(2):105-12; quiz 113. [doi: 10.1097/NCQ.0000000000000161] [Medline: 26447343]
18. University of California, San Francisco MS-EPIC Team, Cree BA, Gourraud P, Oksenberg JR, Bevan C, Crabtree-Hartman E, et al. Long-term evolution of multiple sclerosis disability in the treatment era. Ann Neurol 2016 Oct;80(4):499-510 [FREE Full text] [doi: 10.1002/ana.24747] [Medline: 27464262]
19. University of California San Francisco. UCSF EPIC Study URL: https://epicstudy.ucsf.edu/ [accessed 2019-07-25]
20. Pletcher MR, Robinson A, Najafi N, Patel S, Franzon D, Bove R. How Do We Deliver AI in Real Time to Clinicians? In: Proceedings of the UC-Wide AI in Biomedicine Conference. 2019 Presented at: UCAI'19; June 17-18, 2019; California, USA URL: https://www.slideshare.net/lesliey/delivering-ai-in-real-time-to-clinicians
21. Shah SG, Robinson I. Medical device technologies: who is the user? Int J Healthc Tech Manag 2008;9(2):181. [doi: 10.1504/ijhtm.2008.017372]
22. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
23. Balsamiq. Rapid, Effective and Fun Wireframing Software. 2019. URL: https://balsamiq.com [accessed 2019-07-25]
24. InVision | Digital Product Design, Workflow & Collaboration. 2019. URL: https://www.invisionapp.com [accessed 2019-07-25]
25. Ruby. 2019. URL: https://www.ruby-lang.org/en/ [accessed 2019-07-25]
26. Stoll S, Litchman T, Wesley S, Litchman C. Multiple sclerosis apps: the dawn of a new era: a comprehensive review. Neurology 2019 Apr 16;92(15 Supplement):- [FREE Full text]
27. Irody I. AppAdvice. 2019. MyMSandMe URL: https://appadvice.com/app/mymsandme/459683919 [accessed 2020-04-20]
28. Pelletier D, Rennert T, Chang M, Decunto S, Cen S. myMS; a comprehensive patient-centered mobile app to monitor MS at home (P3.2-003). Neurology 2019 Apr 9;92(15 Supplement):- [FREE Full text]
29. Hillert J, Stawiarz L. The Swedish MS registry–clinical support tool and scientific resource. Acta Neurol Scand 2015;132(199):11-19 [FREE Full text] [doi: 10.1111/ane.12425] [Medline: 26046553]
30. Genentech I. Floodlight Open. Understanding MS Together. 2019. URL: https://floodlightopen.com [accessed 2019-07-25]
31. The Aby App. 2019. URL: https://aby-app.ca [accessed 2019-07-25]
32. BeCareLink: Connecting Data, Medicine & Technology. 2019. Multiple Sclerosis URL: https://becarelink.com/our-products/multiple-sclerosis/ [accessed 2019-07-25]
33. Goodin DS. A questionnaire to assess neurological impairment in multiple sclerosis. Mult Scler 1998 Oct;4(5):444-451. [doi: 10.1177/135245859800400508] [Medline: 9839306]
34. Glaser BG, Strauss AL. The Discovery of Grounded Theory. Abingdon, UK: Routledge; 2017.
35. Block VJ, Bove R, Zhao C, Garcha P, Graves J, Romeo AR, et al. Association of continuous assessment of step count by remote monitoring with disability progression among adults with multiple sclerosis. JAMA Netw Open 2019 Mar 1;2(3):e190570 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0570] [Medline: 30874777]
36. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007 Dec;6(12):1094-1105. [doi: 10.1016/S1474-4422(07)70290-9] [Medline: 18031706]
37. Major EO, Yousry TA, Clifford DB. Pathogenesis of progressive multifocal leukoencephalopathy and risks associated with treatments for multiple sclerosis: a decade of lessons learned. Lancet Neurol 2018 May;17(5):467-480. [doi: 10.1016/S1474-4422(18)30040-1] [Medline: 29656742]
38. Berry AB, Butler KA, Harrington C, Braxton MO, Walker AJ, Pete N, et al. Using conceptual work products of health care to design health IT. J Biomed Inform 2016 Feb;59:15-30 [FREE Full text] [doi: 10.1016/j.jbi.2015.10.014] [Medline: 26528606]
39. Giunti G, Kool J, Romero OR, Zubiete ED. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 9;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
40. Giunti G, Mylonopoulou V, Romero OD. More stamina, a gamified mhealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 2;6(3):e51 [FREE Full text] [doi: 10.2196/mhealth.9437] [Medline: 29500159]
41. Thirumalai M, Rimmer JH, Johnson G, Wilroy J, Young H, Mehta T, et al. TEAMS (tele-exercise and multiple sclerosis), a tailored telerehabilitation mhealth app: participant-centered development and usability study. JMIR Mhealth Uhealth 2018 May 24;6(5):e10181 [FREE Full text] [doi: 10.2196/10181] [Medline: 29798832]
42. Thomas S, Pulman A, Thomas P, Collard S, Jiang N, Dogan H, et al. Digitizing a face-to-face group fatigue management program: exploring the views of people with multiple sclerosis and health care professionals via consultation groups and interviews. JMIR Form Res 2019 May 22;3(2):e10951 [FREE Full text] [doi: 10.2196/10951] [Medline: 31120021]

XSL•FO
RenderX

Abbreviations

CIS: clinically isolated syndrome
DMT: disease-modifying therapy
EDSS: expanded disability status scale
EHR: electronic health record
EPIC: expression, proteomics, imaging, clinical longitudinal study
IRB: institutional review board
MRI: magnetic resonance imaging
MS: multiple sclerosis
MVP: minimally viable product
NMSS: National Multiple Sclerosis Society
prEDSS: patient-reported expanded disability status scale
RR: relapsing-remitting (multiple sclerosis)
SP: secondary progressive (multiple sclerosis)
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 25.07.19; peer-reviewed by P Wicks, R Haase, SGS Shah, F Lanfranchi; comments to author 21.10.19; revised version received 16.12.19; accepted 04.02.20; published 06.07.20.

Please cite as:
Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, Block VJ, Santaniello A, Renschen A, Gomez R, Keshavan A, Gelfand JM, Henry RG, Hauser SL, Bove R
A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open MS BioScreen): Human-Centered Design and Development
J Med Internet Res 2020;22(7):e15605
URL: https://www.jmir.org/2020/7/e15605
doi:10.2196/15605
PMID:32628124

©Erica Schleimer, Jennifer Pearce, Andrew Barnecut, William Rowles, Antoine Lizee, Arno Klein, Valerie J Block, Adam Santaniello, Adam Renschen, Refujia Gomez, Anisha Keshavan, Jeffrey M Gelfand, Roland G Henry, Stephen L Hauser, Riley Bove. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Promise of Patient Portals for Individuals Living With Chronic Illness: Qualitative Study Identifying Pathways of Patient Engagement

Maureen T Stewart[1,2], PhD; Timothy P Hogan[2,3], PhD; Jeff Nicklas[2], MS; Stephanie A Robinson[2,4], PhD; Carolyn M Purington[2], MPH; Christopher J Miller[5,6], PhD; Varsha G Vimalananda[2,4], MD; Samantha L Connolly[5,6], PhD; Hill L Wolfe[2,7], MPA; Kim M Nazi[8], PhD; Dane Netherton[2,9], PhD; Stephanie L Shimada[2,7,10], PhD

[1]Institute for Behavioral Health, The Heller School for Social Policy and Management, Brandeis University, Waltham, MA, United States

[2]Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[3]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States

[4]Department of Medicine, Boston University School of Medicine, Boston, MA, United States

[5]Center for Healthcare Organization and Implementation Research, Boston VA Healthcare System, Boston, MA, United States

[6]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[7]Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA, United States

[8]Independent Consultant, Coxsackie, NY, United States

[9]Division of Biostatistics and Health Services Research, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

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

Corresponding Author:
Stephanie L Shimada, PhD
Center for Healthcare Organization and Implementation Research
Edith Nourse Rogers Memorial Veterans Hospital
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 7816872208
Email: stephanie.shimada@va.gov

Abstract

Background: Patients play a critical role in managing their health, especially in the context of chronic conditions like diabetes. Electronic patient portals have been identified as a potential means to improve patient engagement; that is, patients' involvement in their care. However, little is known about the pathways through which portals may help patients engage in their care.

Objective: Our objective is to understand how an electronic patient portal facilitates patient engagement among individuals with diabetes.

Methods: This qualitative study employed semistructured telephone interviews of 40 patients living with diabetes since at least 2011, who had experienced uncontrolled diabetes, and had used secure messaging through a portal at least 4 times over 18 months. The interviews were recorded, transcribed, coded, and analyzed using primarily an inductive approach to identify how patients living with diabetes use an online health portal to support diabetes self-management.

Results: Overall, patients who used the portal reported feeling engaged in their health care. We identified four pathways by which the portal facilitates patient engagement and some challenges. The portal provides a platform that patients use to (1) better understand their health by asking questions about new symptoms, notes, or labs, (2) prepare for medical appointments by reviewing labs and notes, (3) coordinate care between VA (Veterans Affairs) and non-VA health care teams, and (4) reach out to providers to request help between visits. Several patients reported that the portal helped improve the patient-provider relationship; however, aspects of the portal design may hinder engagement for others. Patients reported challenges with both secure messaging and access to medical records that had negative impacts on their engagement. Benefits for patient engagement were described by many types of portal users with varying degrees of diabetes control.

XSL•FO
RenderX

Conclusions: Patient portals support engagement by facilitating patient access to their health information and by facilitating patient-provider communication. Portals can help a wide range of users engage with their care.

(J Med Internet Res 2020;22(7):e17744) doi:10.2196/17744

KEYWORDS

patient engagement; patient portal; chronic disease; diabetes; secure messaging; veterans

Introduction

Patients play a crucial role in managing their health, and that role is even more central in the context of chronic conditions that require ongoing coping skills and self-management efforts [1,2]. Broadly, the term "patient engagement" refers to patients taking part in or actively participating in their care [3], including how patients work with health care providers and systems to manage and improve their health. Hibbard [4] conceptualized patient engagement as degrees of active involvement in their care, termed patient activation. Patient engagement can operate at multiple levels and exists on a continuum [5]. We focus primarily on the level of direct patient care where engagement can range from patients receiving information about a diagnosis to patients participating in treatment plan decisions [5]. Although it has been characterized in a variety of ways, a common thread is that patient engagement is a multi-faceted concept incorporating actions on multiple levels to promote patient-centered care and self-management.

Patient-facing eHealth technologies are often described as a means to improve patient engagement [6,7]. Online patient portals have received particular attention since their features often support healthcare system transactions (eg, prescription refills, scheduling), healthcare team communication, access to patient medical records, and the delivery of health education resources [7,8]. Previous studies of patients living with diabetes, for example, suggest that individuals who use online patient portals are more likely to achieve desirable outcomes, including improved glycemic control [9-12].

Few studies have attempted to specify how engagement may be realized through patient portals or the pathways through which portals may help patients engage in their care and self-management. One study of direct-release of test results via patient portals found that this access to information improved engagement by helping patients to monitor results over time and prepare before communicating with their provider [13]. However, a greater understanding of pathways through which a range of portal features support engagement is needed.

A framework for patient engagement by Barello and colleagues may help us understand the different ways in which a patient portal may support patient engagement [14]. This framework incorporates patients' actions, thoughts, and feelings and considers patient engagement within three dimensions: emotional, cognitive, and behavioral [14]. In this model, the emotional dimension of engagement encompasses the patients' emotional state and feelings about their care. The cognitive dimension captures patients' understanding of their disease and treatment. The behavioral dimension of engagement captures patients' actions to address their health.

In this paper, we examine whether and how an online patient portal supports patient engagement for individuals living with diabetes. We take a qualitative approach to this by reviewing patient portal experiences among veterans living with diabetes and describe several ways in which an electronic patient portal facilitates patient engagement.

Methods

Data were collected in a more extensive study examining the role of patient portals in diabetes management, described previously [15]. Briefly, from November 2017 to January 2018, 1200 patients living with diabetes were invited to participate in a mailed survey about their diabetes self-management and use of the United States Department of Veterans Affairs (VA) online patient portal, My Health*e*Vet [16]. The My Health*e*Vet portal offers several features to help patients manage their care. Patients can request medication refills, schedule appointments, receive appointment reminders, communicate electronically with providers through secure messaging, and use the Blue Button feature to access portions of their medical records including labs and clinical notes [17].

All 1200 patients had a diabetes diagnosis in their record since at least 2011, had experienced uncontrolled diabetes in 2012 (mean HbA_{1c} >8.0% and less than 25% of the year with an HbA_{1c} <8.0%) and were actively engaged patient portal users who used secure messaging at least four times between January 2016 and June 2017. Since 2012, half had achieved glycemic control, defined as having mean HbA_{1c} <8% in 2016 and HbA_{1c} <8% for at least 75% of 2016, and half had not. We stratified each group based on urban/rural residence, according to the Rural-Urban Commuting Areas [18], and the presence of comorbid mental health diagnoses. Individuals were considered to have a comorbid mental health diagnosis if they had two outpatient diagnoses or one inpatient diagnosis within each diagnostic group. The diagnostic groups selected using International Classification of Diseases (ICD), 9th and 10th Revision (ICD-9 and ICD-10) codes were anxiety, bipolar, depression, personality disorder, post-traumatic stress disorder, schizophrenia, and substance use disorders. We received 449 completed surveys, of which 350 patients responded that they would be interested in participating in an interview about how they managed their diabetes. This study was approved by the Edith Nourse Rogers Memorial Veterans Hospital Institutional Review Board.

Purposeful sampling was used to identify and select patient portal users for telephone interviews. In total, 160 respondents provided positive responses to an open-ended survey question, "Can you tell us about an 'A-Ha!' moment when you realized you could use the My Health*e*Vet portal to better manage your

XSL•FO
RenderX

diabetes?" We selected interviewees to represent a variety of responses to this and other survey items about My Health*e*Vet use. Interviewees were selected to represent those who used a variety of My Health*e*Vet portal features, those with controlled and uncontrolled diabetes, urban and rural patients, and those with and without comorbid mental health diagnoses. Women and minority veterans were oversampled to broaden the representation of patient demographics.

Forty telephone interviews were completed between February and May of 2018. Each interview was conducted by two researchers and averaged 70 minutes long. The interview covered how the patient managed their diabetes, including their feelings about the efficacy of their self-management, how they gathered and used information, how portal features and other technologies supported their diabetes management, and their suggestions for improving the portal. Interviewers were intentionally blinded to the patient's glucose control at the time of the interview.

Participants

The demographic and health characteristics of our patient sample, obtained from the survey data and health records, are reported in Table 1. The mean age of participants was 65.9 years. The majority were white (85%) and male (80%); 60% had a mental health diagnosis. Participants' mean HbA_{1c} tested closest to the interview (recent HbA_{1c}) was 8.2% (SD 1.4%), over our threshold for diabetes control (<8.0).

Approach

All interviews were recorded, transcribed, and double-coded for multiple themes, using a coding scheme developed by the team. Several coders collaborated to establish the list of codes using both deductive and inductive thematic coding [23]. Deductive coding was used initially to create a list of preliminary codes from the topic areas of our interview guide. These codes included the benefits and challenges of each specific portal feature. Inductive codes were developed as coders reviewed the narrative, and new themes emerged from the interview transcripts. Coders met regularly to discuss the codes and themes, and to ensure consensus on meaning was achieved. Coding discrepancies were resolved with discussion among three authors (MTS, SLS, TPH). Coders were blinded to the interviewee's glucose control status.

We selected eight codes related to portal use and patient engagement for in-depth analysis and reporting here. The following inductively developed codes were used to identify pathways for engagement: (1) patient-team relationship (portal use impact on the patient-healthcare team relationship); (2) empowerment (patients feeling empowered through portal use); (3) care collaboration (patients using the portal to coordinate care with their healthcare teams); (4) impact on care plan (how portal use changes patients' care plans between visits); (5) clarification (patient-initiated communication through the portal for explanations of information). The following deductively developed codes were used to identify challenges to patient engagement: (1) secure messaging challenges, (2) medication refill challenges, (3) BlueButton challenges. The interview text excerpts identified by these codes were analyzed, and themes around how the patient portal facilitates engagement were elicited and used to organize our results.

XSL•FO
RenderX

Table 1. Sample characteristics (N=40).

Characteristic	Value
Diabetes Self-Management Questionnaire (DSMQ) Score[a], mean (SD)	8.0 (0.79)
Mean diabetes self-efficacy score (DSES)[b], mean (SD)	7.4 (1.6)
Recent HbA_{1c}[c], mean (SD)	8.2 (1.4)
Nosos risk score[d], mean (SD)	2.4 (2.7)
Age (years), mean (SD)	65.9 (6.5)
Male, n (%)	32 (80)
Race/Ethnicity, n (%)	
White	33 (85)
Black	5 (13)
Latino, n (%)	2 (5)
Rural, n (%)	21 (53)
Health information, n (%)	
Has a mental health diagnosis	24 (60)
Recent HbA_{1c} in control (<8.0%)	22 (55)
Health literacy[e], n (%)	
Inadequate	2 (5)
Marginal	4 (10)
Adequate	34 (85)
Income ($US), n (%)	
<$25,000	9 (23)
$25,000-$49,000	10 (25)
$50,000-$149,000	15 (38)
>$150,000	3 (8)
Income not reported	3 (8)

[a]Diabetes Self-Management Questionnaire (DSMQ): a global measure of diabetes self-management comprised of 16 items to assess activities related to glycemic control in patients with diabetes. Scaled scores range from 0-10 and higher values indicate more effective self-management [19].

[b]Diabetes Self-efficacy Scale (DSES): measures how confident patients are in their ability to do certain activities related to managing their diabetes. Scores range from 1-10 and higher values indicate higher self-efficacy [20].

[c]HbA_{1c}: glycated hemoglobin.

[d]Nosos risk score: VA's modified version of Medicare's Hierarchical Condition Category. A measure of expected health care costs based on demographic, pharmacy, psychiatric and health care utilization data; mean for a population equals 1.0 and scores >1.0 indicate the patient is expected to have health care costs that much higher than the average VA patient [21].

[e]Health literacy: determined by the patient's response to the question, "How often do you have someone help you read hospital materials? [22]" Health literacy was considered inadequate when patients responded "Often" or "Always"; "Sometimes" was considered marginal health literacy; "Never" was considered adequate health literacy.

Results

Across all 40 patient interviews, 30 touched on concepts related to patient engagement through portal use. Overall, patients reported feeling engaged by their use of the portal. We describe this below, followed by illustrations of the four pathways of interaction with the portal that support patients' feelings of engagement. Finally, we share patient insight on how portal functionality may hinder engagement. To further contextualize our data, we describe select patient characteristics after each illustrative quote.

Feeling Engaged by Use of the Portal

Patients explained that the portal helped to support their engagement with their care and improve their health by having a provider available to give consistent feedback. One patient reported that interacting with providers through the portal improved their attitude and health:

> *The best thing I ever did was when I enrolled in (My HealtheVet). It helped me be in better control of my attitude, my depression, my diabetes. When I deviate a little bit left or right, there's always somebody on*

XSL•FO
RenderX

> *the other end going, 'uh oh, you need to go down that straight and narrow path again.' It's a great program. (60-year-old Latino male, white, urban, mental health diagnosis, recent HbA_{1c} 7.3, in control)*

Patients reported feeling that interacting with providers through the portal helped build patient-provider relationships and that these relationships were vital to receiving better care:

> *The more interest you show, the more interest they show in you. I think the secure messaging helps you establish that kind of relationship. Like, 'Oh here's a guy that's trying to take care of himself, let's help him. (82-year-old white male, rural, no mental health diagnosis, recent HbA_{1c} 7.5, in control)*

How the Portal Supported Patient Feelings of Engagement

Our analysis identified four key pathways by which the portal supported patients' engagement in their care and some difficulties with the portal that may hinder engagement. The portal provides support for patients to (1) work to better understand their health, (2) prepare for medical appointments, (3) coordinate care and share health information between VA and non-VA health care teams, and (4) reach out to providers to request help between visits. These pathways are illustrated below.

Working to Better Understand Their Health

Reading clinical notes and test results helps patients understand their health information in their own time, gauge the seriousness of health issues they have, keep things in perspective, and make decisions about appropriate next steps. For example, patients reported that reviewing clinical notes using the Blue Button portal feature after an appointment helped them to understand what their provider was telling them. "It doesn't always sink in right away what they are telling you. So I've used the Blue Button notes." (69-year-old white male, rural, no mental health diagnosis, recent HbA_{1c} 7.5, in control)

Making test results available through the patient portal provides an opportunity for patients to be proactive based on test results:

> *I think that was the "A-ha" moment when I said, 'now I can see my results of 50 different tests'. And all at once and go over them and pick out the ones that are too high or too low. Then if I see something way off, I can make an appointment. (72-year-old black male, urban, no mental health diagnosis, recent HbA_{1c} 6.9, in control)*

Sharing health information through patient portals can also cause some distress when patients do not understand the information. One participant described a situation like this, but said they use the secure messaging feature to obtain more information and reassurance from their provider:

> *I've had issues where I [view a test result] and it sounds really bad but the note says it's no big deal so I would send a secure message to my doctor and ask them why they're not concerned about this level being high, and she would explain it more to me. So that's helped. (51-year-old black female, urban, mental health diagnosis, recent HbA_{1c} 9.2, not in control)*

Patients reported finding the portal helpful because they could ask questions before they forget. The portal helps reduce the chance of a patient forgetting a health concern or inquiry by enabling them to send providers questions via secure messaging right when patients think of them:

> *Sometimes you can't get an appointment for a week or so and by then you've forgotten what question you have! So it's so much easier just writing it down in that secure message and sending it off before you forget what you want to ask. (61-year-old white female, urban, no mental health diagnosis, recent HbA_{1c} 7.0, in control)*

Patients also used the information available through the portal together with information available from other sources to better understand their conditions:

> *There've been times when I look at my blood work and I see something I don't understand; medical jargon. I've been able to Google it and find out what it means. That's empowering. (51-year-old black female, urban, mental health diagnosis, recent HbA_{1c} 9.2, not in control)*

Preparing for Medical Appointments

Patients described reviewing notes to prepare for their visit to be proactive about their care. One patient reported using the patient portal to prepare for their appointment so that they would be able to ask thoughtful and helpful questions:

> *I'm always checking on my lab results. I ask the doctor if something's high or low because I know the results before I go in and see her. That's what I like about [My HealtheVet], I know the lab results, I know what questions to ask. (72-year-old black male, urban, no mental health diagnosis, recent HbA_{1c} 6.9, in control)*

Another patient described using the portal to review information and vocabulary to engage more meaningfully during appointments in the conversations with their provider:

> *It's easier to sit there and look at [My HealtheVet] and have an idea of what's going on before you talk to the people who know more about it than you do… So that I have an understanding of what terms that they're using. So that they aren't snowballing me or going over my head. (69-year-old white male, rural, no mental health diagnosis, recent HbA_{1c} 7.5, in control)*

Coordinating Care and Sharing Health Information Between VA and Non-VA Health Providers

Having access to the patient portal puts control over health information into the hands of patients. When patients want to share information between providers, patients can send information themselves. Patients reported sharing lab results between VA and non-VA providers to avoid duplicate labs and

XSL•FO
RenderX

unnecessary testing. This access saves time and health care resources.

> *With the VA I usually get labs done every 2 months, give or take. Privately it's probably about the same, every 2-3 months. It goes back and forth. That's the nice thing about this. Go from private to the VA. VA to private. You can take that information and if the doctors need it or want it, you can transfer it to them. (71-year-old white male, rural, mental health diagnosis, recent HbA_{1c} 8.2, not in control)*

Patients also reported using the portal to coordinate care between VA and non-VA providers because the providers do not coordinate themselves. When patients have to take on the care coordination role, having information in writing from each doctor supports accurate sharing of information between providers:

> *I have a VA doctor and I have one through Medicare, I've got to be very careful that I have both of them in agreement. So what I end up doing is if one of them recommends a different type of medication … I can email both of them … they don't seem to want to talk, for whatever reason they don't talk on the phone. So I will literally copy and send emails back and forth with the pros and cons … I have something in writing that I can actually communicate back, rather than me trying to remember. (68-year-old white male, urban no mental health diagnosis, recent HbA_{1c} 7.8, in control)*

Reaching Out to Providers to Request Help or Changes to Care Plan Between Visits

Close collaboration with providers may facilitate improved health and reduce the amount of in-person medical visits, which may be especially burdensome for rural patients. Patients reported using the portal to reach out to providers for information about how to address symptoms and manage their blood sugar levels. Patients described they found secure messaging supportive as a tool to reach out to their providers to understand why their blood sugar is too low or too high and facilitates their ability to receive guidance about what to do differently. Patients use secure messaging to help them as they work to manage their health and understand why they may have certain symptoms,

> *I sent a question to my care team [through secure messaging] asking about what I'm doing wrong because I'm ending up with morning [blood sugar] numbers that are too low…I like that because they're usually pretty prompt at getting back to me. (59-year-old white male, rural, no mental health diagnosis, recent HbA_{1c} 6.5, in control)*

Variations of this comment were heard from many patients:

> *If I thought my [A_{1c}] was out of whack or … that I think it's high, then I might say something. Especially when I don't understand why it's high. (72-year-old black male, urban, no mental health diagnosis, recent HbA_{1c}, 6.9, in control)*

Between visit communication between patients and providers through secure messaging affords an opportunity to more quickly address issues, including potentially changing medication management plans, to try to gain control over blood sugar:

> *I could send the readings via secure message and, then they'd say, 'Okay, well you can drop this dose or you need to add this dose. (60-year-old white female, rural, mental health diagnosis, recent HbA_{1c} 9.9, not in control).*

Patients reported that they used portal interactions to advocate for themselves and be proactive about getting what they need, including asking about how to manage blood sugar better:

> *If I notice that my blood sugar stays high and I can't seem to regulate it, I use My HealtheVet to do the secure messaging and I'll send my care provider [a message]. I usually let it go for a couple of days so that I can see an average. And, then I'll contact them and maybe they'll contact me right back and say, 'Listen, I need you to increase it by two units. And let's see if we got you on track now.' (61-year-old black male, urban, mental health diagnosis, recent HbA_{1c} 7.7, in control)*

Some patients also reported these interactions reduced the number of in-person visits required.

> *If I punched in high numbers into telehealth, [my doctor] would message me, find out what I was doing and tell me what I needed to adjust. I have to say that a lot of [my motivation to use My HealtheVet] had to do with being able to have more interaction directly with my doctor straight from my house. I almost don't even have to go to the VA anymore. (54-year-old white male, rural, mental health diagnosis, recent HbA_{1c}, 7.2, in control)*

Patients also found secure messaging to be an efficient way to ask about how to address medication side effects they were experiencing:

> *You know, I can ask her, 'This new medication is making me a little light-headed or whatever. What should I do.' And I get an answer the next day. (56-year-old white male, rural, mental health diagnosis, recent HbA_{1c} 9.7, not in control)*

Portal Functionality Hinders Engagement

Patients identified some issues with portal features that may have dampened their engagement. First, multiple patients reported being unable to use secure messaging with all of their providers. Patients do not always understand that this is a portal design issue; some misinterpret this as a purposeful act on the part of their providers and may see it as a form of rejection:

> *On secure messaging I have 6 groups of different people including my primary doctor and my pharmacist. I used to have my hepatologist and somehow they took that off of there. I do a lot of work through my hepatologist and I can't secure message*

XSL•FO
RenderX

> *him because he's not on my board anymore, which makes me very mad. (67 year old male, urban, mental health diagnosis, recent HbA_{1c} 6.1, in control)*

Second, patients reported that they could not send secure messages to individual providers, but instead had to message the team and that their uncertainty about who might read the message made them uncomfortable.

> *They call it secure messaging but it doesn't go to the specific person, it just goes to the department and then it gets trickled down from there. So you don't really know who's going to be reading all that. I have to write to the purple team, not my primary doctor. (57-year-old American Indian/Alaska Native female, urban, no mental health diagnosis, recent HbA_{1c} 7.4, in control)*

Finally, patients described challenges with the BlueButton feature that may hinder engagement, either by preventing access to information or generating frustration so that patients stop trying to engage. One patient said the information in BlueButton is not helpful and feels impersonal, "There is a lot of it that feels cookie-cutter, nothing new" (54-year-old white male, rural, no mental health diagnosis, recent HbA_{1c} 7.1, in control). Several patients reported the interface is difficult to navigate, and they expressed frustration with not knowing how to download their notes. Such challenges prevent them from obtaining information that may facilitate engagement. One reported, "finding medical record notes, that's been hard for me. I couldn't find that so I just said the heck with it." (72-year-old white male, urban, mental health diagnosis, recent HbA_{1c} 8.1, not in control)

Discussion

Through 40 interviews with patients with diabetes who used a patient portal, we engaged in a wide-ranging discussion regarding patient engagement and pathways by which portals may facilitate engagement. Access to detailed health information in the portal facilitates engagement by allowing patients to learn about their condition, remember information from provider visits that they otherwise may forget, and prepare for medical appointments. Tools to facilitate communication through the portal allow patients to partner with their providers to manage their health. Many patients found secure messaging and Blue Button features supportive of engagement; difficulties that patients reported with these features may, by extension, be barriers to patient engagement.

Patient narratives in this study align with Barello and colleagues' framework [14] of three engagement dimensions (emotional, cognitive, and behavioral) and highlight an interconnectedness between cognitive and behavioral engagement. In terms of emotional engagement, patients felt that interacting with providers via the portal improved the patient-provider relationship. Portal use cognitively engaged patients to understand their health by using portal tools such as secure messaging to ask questions and Blue Button to review clinical notes. Patient narratives revealed how cognitive engagement is translated into behavior. Access to information through the patient portal led patients to engage in behaviors to support their health and self-management. Actions facilitated by the portal's information and communication platform included preparing for visits, coordinating care among providers, and making lifestyle or medication changes between visits.

There are concerns in the literature that eHealth initiatives may worsen disparities [24] or weaken patient-provider relationships [12]. Disparities could worsen if vulnerable groups have less access to technology that promotes engagement or if vulnerable populations are less likely to use patient portals [25]. Our data suggest varied backgrounds and degrees of glucose control among patients using the portal to enhance their engagement. Portal benefits for engagement seem to be experienced by all types of patients, including individuals in urban and rural settings, with and without mental health conditions, those with diabetes in control, and those with diabetes not in control. Furthermore, several patients who were not in control described activities related to engagement. A portal may offer further opportunities to work with these patients to improve diabetes control. However, engaging patients via portal use does not necessarily indicate they will have better outcomes. In this analysis, we do not measure whether engagement facilitated by the portal influenced diabetes outcomes or patient satisfaction scores or whether this varied based on additional patient characteristics. The interviews were overwhelmingly positive about the benefits of the portal and suggested patient satisfaction is improved among those who use it. Future research should explore whether portal engagement predicts health outcomes and patient satisfaction and whether the findings are consistent among patients with other chronic conditions.

Patient-provider relationships are essential for individuals with chronic conditions like diabetes. There is some concern that portals may hurt the patient-provider relationship if portal interactions replace face-to-face interactions [12,26]. However, in our study, patients described the portal interactions between patient and provider as a key factor that facilitated their engagement. Some patients reported feeling that their interactions with providers through the portal help strengthen their relationship with providers. A point of caution: we must recognize the unintended consequence that a portal's design may have on patient perceptions of their provider's willingness to communicate with them. In our analysis, most patients found the portal helpful, but the design of the secure messaging feature led some patients to feel their providers may not want to communicate with them. Employing human-centered design may be one strategy to help mitigate portal design issues by engaging patients and providers early in the design process. One patient also complained about the impersonal, "cookie-cutter" nature of some clinical notes, which could make patients feel like their providers didn't know them as individuals, or that the notes did not capture the essence of what was discussed during the clinical encounter. Future qualitative work should continue to explore patient and provider perspectives on how portals affect their relationships and could examine provider attitudes and approaches to the portal as a possible moderator of patient engagement.

This study has several limitations. The study population was limited to United Stated military veterans with diabetes who used the My Health*e*Vet patient portal. We found

XSL•FO
RenderX

overwhelmingly positive comments regarding the portal, perhaps because all interviewees were portal users. Our study was not designed to determine the prevalence of patient engagement, but our findings do reveal potential ways in which a portal may facilitate engagement.

In this study, patient narratives helped identify pathways by which a portal may facilitate patient engagement. Patients found the portal helped strengthen their relationship with providers and helped the patient feel engaged. Patients reported finding the portal useful for receiving help in managing symptoms, coordinating their care, and learning about their health. Patient portal users included a variety of individuals who described the engagement benefits of the portal. The group included individuals living in urban and rural settings, those with and without mental health conditions, and those with controlled and uncontrolled diabetes. Thus, the portal may help a wide range of portal users engage with their care.

Acknowledgments

This study was supported by the United States Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development Service through a grant to SLS (IIR 15-307). SR and SC were supported by the United States Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship in Health Services Research. VV is supported through a United States Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development Career Development Award (CDA5IK2HX002137-02). HLW is supported by the National Institute on Drug Abuse Predoctoral training grant at Boston University School of Public Health (T32DA041898-03).

Conflicts of Interest

None declared.

References

1. Barello S, Graffigna G, Vegni E. Patient engagement as an emerging challenge for healthcare services: mapping the literature. Nurs Res Pract 2012;2012:905934 [FREE Full text] [doi: 10.1155/2012/905934] [Medline: 23213497]
2. James J. Health Policy Brief: Patient Engagement. Health Affairs 2013 Feb 14;PMID:1-6.
3. World Health Organization. Patient Engagement: Technical Series on Safer Primary Care. Technical Series on Safer Primary Care. Geneva: World Health Organization; 2016 Dec. URL: https://apps.who.int/iris/bitstream/handle/10665/252269/9789241511629-eng.pdf;jsessionid=0E3D99DF9B72018347429F78FD9A070A?sequence=1 [accessed 2020-06-15]
4. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
5. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231 [FREE Full text] [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
6. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
7. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
8. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
9. Kuo A, Dang S. Secure messaging in electronic health records and its impact on diabetic clinical outcomes: a systematic review. Telemed J E Health 2016 Sep;22(9):769-777. [doi: 10.1089/tmj.2015.0207] [Medline: 27027337]
10. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
11. McMahon GT, Gomes HE, Hickson Hohne S, Hu TM, Levine BA, Conlin PR. Web-based care management in patients with poorly controlled diabetes. Diabetes Care 2005 Jul;28(7):1624-1629 [FREE Full text] [doi: 10.2337/diacare.28.7.1624] [Medline: 15983311]
12. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]
13. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
14. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]

XSL•FO
RenderX

15. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure messaging, diabetes self-management, and the importance of patient autonomy: a mixed methods study. J Gen Intern Med 2020 May 21:1-8. [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
16. Etingen B, Hogan TP, Martinez RN, Shimada S, Stroupe K, Nazi K, et al. How do patients with mental health diagnoses use online patient portals? An observational analysis from the Veterans Health Administration. Adm Policy Ment Health 2019 Sep;46(5):596-608. [doi: 10.1007/s10488-019-00938-x] [Medline: 31065908]
17. Hogan TP, Nazi KM, Luger TM, Amante DJ, Smith BM, Barker A, et al. Technology-assisted patient access to clinical information: an evaluation framework for blue button. JMIR Res Protoc 2014 Mar 27;3(1):e18 [FREE Full text] [doi: 10.2196/resprot.3290] [Medline: 24675395]
18. US Department of Veterans Affairs, Office of Rural Health. VHA and ORH Adopt New System to Define "Rural". US Department of Veterans Affairs. 2014. URL: https://www.ruralhealth.va.gov/rural-definition.asp [accessed 2020-06-15]
19. Schmitt A, Gahr A, Hermanns N, Kulzer B, Huber J, Haak T. The Diabetes Self-Management Questionnaire (DSMQ): development and evaluation of an instrument to assess diabetes self-care activities associated with glycaemic control. Health Qual Life Outcomes 2013 Aug 13;11:138 [FREE Full text] [doi: 10.1186/1477-7525-11-138] [Medline: 23937988]
20. Lorig K, Ritter PL, Villa FJ, Armas J. Community-based peer-led diabetes self-management: a randomized trial. Diabetes Educ 2009;35(4):641-651. [doi: 10.1177/0145721709335006] [Medline: 19407333]
21. Rosen AK, Wagner TH, Pettey WBP, Shwartz M, Chen Q, Lo J, et al. Differences in risk scores of veterans receiving community care purchased by the Veterans Health Administration. Health Serv Res 2018 Dec;53 Suppl 3:5438-5454 [FREE Full text] [doi: 10.1111/1475-6773.13051] [Medline: 30251367]
22. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
23. Lune H, Berg L. Qualitative Research Methods for the Social Sciences, Ninth Edition. Boston: Pearson Education, Inc; 2017.
24. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
25. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
26. Zickmund SL, Hess R, Bryce CL, McTigue K, Olshansky E, Fitzgerald K, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med 2008 Jan;23 Suppl 1:20-26 [FREE Full text] [doi: 10.1007/s11606-007-0273-6] [Medline: 18095039]

Abbreviations

DMSQ: Diabetes Self-Management Questionnaire
DSES: Diabetes Self-efficacy Scale
ICD: International Classification of Diseases
VA: Veterans Affairs

Edited by G Eysenbach; submitted 12.01.20; peer-reviewed by D Klein, T Risling; comments to author 31.01.20; revised version received 26.03.20; accepted 03.06.20; published 17.07.20.

Please cite as:
Stewart MT, Hogan TP, Nicklas J, Robinson SA, Purington CM, Miller CJ, Vimalananda VG, Connolly SL, Wolfe HL, Nazi KM, Netherton D, Shimada SL
The Promise of Patient Portals for Individuals Living With Chronic Illness: Qualitative Study Identifying Pathways of Patient Engagement
J Med Internet Res 2020;22(7):e17744
URL: https://www.jmir.org/2020/7/e17744
doi:10.2196/17744
PMID:32706679

©Maureen T Stewart, Timothy P Hogan, Jeff Nicklas, Stephanie A Robinson, Carolyn M Purington, Christopher J Miller, Varsha G Vimalananda, Samantha L Connolly, Hill L Wolfe, Kim M Nazi, Dane Netherton, Stephanie L Shimada. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted

XSL•FO
RenderX

use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Features of a Patient Portal for Blood Test Results and Patient Health Engagement: Web-Based Pre-Post Experiment

Bas Struikman[1], MA; Nadine Bol[2,3], PhD; Annelijn Goedhart[4], BSc; Julia C M van Weert[3], PhD; Esther Talboom-Kamp[4,5], PhD; Sanne van Delft[4], PhD; Anne E M Brabers[1], PhD; Liset van Dijk[1,6], PhD

[1]Nivel (Netherlands Institute for Health Services Research), Utrecht, Netherlands

[2]Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

[3]Department of Communication Science, Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[4]Saltro, Diagnostics Center for Health Care, Utrecht, Netherlands

[5]Department National eHealth Living Lab, Leiden University, Leiden, Netherlands

[6]Department of PharmacoTherapy, -Epidemiology & -Economics, Groningen Research Institute of Pharmacy, Faculty of Science and Engineering, University of Groningen, Groningen, Netherlands

Corresponding Author:
Liset van Dijk, PhD
Nivel (Netherlands Institute for Health Services Research)
PO Box 1568
3500 BN
Utrecht
Netherlands
Phone: 31 30 2729755
Email: l.vandijk@nivel.nl

Abstract

Background: The use of patient portals for presenting health-related patient data, such as blood test results, is becoming increasingly important in health practices. Patient portals have the potential to enhance patient health engagement, but content might be misinterpreted.

Objective: This study aimed to discover whether the way of presenting blood test outcomes in an electronic patient portal is associated with patient health engagement and whether this varies across different blood test outcomes.

Methods: A 2x3 between-subjects experiment was conducted among members of the Nivel Dutch Health Care Consumer Panel. All participants read a scenario in which they were asked to imagine themselves receiving blood test results. These results differed in terms of the presented blood values (ie, normal vs partially abnormal vs all abnormal) as well as in terms of whether the results were accompanied with explanatory text and visualization. Patient health engagement was measured both before (T0) and after (T1) participants were exposed to their fictive blood test results.

Results: A total 487 of 900 invited members responded (response rate 54%), of whom 50.3% (245/487) were female. The average age of the participants was 52.82 years (SD 15.41 years). Patient health engagement saw either a significant decrease or a nonsignificant difference in the experimental groups after viewing the blood test results. The mean difference was smaller in the groups that received blood test results with additional text and visualization ($mean_{T0}$ 5.33, SE 0.08; $mean_{T1}$ 5.14, SE 0.09; mean difference 0.19, SE 0.08, P=.02) compared with groups that received blood test results without explanatory text and visualization ($mean_{T0}$ 5.19, SE 0.08; $mean_{T1}$ 4.55, SE 0.09; mean difference 0.64, SE 0.08, P<.001). Adding text and visualization, in particular, reduced the decline in patient health engagement in participants who received normal results or mixed results (ie, combination of normal and abnormal results).

Conclusions: Adding text and visualization features can attenuate the decrease in patient health engagement in participants who receive outcomes of a blood test via a patient portal, particularly when blood test results are (partly) normal. This suggests that explanatory text and visualization can be reassuring. Future research is warranted to determine whether these results can be generalized to a patient population who receive their actual blood test results.

(J Med Internet Res 2020;22(7):e15798) doi:10.2196/15798

XSL•FO
RenderX

KEYWORDS

online patient portals; patient health engagement; blood testing; blood test results; consumer panel, visualization; patient information

Introduction

Patient health engagement (PHE) can play an important role in personal health and can improve adherence and satisfaction with received care [1,2]. From a holistic perspective, PHE can be defined as a multidimensional process, including elements of cognition (thinking), emotion (feeling), and behavior (acting) of a patient towards his or her role in health management [3]. Patients' engagement in their health management is dependent on the stability of these dimensions. A lack of understanding of their position in health care can immobilize patients, making them indecisive or apathetic, while a better understanding makes them more active [3]. Engaged patients who want to be informed about and take an active role in their own health care are more likely to show healthy behaviors, have better self-management, and achieve better health-related outcomes [2,4]. Vice versa, lower PHE has been related to preventable deaths and unnecessary costs [5]. One stage where PHE is important is the stage of diagnosis [6]. When patients take an active role by asking questions and voicing their opinions, they enhance their role in their health management and increase their empowerment [7]. An important part of the stage of the diagnostic process is the information that becomes available from blood tests. However, this information can cause insecurity and uncertainty for the patient. This can negatively impact PHE as patients can become emotionally destabilized by the confusion or impact of the test results [7]. This risk is bigger when patients find the blood test values difficult to interpret [8].

Currently, electronic patient portals are increasingly offered by health professionals to communicate blood test results to the patient. Even though these portals are not optimally used yet [9], patient portals are becoming increasingly important in the health care sector and valued by both patients and practitioners [10]. Patient portals have been designed to encourage patient involvement. Yet, the way the content is presented in a portal and the way the patient interprets such content impact the overall usefulness of the information [10]. To prevent unnecessary anxiety, blood test results must be presented properly to patients [7] and needs to be done in such a way that it does not jeopardize PHE. In health practices, blood test results in a standard portal are usually communicated by giving the quantitative results (ie, the patient's value) plus reference values (ie, a range that expresses the normal values for that test). Added features such as text, symbols, or visuals to help interpret blood test results are usually not provided. However, problems have been reported about the usage of such a basic portal. For instance, the limited amount of information makes interpretation of results complex, which makes the blood test results only useful for patients with high health literacy [11]. Furthermore, misinterpretation has led to patients underestimating the severity of blood test outcomes [12].

One of the main problems is that patient portals often rely on numerals to purvey information. This is a concern for people with low numeracy skills (ie, people who lack the ability to use and draw meaning from numbers). People with low numeracy skills have shown more difficulties with identifying out-of-range test results [13]. This is likely due to the unfamiliar abbreviations, unfamiliar units, and little guidance as to whether higher numbers represent more positive or negative outcomes. This problem increases when larger sets of values are displayed at once [13]. These findings raise concerns for patient safety. Arguments for keeping this type of portal can be that people, who searched online for information, found that websites often have too much information for them to comprehend [14]. Furthermore, detailed information was found to be overwhelming for certain patients [15]. Thus, a patient portal can be useful to provide patients with individualized information (in this case, blood test results) without irrelevant information. However, this information has to be presented in a comprehensible way.

There are good reasons to believe that textual explanations with explanatory visual aids can benefit patients [16], both online and offline. For example, visualization to communicate different levels of driving risks (ie, yellow, orange, and red bars) provided good insight into the risk level of driving while using a specific medicine [17]. When risk communication is done using well-designed visual aids, information through patient portals could be received with fewer problems and enhance consent to further treatment [18]. Infographics have also been shown to be of added value in delivering complex information [19], and graphs helped a third of patients with lower numerical skills in transferring risk information [20]. When people do not understand the information, they will often use their "gut" feeling to make decisions about uncertain situations [21]. This can be a problem for PHE, as this feeling can make a patient more passive in their health management. Little is known about how presenting blood test values in a patient portal can influence PHE. The aim of this study was to explore whether the way blood test outcomes are presented in a patient portal is associated with PHE and whether this varies across different blood test outcome combinations. Exploratory research can be a crucial step to further develop scientific knowledge by laying groundwork for research topics that do not yet have a strong basis for hypotheses [22]. An experiment was conducted to systematically test the effect of different blood test results presented in a basic patient portal as well as in a patient portal in which text and visualization was added to the standard way of presenting blood test results.

Methods

Portal

The basis of our study was the comparison of two portals. Both portals communicated blood test results accompanied by reference values (ie, the range that expresses the normal values for that test). These types of portals are most often used in Dutch clinical practice, which was the setting of our study. The first portal was a fictive basic portal only providing the patient's blood values with the corresponding reference categories. The second portal was based on a more sophisticated portal as

XSL•FO
RenderX

developed by Saltro, one of the largest diagnostic centers in the Netherlands. The Saltro portal adds two main features compared to the first portal: (1) textual understandable information explaining the test and its outcomes and the action the patient can take afterwards and (2) visual support by using traffic light colors to indicate whether the outcome is within the normal range. The content of this portal was cocreated with health care professionals, communication experts, and patients. The text was written to be understandable for the majority of people. The level of health literacy of the results information has been estimated at communication level 1B on the scales of the Common European Framework of Reference for Languages. Furthermore, the content is frequently evaluated by patients and adapted according to their recommendations.

Design and Procedure

A 2x3 between-subjects experimental design was employed to test the effects of the blood test results outcome and the addition of explanatory text and visualization on PHE. Participants were so-called analog or simulated patients (ie, people who imagine themselves being in a hypothetical health care situation) presented with a hypothetical case (see Textbox 1). We opted for analog patients for two reasons. First, since we manipulated the blood test outcomes, it was unethical to use "real" patients receiving their own blood test results. Second, analog patients can be used in study designs such as ours, based on a meta-analysis that demonstrated the validity of using analog patients by showing insignificant discrepancies between the perceptions of analog patients and clinical patients [23]. The hypothetical case was identical for every participant (Textbox 1). Fatigue was chosen as a health problem as it is easy to relate to. In short, the case description stated that participants had to envision they have been tired for a couple of months now and it is not getting any better. We chose an excessive form of fatigue in order to arouse feelings of fear or worry in the participant. They went to a general practitioner who ordered some blood tests. The outcomes were to be communicated through a web portal.

We tested the two types of portals (ie, with vs without explanatory text and visualization), within which we distinguished three possible outcomes of the blood tests: all values within range (normal), partially deviating values (partially abnormal), and all deviating values (all abnormal). Table 1 shows the 6 conditions that were tested. The outcomes per blood value group were identical. Participants were randomly assigned to 1 of the 6 experimental conditions without knowing about the other 5 conditions.

Textbox 1. Fictional case presented to each participant, translated from Dutch to English.

You have been tired for a couple of months now, and it does not get any better. No matter how much you sleep, you remain tired. You even fell asleep at work once. You do not feel tense or stressed. It does bother you that the tiredness does not go away. You decide to visit your general practitioner (GP) to describe your symptoms to better understand/get a grip on your situation. The GP asks if you have any thoughts on the cause of the tiredness. You have no clue. Your private life is fine, you have never been this tired before, your diet is healthy, and you do not smoke or use medication. You have become worried about the situation; what is going on? The GP suggests to first run some blood tests before making any decisions. The GP explains that she wants to measure 3 types of blood values: tests for an underactive thyroid (thyroid stimulating hormone), anemia (hemoglobin), and a deficit of vitamin D. In the morning, your blood gets drawn, and you are told that your results are available at your convenience on the website of your GP through a patient portal the same day.

Table 1. The 6 groups of the 2x3 design.

Portal type	All normal values	Partially abnormal values	All abnormal values
No features	"Green basic"	"Partial orange basic"	"Orange basic'"
Added features[a]	"Green Saltro"	"Partial orange Saltro"	"Orange Saltro"

[a]Explanatory text with added visuals (see Figure 1).

XSL•FO
RenderX

Figure 1. Basic portal (top) and Saltro portal (bottom), both displaying the same partially abnormal results. All text is in Dutch and is shown in a similar fashion for hemoglobin and thyroid stimulating hormone.



Stimulus Materials

The Saltro portal adds explanatory text that gives an explanation about which function the substance has in the body and what exactly is being measured, while showing a bar that presents the patient's blood value in an orange (abnormal) or green (normal) range. A marker shows where the patient's blood test result falls within the range (see Figure 1). The text also offers an explanation on why measuring this particular substance is important, what it could mean if a value is below or above the normal range, and an indication of possible causes for the abnormal result. The text encourages the patient to contact their physician for further questions. The text ends with a conclusion stating whether the blood test result of the patient is below, above, or within the normal value range. Participants were exposed to 3 blood tests: hemoglobin, thyroid stimulating hormone, and vitamin D. The total word counts of the explanatory text were 271 for hemoglobin, 175 for thyroid stimulating hormone, and 194 for vitamin D.

Participants

We approached a sample of 900 participants of the Dutch Health Care Consumer Panel to complete an online questionnaire. This panel aims to measure opinions on and knowledge of health care as well as the expectations of and experiences with health care among a cross-section of the Dutch population [24]. The Dutch Health Care Consumer Panel is an access panel consisting of people who have agreed to answer questionnaires on a regular basis. Some sociodemographic details of the participants are known, such as age, gender, and highest level of education. In June 2018, the Consumer Panel consisted of about 12,000 people. To be included in the panel, a respondent has to be 18 years or older. All participants typically are asked to complete a questionnaire 3-4 times a year. Participants have complete liberty to decide whether to only answer certain questions or to participate at all. The technical functionality of the questionnaires is tested by the researcher before sending out the questionnaire. Resigning from the panel can be done at any time. People cannot sign up for the panel on their own initiative, and as such, we used a closed survey. The response rate for this study was stimulated by sending two electronic reminders to panel members who had not responded yet. The closing date of the questionnaire was at the end of May 2018. Panel members did not receive financial compensation.

Data were analyzed anonymously and processed according to the privacy policy of the Dutch Healthcare Consumer Panel, which complies with the General Data Protection Regulation. According to Dutch legislation, there is no legal requirement

XSL•FO
RenderX

to obtain informed consent nor approval by a medical ethics committee for conducting research through the panel [25]. Privacy regulation is available.

Measurements

We used the original 9-item version of the PHE scale as the starting point for our measurement [3]. The PHE scale defines patient health engagement as a multidimensional process, including elements of emotion (feeling), cognition (thinking), and behavior (acting) towards his or her role in health management. Of the 9 original items, 5 items were considered appropriate for this experiment. For 2 items, the answer categories were slightly adapted to make it more applicable to the case in our study. Of the original PHE scale, 4 items were considered irrelevant for this study, as they specified feelings about the illness itself and could not be used due to the fictional nature of the case study. Participants were asked to indicate the extent to which the 7-point semantic differential scales applied to them. The items were: "I feel in blackout" versus "I feel positive," "I feel dazed" versus "I feel serene," "I can't understand what happened to me" versus "I understand what this situation means to me," "I feel totally messed up" versus "I know what to do in this situation," and "I let others take care of me" versus "I can autonomously manage my medical regime." The participants filled out the 5-item version of the PHE scale both before (T0) and after (T1) exposure to the patient portal. All 5 items were averaged into one mean scale of PHE at T0 (Cronbach α=.90) and T1 (Cronbach α=.92). As individual items showed similar trends to those of the overall items, we did not analyze those separately. A higher score on the mean scale means a higher PHE. Because T0 was the baseline measurement (after exposure to the case scenario but before the patient portal with the blood test results), it could be expected that the health engagement score would not increase after exposure to the portal with the blood test results, but either remain the same or decrease, depending on the extent to which the blood test results would evoke elements of emotion, cognition, or behavior that could hinder patients' health engagement. The reason for this is that when the results are abnormal, the level of PHE is expected to be lower after receiving the test. Also, for those who receive normal results, an increase is not obvious because it means that the cause for the long-term tiredness is not clear.

Statistical Analysis

For the randomization check, we examined whether participant characteristics were equally divided across experimental conditions using *F* tests and chi-square tests. For age, education, and health status, we conducted 3 two-way analyses of variance with differences between "text/visualization added" and between "outcome of blood test results" as the between-subjects factors and age, education, and health status as the dependent variables. Health status was measured with a single question asking how the participants would rate their own health (1=excellent, 5=poor). For gender, we conducted a chi-square test with the combined experimental factors as one variable and gender as the other variable.

For the main analysis, a mixed analysis of variance was conducted with "text/visualization added" and "outcome of blood test results" as the between-subjects factors and "pretest measure of patient health engagement" versus "posttest measure of patient health engagement" as the within-subjects factor or repeated measure. Main effects of the between-subjects factors (ie, "text/visualization added" and "outcome of blood test results") and within-subjects factor (ie, "pretest measure of patient health engagement" versus "posttest measure of patient health engagement") were calculated. Furthermore, the interaction effect between the between-subjects factors (ie, "text/visualization added" × "outcome of blood test results"), two-way interaction effects between the between-subjects and within-subjects factors (ie, "text/visualization added" × "pretest vs posttest" and "outcome of blood test results" × "pretest vs posttest"), and three-way interaction effect between factors (ie, "text/visualization added" × "type of blood test results" × "pretest vs posttest") were assessed. Simple effects analyses were conducted in case of significant interactions between variables.

Results

Sample Characteristics

Of the 519 participants that started the survey, 487 had sufficiently complete data (ie, ≥60% of the PHE measure completed). Table 2 shows the descriptive statistics of these 487 participants. Participants were, on average, 53 years old (mean 52.82 years, SD 15.41 years) and reported to be in good to very good health (mean 2.65, SD 0.87). About half of the participants were female (245/487, 50.3%), and most people had completed a middle level (219/487, 45.8%) or higher level (222/487, 46.4%) of education. The randomization check presented no significant differences between the 6 experimental conditions with respect to gender (χ^2_5=7.52, *P*=.19), age, ($F_{2,472}$=0.32, *P*=.73, η^2=.00), education ($F_{2,472}$=0.00, *P*=1.00, η^2=.00), and health status ($F_{2,472}$=0.13, *P*=.88, η^2=.00). No control variables were included in the analyses.

XSL•FO
RenderX

Table 2. Sample characteristics (N=487).

Characteristics	Values
Age (years), mean (SD; range)	52.82 (15.41; 24-90)
Gender, n (%)	
Male	242 (49.7)
Female	245 (50.3)
Education[a], n (%)	
Low	37 (7.7)
Middle	219 (45.8)
High	222 (46.4)
Health status[b], mean (SD; range)	2.65 (0.87; 1-5)

[a]Values for education do not add up to 487 due to missing data.

[b]Self-reported health status (how would you rate your own health) ranges from 1 = excellent to 5 = poor.

Effects of the Outcome of the Blood Test Results

We found that the outcome of the blood test results, regardless of the portal design, impacted PHE after exposure to the blood test results ($F_{2,481}$=6.65, P<.001, η_p^2=.03). Receiving normal blood test results did not significantly decrease PHE (mean$_{T0}$ 5.27, SE 0.10; mean$_{T1}$ 5.10, SE 0.11; mean difference 0.17, SE 0.09, P=.07), but PHE significantly decreased after receiving abnormal blood test results (mean$_{T0}$ 5.35, SE 0.10; mean$_{T1}$ 4.69, SE 0.11; mean difference 0.66, SE 0.10, P<.001) or partially abnormal blood test results (mean$_{T0}$ 5.16, SE 0.10; mean$_{T1}$ 4.75, SE 0.11; mean difference 0.41, SE 0.10, P<.001).

Effect of Textual and Visual Explanation

Furthermore, adding text and visualization that explained the blood test results impacted PHE after exposure to blood test results ($F_{1,481}$=16.83, P<.001, η_p^2=.03). Although receiving blood test results with additional text and visualization significantly decreased PHE after receiving the results (mean$_{T0}$ 5.33, SE 0.08; mean$_{T1}$ 5.14, SE 0.09; mean difference 0.19, SE 0.08, P=.02), this decrease was significantly larger when blood test results were presented without explanatory text and visualization (mean$_{T0}$ 5.19, SE 0.08; mean$_{T1}$ 4.55, SE 0.09; mean difference 0.64, SE 0.08, P<.001).

Interaction Between the Type of Outcome of the Blood Test Result and Textual and Visual Explanation

A significant interaction effect between the outcome of the blood test results and the addition of explanatory text and visualization revealed a nuanced insight into how PHE develops after being exposed to blood test results. It showed that, for all outcomes of blood test results, a lack of explanatory text and visualization decreased PHE after being exposed to the results ($F_{2,481}$=3.83, P=.02, η_p^2=.02). More specifically, this decline occurred for normal blood test results (mean$_{T0}$ 5.02, SE 0.14; mean$_{T1}$ 4.73, SE 0.16; mean difference 0.29, SE 0.13, P=.03), abnormal blood test results (mean$_{T0}$ 5.39, SE 0.14; mean$_{T1}$ 4.61, SE 0.09; mean difference 0.77, SE 0.13, P<.001), and partially abnormal results (mean$_{T0}$ 5.16, SE 0.15; mean$_{T1}$ 4.31, SE 0.16; mean difference 0.85, SE 0.14, P<.001). However, when explanatory text and visualization were added to the blood test results, we found no significant decline in PHE for normal (mean$_{T0}$ 5.51, SE 0.14; mean$_{T1}$ 5.46, SE 0.16; mean difference 0.05, SE 0.13, P=.71) and partially abnormal (mean$_{T0}$ 5.16, SE 0.14; mean$_{T1}$ 5.19, SE 0.16; mean difference –0.03, SE 0.13, P=.81) blood test results. Yet, in the case of abnormal blood test results, even with explanatory text and visualization, PHE significantly decreased after receiving the blood test results (mean$_{T0}$ 5.31, SE 0.15; mean$_{T1}$ 4.73, SE 0.17; mean difference 0.54, SE 0.14, P<.001). These findings are presented in Table 3.

XSL•FO
RenderX

Table 3. Descriptive statistics for patient health engagement (PHE) before (T0) and after (T1) exposure to the patient portal across the experimental conditions (N=487).

Experimental conditions	n	PHE at T0, mean (SE)	PHE at T1, mean (SE)	*P* value
Normal blood test results				
Without text	83	5.02 (.14)	4.73 (.16)	.03
With text	82	5.51 (.14)	5.46 (.16)	.71
Partially abnormal blood test results				
Without text	87	5.16 (.15)	4.31 (.16)	<.001
With text	74	5.16 (.14)	5.19 (.16)	.81
Abnormal blood test results				
Without text	78	5.39 (.14)	4.61 (.09)	<.001
With text	83	5.31 (.15)	4.73 (.17)	<.001

Discussion

Principal Findings

The aim of this study was to discover whether the way in which blood test outcomes are presented in a patient portal is associated with PHE and whether this varies across different blood test outcomes. Adding textual and visual explanations to blood test results minimizes the decline in PHE when receiving blood test results in an electronic patient portal. When presenting blood test results through an existing patient portal, the group that received explanatory text and visualization in addition to their results experienced less of a decline in PHE than the group without these features. This was particularly true for patients who received normal and partially abnormal results (ie, combination of normal and abnormal results). For patients who received abnormal results for all three blood tests, health engagement significantly decreased independent of whether explanatory text and visualization were added. It can be concluded that adding text and visualization to a patient portal can attenuate PHE and therefore involve patients more in their health management, but only when blood test results are normal or partially abnormal.

Comparison With Prior Work

The focus of our study was to gain insight into how providing explanatory text and visualization, when presenting blood test results via a patient portal, influences PHE. While previous research has shown that patients are generally satisfied with the use of a patient portal to check their blood test results [16], there are also concerns with such portals.

One recurring concern with patient portals is the fear of the misinterpretation of results [9,10]. For example, in a study by Korngiebel et al [26], clinicians' main concern was that providing blood test results to patients without explanation could lead to confusion due to the sensitive and complicated nature of the test results. Especially divergent results would have to be carefully shown. Although our results cannot confirm if the nature of the feelings of our participants are due to confusion, it does show signs of fear, as lower PHE is associated with greater emotional immobilization. Participants who received normal or partially abnormal results in the "standard" way showed a significant decline in PHE, while health engagement in patients who received explanatory information and visual support remained at the same level as at baseline.

In this study, we did not distinguish between the added value of explanatory text only and the added value of visualization only. Hence, we do not know which of these two features was the "active ingredient" or whether the combination of the features caused the effectiveness of the patient portal. In two previous studies, information provision in a patient portal was evaluated in isolation. In the first study, visual support using a color scheme to differentiate between normal results and divergent results was positively evaluated by elderly patients [27]. In the second study, veteran patients did not have extra text available but instead were given a search bar to find relevant health information themselves [28]. The veterans evaluated the search bar positively. Both these studies based the evaluation of the patient portal on opinions, whereas our study measured effectiveness in terms of PHE.

In addition, our study measured PHE using a questionnaire. Previous studies have used different ways to measure (patient) engagement. To illustrate, Phelps et al [29] defined engagement in terms of portal usage and measured it by monitoring the number of logins. They found an increase in logins, for instance shortly before meeting a physician, and results were therefore seen as positive. We used a different operationalization of engagement, including emotions and cognitions that test results can evoke, which resulted in a more nuanced picture of how people can respond upon being confronted with blood test results. The explanatory text and visualization had a positive influence on PHE, such that it reduced the decline in engagement after being exposed to blood test results. It did not improve PHE compared to the baseline measure. However, since we were studying how people react upon exposure to (potential) risk information, it seems to make sense to aim for a reduction of negative reactions rather than for an increase of positive reactions.

Strengths and Limitations

A strength of our study is that it systematically tested the impact of patient portal design on PHE in a highly controlled experimental setting. Although patient portal designs have been introduced to patients, these have not been studied yet in terms of how their design can impact PHE. In addition, our

XSL•FO
RenderX

measurement of PHE gave a broad depiction of PHE through patients' feelings, thoughts, and behaviors. Therefore, our study design gives this study a good foundation for its conclusions and implications.

There were also some limitations to our study. The first one is the potential bias in our sample. Although the sample was randomly selected from the Nivel Consumer Panel, which represented a relatively diverse group of patients, the participants of this panel are people who agreed to fill out questionnaires about health and health care. This means that there is good reason to believe that a majority of the people have, at least to some extent, affinity with their personal health (and health care). Furthermore, it can be assumed that the average panel user has a higher health literacy than normal. Health literacy entails the capability of obtaining, processing, and understanding information about health and health services [30]. If this assumption holds true, it is relevant to consider that affinity with health care might coincide with a better understanding and processing of health care information. Thus, extra information might not be as vital to them as to those with lower health literacy. In that sense, our sample resembled the target group of the portal with added text, as both our questionnaire and the portal require some level of functional literacy.

The necessity for participants to empathize with a fictional case was another limitation. Actual patients, for whom results might indeed have an impact on their life, might have reacted differently to the results of the blood test as they might be expected to be more emotionally involved. Based on the meta-analysis by Van Vliet et al [23], we do not expect that the results in actual patients will be weaker than in the current sample of analog patients, and they might even be stronger in actual involved patients. Thus, a replication of our study with actual patients would be of added value. Moreover, we used the original 9-item PHE scale [3] as the starting point for our measurement while rephrasing and shortening it for the purpose of our study, resulting in a scale with high internal consistency. The reason for doing this was that there is no scale available that is validated for the case in our study (tiredness).

Another possible limitation is the specific design of our portal. Our results could not cover all portal designs, so the scope of the study may be limited and not generalizable to other patient portals. For instance, the colors of our "complementary" portal design could have influenced the feelings of the participants. Research has shown that the color red can evoke feelings with patients [31]. The color orange might cause an unnecessary reaction as well, since it can be associated with danger or a sign for increased attention. It could therefore be that a participant who viewed a full or partially orange result were more anxious, although an abnormal result does not automatically mean a danger to your health or the need for immediate medical intervention. Future research can address the consequences by implementing more neutral colors. This way, a patient's anxiety can be ascribed to the conclusion of the blood test result and not stimulation by a bright color. Finally, as mentioned before, our existing portal had two features: additional text for explanation and added visualization for understanding the blood levels. The downside of this is that a distinction cannot be made between the impact of the text and that of the visualization. If electronic patient portals are to be improved, it is necessary to understand if the limiting impact on the decline in PHE is due to either one of the features or perhaps the interaction between them.

Clinical and Research Implications

Our study has provided insight into how portal design can benefit PHE. Our results suggest that institutions that use a patient portal for blood test results might want to consider adding complementary information and visual support. Further testing of patient portal features is recommended, but institutions would do well to start adding visualization and textual support to their portals as this can benefit PHE. Adding text and visualization that explain the blood values and the implications of an abnormal result might support their patients in their health management. It might lead to patients who are more involved in their diagnosis and treatment, which could lead to, for instance, patients being more comfortable voicing concerns or asking questions. As mentioned, there is a possibility that our sample consisted of participants with an above-average health literacy. Therefore, future research could focus on patients with below-average health literacy who might need a different portal design. Yet, more research among actual patients is needed to test the portal with patients who are personally involved and therefore more concerned about the outcomes of blood test results. For example, in case a fictional study is chosen, this fictional case could include a disease with a higher emotional impact and familiarity, such as cancer. For such a disease, a more intense reaction could occur while viewing blood test results, such as relief or anxiety, which in turn could have a stronger effect on PHE. However, ethical considerations should be taken into account, as participants often do not know that they are being confronted with such an emotional case. A warning up front for the participant might be considered if the need for a more emotional investment is deemed necessary. Lastly, we used an adaptive version of the validated, original PHE scale. By deleting 4 items, it is possible that it did not cover the full domain of PHE (ie, did not measure the domains of feelings, thoughts, and behaviors as fully as the original scale).

Conclusion

Patient portals have been designed to improve patient involvement. When blood test results are communicated to patients, it can negatively affect their PHE and consequently their involvement in their health management. However, when these outcomes are supported by explanatory text and visualization to help interpret the outcomes, the decrease in PHE can be attenuated, especially when test results are partially normal. As receiving test results can cause feelings of uncertainty in patients, which can lead to lower PHE, our results suggest that explanatory text with visualization can cause feelings of relief in patients. Future research should focus on repeating the experiment with actual patients who receive their own blood test results to test whether the results hold in more ecologically valid settings.

XSL•FO
RenderX

Conflicts of Interest

BS, NB, JvW, AB, and LvD do not have conflicts of interest to declare. AG, ETK, and SvD are employees of Saltro, the organization that developed the portal that was tested in this experiment.

References

1. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
2. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
3. Graffigna G, Barello S, Bonanomi A, Lozza E. Measuring patient engagement: development and psychometric properties of the Patient Health Engagement (PHE) Scale. Front Psychol 2015;6:274 [FREE Full text] [doi: 10.3389/fpsyg.2015.00274] [Medline: 25870566]
4. AARP Public Policy Institute. Beyond 50.09 Chronic Care: A Call to Action for Health Reform. URL: https://assets.aarp.org/rgcenter/health/beyond_50_hcr.pdf [accessed 2020-06-11]
5. Laurance J, Henderson S, Howitt PJ, Matar M, Al KH, Edgman-Levitan S, et al. Patient engagement: four case studies that highlight the potential for improved health outcomes and reduced costs. Health Aff (Millwood) 2014 Sep;33(9):1627-1634. [doi: 10.1377/hlthaff.2014.0375] [Medline: 25201668]
6. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington DC: The National Academies Press; 2015.
7. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii33-ii39 [FREE Full text] [doi: 10.1136/bmjqs-2012-001623] [Medline: 23893394]
8. Garrido T, Jamieson L, Zhou Y, Wiesenthal A, Liang L. Effect of electronic health records in ambulatory care: retrospective, serial, cross sectional study. BMJ 2005 Mar 12;330(7491):581 [FREE Full text] [doi: 10.1136/bmj.330.7491.581] [Medline: 15760999]
9. Wouters M, Swinkels I, van Lettow B, de Jong J, Sinnige J, Brabers A, Friele R, van Gennip L. E-health in verschillende snelheden, eHealth-monitor 2018. The Hague / Utrecht: Nictiz/Nivel; 2018.
10. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
11. Parker RM, Ratzan SC, Lurie N. Health literacy: a policy challenge for advancing high-quality health care. Health Aff (Millwood) 2003;22(4):147-153 [FREE Full text] [Medline: 12889762]
12. Fraccaro P, Vigo M, Balatsoukas P, van der Veer SN, Hassan L, Williams R, et al. Presentation of laboratory test results in patient portals: influence of interface design on risk interpretation and visual search behaviour. BMC Med Inform Decis Mak 2018 Feb 12;18(1):11 [FREE Full text] [doi: 10.1186/s12911-018-0589-7] [Medline: 29433495]
13. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and literacy independently predict patients' ability to identify out-of-range test results. J Med Internet Res 2014;16(8):e187 [FREE Full text] [doi: 10.2196/jmir.3241] [Medline: 25135688]
14. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
15. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013 Aug 26;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
16. McNamara M, Arnold C, Sarma K, Aberle D, Garon E, Bui AAT. Patient Portal Preferences: Perspectives on Imaging Information. J Assoc Inf Sci Technol 2015 Aug 01;66(8):1606-1615 [FREE Full text] [doi: 10.1002/asi.23269] [Medline: 26191535]
17. Monteiro SP, Huiskes R, Van Dijk L, Van Weert JCM, De Gier JJ. How effective are pictograms in communicating risk about driving-impairing medicines? Traffic Inj Prev 2013;14(3):299-308. [doi: 10.1080/15389588.2012.710766] [Medline: 23441949]
18. Garcia-Retamero R, Cokely ET. Designing Visual Aids That Promote Risk Literacy: A Systematic Review of Health Research and Evidence-Based Design Heuristics. Hum Factors 2017 Jun;59(4):582-627. [doi: 10.1177/0018720817690634] [Medline: 28192674]
19. Otten JJ, Cheng K, Drewnowski A. Infographics And Public Policy: Using Data Visualization To Convey Complex Information. Health Aff (Millwood) 2015 Nov;34(11):1901-1907. [doi: 10.1377/hlthaff.2015.0642] [Medline: 26526248]
20. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med Decis Making 2011;31(3):444-457. [doi: 10.1177/0272989X10373805] [Medline: 20671213]
21. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. Science 2011 Sep;333(6048):1393-1400. [doi: 10.1126/science.1191181] [Medline: 21903802]
22. Cannon HM, Williams DL. Toward a Hierarchical Taxonomy of Magazine Readership. Journal of Advertising 1988 Jun;17(2):15-25. [doi: 10.1080/00913367.1988.10673109]

XSL•FO
RenderX

23. van Vliet LM, van der Wall E, Albada A, Spreeuwenberg PMM, Verheul W, Bensing JM. The validity of using analogue patients in practitioner-patient communication research: systematic review and meta-analysis. J Gen Intern Med 2012 Nov;27(11):1528-1543 [FREE Full text] [doi: 10.1007/s11606-012-2111-8] [Medline: 22700392]
24. Brabers A, Reitsma-van Rooijen M, De Jong J. Consumentenpanel Gezondheidszorg: basisrapport met informatie over het panel. NIVEL: Utrecht; 2015:1-55.
25. Your research: does it fall under the WMO? CCMO. The Hague: CCMO; 2019. URL: http://www.ccmo.nl/en/your-research-does-it-fall-under-the-wmo
26. Korngiebel DM, West KM, Burke W. Clinician-Stakeholders' Perspectives on Using Patient Portals to Return Lynch Syndrome Screening Results. J Genet Couns 2018 Apr;27(2):349-357 [FREE Full text] [doi: 10.1007/s10897-017-0179-3] [Medline: 29159545]
27. Gradinar A, Davenport J, Hill H, Coulton P. Improving the Visualisation of Renal Blood Test Results to Enhance Patient – Clinician Communication. The Design Journal 2017 Sep 06;20(sup1):S2363-S2374 [FREE Full text] [doi: 10.1080/14606925.2017.1352751]
28. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to Electronic Health Records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
29. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
30. Williams MV, Davis T, Parker RM, Weiss BD. The role of health literacy in patient-physician communication. Fam Med 2002 May;34(5):383-389. [Medline: 12038721]
31. Pravossoudovitch K, Cury F, Young SG, Elliot AJ. Is red the colour of danger? Testing an implicit red-danger association. Ergonomics 2014;57(4):503-510. [doi: 10.1080/00140139.2014.889220] [Medline: 24588355]

Abbreviations

GP: general practitioner
PHE: patient health engagement

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by E Ahn, T Irizarry, R Williams, E Metting; comments to author 21.10.19; revised version received 15.01.20; accepted 22.03.20; published 20.07.20.

Please cite as:
Struikman B, Bol N, Goedhart A, van Weert JCM, Talboom-Kamp E, van Delft S, Brabers AEM, van Dijk L
Features of a Patient Portal for Blood Test Results and Patient Health Engagement: Web-Based Pre-Post Experiment
J Med Internet Res 2020;22(7):e15798
URL: https://www.jmir.org/2020/7/e15798
doi:10.2196/15798
PMID:32706704

©Bas Struikman, Nadine Bol, Annelijn Goedhart, Julia C M van Weert, Esther Talboom-Kamp, Sanne van Delft, Anne E M Brabers, Liset van Dijk. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Untapped Potential in Primary Care: Semi-Structured Interviews with Clinicians on How Patient Portals Will Work for Caregivers in the Safety Net

Alejandra Casillas[1], MD, MSc; Anupama Gunshekar Cemballi[2], BA, MA; Anshu Abhat[3], MD, MPH; Miya Lemberg[2]; Jennifer D Portz[4], MSW, PhD; Shobha Sadasivaiah[5], MD, MPH; Neda Ratanawongsa[2], MD; Wagahta Semere[2], MD; Arleen Brown[1], MD, PhD; Courtney Rees Lyles[2], PhD

[1]Division of General Internal Medicine and Health Services Research, Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, United States

[2]UCSF Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, UCSF Center for Vulnerable Populations, San Francisco, CA, United States

[3]Los Angeles County Department of Health Services, Harbor-UCLA Medical Center, Los Angeles, CA, United States

[4]Division of General Internal Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[5]Office of Health Informatics, San Francisco Health Network, San Francisco, CA, United States

Corresponding Author:
Alejandra Casillas, MD, MSc
Division of General Internal Medicine and Health Services Research
Department of Medicine
UCLA David Geffen School of Medicine
1100 Glendon Avenue
Suite 850
Los Angeles, CA, 90278
United States
Phone: 1 310 794 0150
Email: acasillas@mednet.ucla.edu

Abstract

Background: Patients within safety-net settings are less likely to access health information on patient portals, despite expressed interest. Family and friends are important resources to assist these patients (ie, Medicaid recipients, older patients, patients with limited English proficiency) in navigating health systems, and provider support of the use of patient portals among these groups may also facilitate caregivers' use of their patients' portal.

Objective: Because safety net providers work closely with caregivers to care for their patients, we used qualitative methods to explore safety net providers' perspectives on portal use among caregivers for their patients, especially as there is limited literature about caregivers' use of portals in the safety net.

Methods: We conducted 45- to 60-min semistructured telephone interviews with providers from three large California safety-net health systems. The interviews focused on providers' experiences with caregivers, caregiver roles, and how the portal could be leveraged as a tool to support caregivers in their responsibilities. A total of three coders analyzed the interview transcripts using both deductive and inductive approaches and established a consensus regarding major themes.

Results: Of the 16 participants interviewed, 4 specialized in geriatrics, and all held a leadership or administrative role. We described themes highlighting providers' recognition of potential benefits associated with caregiver portal use and specific challenges to caregiver engagement.

Conclusions: Providers recognized the potential for portals to improve information delivery and communication by helping caregivers assist socially and medically complex patients in the safety net. Providers in safety net sites also discussed a clear need for better ways to keep in touch with patients and connect with caregivers, yet security and privacy are perhaps of higher importance in these settings and may pose challenges to portal adoption. They noted that caregivers of patients in the safety net likely face similar communication barriers as patients, especially with respect to digital literacy, health literacy, and English proficiency. Further research is needed to assess and support caregivers' interest and ability to access portals across barriers in health and

XSL•FO
RenderX

digital literacy, and English proficiency. Portal platforms and health systems must also address specific strategies to uphold patient preferences while maintaining privacy and security.

(J Med Internet Res 2020;22(7):e18466) doi:10.2196/18466

KEYWORDS

patient portal; caregivers; vulnerable populations; digital divide; mobile phone

Introduction

Background

Growing evidence on the benefits of patient engagement has fueled health systems' focus on patient portals as a central access point for the future of primary care [1,2]. The financial incentives of the Meaningful Use Program, as part of the US health care reform in 2014, spurred a rapid uptake of patient portals across health systems nationwide [3-5]. Emerging national evidence in the last 5 years indicates that the digital divide is shrinking, as 80% of the US population owns internet-accessible smartphones, without differences by race or ethnicity and with few differences across income categories nationally, especially when it comes to mobile use [6,7].

Despite improved health information technology (IT) access, there is a large body of evidence demonstrating significantly lower use of digital health care tools among underserved populations, underscoring the need for more research to understand the contextual factors affecting their use [8,9]. Vulnerable patients within safety-net settings (eg, low socioeconomic status, under- or uninsured, limited English proficient [LEP], aging, and physically or mentally disabled) are interested in accessing their health information on patient portals but are less likely to actually do so. Numerous studies have shown that there is a high level of patient interest in portals, repeatedly documented among LEP populations, Medicaid/Medicare recipients, and older patients [10-14]. A 2014 survey of Spanish-speaking patients in a Los Angeles Federally Qualified Health Center found that the majority of patients had computer access (66%), internet access (78%), a current email account (78%), or a smartphone (71%) and that 75% of patients were interested in using a patient portal [10].

Family and friends are important resources to assist these underserved populations (especially those that are Medicaid recipients, older patients, or LEP) in navigating health systems, and provider support could facilitate their portal engagement. Caregiver engagement with a patient's portal may range from simply assisting the patient to log on and use the portal to logging onto a patient portal for the patient (unofficial patient *surrogate*) all the way to acting as a registered proxy for portal access (where caregivers are granted their own registered access to a patient's health information on the patient portal with the patient's permission) [15,16]. Older patients, especially those who are LEP, tend to navigate health system processes with family members or close friends who schedule and attend medical visits, coordinate care, manage medications, assist with self-care tasks, and facilitate transitions across care settings [15,16]. Patients with low health literacy similarly benefit from the involvement of trusted family or friends during medical visits [17,18]. Prior studies among older patients have confirmed that family members facilitate patient access and use of the patient portal [19,20] and utilize the patient portal for the patient [21-25]. Given that these patients are interested in sharing electronic health information with family members or close friends [19,20,26,27], tailored strategies are needed to engage family members or friends in accessing health care digitally, especially in safety-net settings with a majority proportion of vulnerable patients and where caregivers play a key role in patient care.

The growing role of caregivers and caregiver/proxy portal use among vulnerable patients presents a potential well-aligned digital strategy for reducing racial/ethnic disparities in health that can be facilitated via already available health IT. However, buy-in from physicians and other health care providers will be important in promoting patient portals on this next level, especially in safety-net settings [28,29]. The aim of this qualitative study was to explore safety-net providers' perspectives on caregiver/proxy portal use.

Methods

Design and Study Setting

The study was approved by the University of California, San Francisco and University of California, Los Angeles institutional review boards. For this descriptive, qualitative study [30], we conducted individual telephone semistructured interviews with a purposive sample of physicians or providers (n=17), working in a California safety-net setting in Los Angeles, San Francisco, or Alameda County. Los Angeles participants were recruited from the Los Angeles County Department of Health Services (LAC DHS) system. LAC DHS forms the core of the health care safety-net for indigent populations in Los Angeles County, the largest, most ethnically diverse county in the United States. LAC DHS serves more than 10 million residents and provides over two and a half million ambulatory visits every year [31-33]. In northern California, providers were recruited from the San Francisco Health Network (SFHN) and the Alameda Health System (AHS). The AHS is the East Bay's (Oakland area) public health system with 5 hospitals and 4 primary care clinics, whereas SFHN is San Francisco City's and county's public health care system, providing primary care at 14 clinics, youth-focused care at 11 clinics, and care at a hospital serving over 1 million residents. The internet patient portals for the safety-net health systems included in the study allow patients to view clinical data (laboratories and radiology reports), refill medications, request/change appointments, message their health care teams, read about their prior and current diagnoses (links to educational content), and view/download their medical records.

XSL•FO
RenderX

Participants and Recruitment

From April to November 2018, we recruited and interviewed providers from the LAC DHS, SFHN, and AHS. Providers were sought if they had clinical activity in a safety-net health setting and/or held a leadership role in any safety-net clinical setting. Los Angeles participants were recruited via personal recommendations from the medical director of digital health services in LAC DHS. The SFHN and the AHS providers were recruited via recommendation from study coinvestigators. We conducted 45- to 60-min semistructured interviews with health providers from these 3 large California safety-net health systems. Participants were interviewed via telephone. AC, AGC, WS, and CL conducted the interviews on multiple days until at least three providers were interviewed from each of the 3 health systems. Participants were provided with a US $25 gift card after completing their scheduled interviews.

Data Collection

The semistructured interviews focused on providers' experiences with caregivers, the roles caregivers play, and how the portal could be leveraged as a tool to support caregivers in their responsibilities. The discussion questions were modified from projects led by study investigators who examined barriers and facilitators to portal use to develop a randomized trial on patient portal training [34]. This prior research highlighted the limited knowledge and understanding of caregivers' use of the patient portal.

Discussion introductions included participants' self-describing clinical and administrative roles. Participants were then asked about their experiences with patient caregivers (definition, types of caregivers, and positive and challenging interactions). They were subsequently tasked to describe strategies that they have observed caregivers take on with patient health management. We asked participants about how the patient portal affects these caregiver strategies, what they thought was useful, and how they envisioned the portal could affect these patients' and caregivers' care. The discussions wrapped up with a conversation about what the health system could do to help improve caregiver and patient engagement with the portal.

Analysis

Interview discussions were audio recorded and transcribed, reviewed for accuracy, and deidentified. AC and AGC independently read and summarized 6 transcripts over multiple theme meetings, applying inductive and deductive methods to identify the spectrum of themes encountered for each of the interviews with accompanying example quotations. AC and AGC then cross-referenced their independently developed list of themes. Through codebook discussions, they reached a consensus for a more final codebook. AC and AGC then coded 3 transcripts independently with this codebook for validity. The definition for each theme (code), with the final example quote, was developed by AC, with iterative feedback from the AGC and the entire group, until a clear consensus was achieved. Using the established codebook and definitions, a third coder (ML) analyzed the transcripts using Dedoose version 8.2.14. AC and AGC reviewed the analysis by ML and established a consensus regarding any other discrepancies in themes and corresponding quotes [35].

Results

Participant Characteristics and Theme Categories

Of the 16 participants interviewed, 4 specialized in geriatrics, and all held a leadership or administrative roles, as shown in Table 1. All the participants worked in primary care settings, including a speech language pathologist.

We described primary themes highlighting safety-net physicians' and providers' recognition of potential strong benefits associated with caregiver portal use and specific barriers to caregiver engagement. The 4 major themes are separated into 2 large categories: (1) positive aspects of portal use by caregivers (Table 2); and (2) challenges to address for portal use by caregivers (Table 3). These show major themes, subthemes, frequencies, and quotes.

XSL•FO
RenderX

Table 1. Safety-net provider participants' characteristics.

Clinical specialty	Safety-net role
Geriatrics	Physician, Medical Director
Geriatrics	Physician, Department Chief
Geriatrics	Physician, Director of Primary Care
Geriatrics	Physician, Outpatient Care Medical Director
Internal medicine	Physician, Clinical Lead
Internal medicine	Physician, Clinical Lead
Internal medicine	Physician, Health Care Executive
Internal medicine	Physician, Medical Director
Internal medicine	Physician, Medical Director
Internal medicine	Physician, Resident Preceptor
Internal medicine	Physician, Director of Outpatient Care
Internal medicine	Physician, Director of Primary Care Quality
Family medicine	Physician, Assistant Medical Director
Family medicine	Physician, Health Care Executive
Family medicine	Physician, Health Care Executive
Occupational therapy	Speech Pathologist, Health Care Executive

XSL•FO
RenderX

Table 2. Themes and subthemes on positive aspects of portal use with exemplar quotes.

Themes/subthemes	Quotes
Caregiver designation in the electronic health record (n=25)	
	• "I record it in my notes...at the first part of workflow...registration staff is collecting information about emergency contacts and alternate contacts...depending on what the patient or their caregiver says at registration, there might be something entered into the chart in that capacity." • "Sometimes, because usually it would be because the note would say, 'Accompanied by daughter' or 'Discussed with daughter', 'Discussed with son', 'Discussed with caregiver', 'Discussed with IHSS [in home support services] worker,' but not all the time." • "I don't think uniformly. You'll see it (caregiver information) in the body of the note, not always in the, social history section. And to tell you the truth, our EHR doesn't have a good section on caregiver information...in terms of like whose number should you call? Do they have a caregiver they can call?" • "[I'm] trying my best to collect the information and actually put it in the EHR so I know who they are...oftentimes, after I did that, [with] a group of patients, I always follow up if they live in supportive housing, or they have a case manager and so that's external, nonfamily caregiver visit their office, followed up by email to whoever their support people are in the community to make sure that they know what the next steps are."
Caregiver use and potential use of the portal (n=55)	
Portal as a tool to assist caregivers with standard health care tasks (n=32)	• "I think caregivers using the patient portal could co-manage patient's health, could be a huge asset to the caregiver and the provider... it's not infrequent that the patient or the patient and the caregiver show up to appointments saying they ran out of their meds two months ago." • "Usually there's more than one caregiver, where there's a network and may have varied involvement. Like with my sisters and me, we have this constant flow of information, and having a way to put it all in one place to share easily is really important." • "...they're just rolling out this new scheduling tool that will allow internal schedulers to be able to search for ways of grouping their appointments. So, the [OT, PT,] speech could all happen on the same day. If that tool could be made available to a caregiver so they can say 'Okay, I wanted to have doctor's appointments at this and this and this all on the same day', I think that would be really powerful."
Portal as a tool to directly support the caregiver (n=7)	• "Empowering the caretakers to be able to use technology decreases their stress level." • "I think respite is also a big thing, caregiver relief and caregiver fatigue is a big problem sometimes, and having access to joint behavioral health services, or like couple counseling, or parent-child counseling in the study of chronic disease, I think it's something super underutilized. That could probably help a lot with the challenges." • "I know there's a lot of groups that are working on this or have published. I know at UCLA in geriatrics, they've done an Internet caregiving education course, and I know at Stanford, they put together Caregiving Ed. And at the VA, there's several well published evidence-based caregiving teaching programs, but our patients certainly don't have—or a lot of my patients don't have access to that." • "I would be really interested in the opportunity to do more training for caregivers. I think that's a great business opportunity, because so many caregivers have no idea what they're doing and would be happy to get trained. You can imagine the range of trainings, but our organization or your organization, or you know some private—any big county—I could imagine offered some form of training for caregivers. I think having a portal will help hopefully and I think having processing in place to clarify what can be shared and what can't be shared with caregivers, if it's documented well."
Expanding portal functionality for caregiver use (n=18)	• "For caregivers who have a homebound elder, to be able to, you know, put on their FaceTime and for me to see what's going on that would be amazing." • "...allowing patients to search and contribute to the medical notes before the visit, the medical records, typing in their symptoms. I think there should be conversations of getting caregivers involved in the patient's care." • "If people just keep track of hey, someone's calling not for themselves and just kept a running list, and then you do active outreach for anyone who is calling not for themselves."

XSL•FO
RenderX

Table 3. Themes and subthemes on challenges to portal use with exemplar quotes.

Themes/subthemes	Quotes
Portal privacy and security in the caregiver-provider-patient triad (n=38)	
Portal triad relationship (n=7)	• "The one downfall of caregivers having access is, is there any loss in translation or, if some plan is made on the portal, is that being followed through by the caregiver?" • "One of the few that I communicate with is a husband and wife patient and this is a challenge...because then I have to document it in her, in his chart, where her messages get documented in his chart... and it's worked because she has direct access to me and then he has direct access too, but she uses her account to...manage his health... And, yeah, I thought it would be nice if she could switch back and forth." • "If it's something I can communicate to them immediately, then I send a message through the portal. Then, if there's anything else that I might need to communicate with them, and I don't need to call them, then I'll do it through the portal. A lot of it is dependent on, 'What can I communicate briefly in writing without creating [confusion] on their part or some more questions than answers?'" • "I'm observing body language, so on rare occasions, if I'm concerned that perhaps the person feels like they can't ask the other person to leave the room, then I go ahead and...When I'm gonna be asking something sensitive in the interview around social history or any history, I'll actually ask the other person to step out of the room and wait in the waiting room and I just normalize that and say, 'You know, I always do this to respect confidentiality'. That's in person, so what do you do when tech comes in?" • "If you build a system that is specifically asking, 'Is this a proxy person?' and kind of asking for that designation right up front and putting it in black and white, you do run the risk of—of having more scrutiny and I think it's appropriate." • "I think that what we're doing right now is we are just in the infancy of using our portal for patient communication." • "And when I get an odd message or a little crazy message, you know, or someone requesting whatever, I screenshot it and I send it to the site managers and I tell them to call them on the phone to get more information."
Control of shared access (n=13)	• "When we see some research that suggests that some patients don't want the entire portal revealed to their loved ones, but maybe part of it...and so patient privacy is—is a big concern. A husband may not want his wife to know that he was a prior IV-drug user from a different life." • "I have mixed views because we try the different role for parents from the portal once the kid at 13. And sometimes they want to come for birth control and they didn't want their moms to know. So, I don't want the moms to see the kid's the portal." • "I mean just mostly with like the proxy access...they've just had a lot of questions and concerns around, 'When is it okay? Do we need a legal document that states, "Yes, ma'am. They are my legal healthcare proxy,"' or if they're, on mild dementia, but they say, 'No, let her do it', is that sufficient?"
General portal security and technology concerns (n=11)	• "I do have some HIV patients who won't join because they're just afraid of the internet, in general...I think it's a broader concern beyond just their HIV status but just that they don't think it's safe to have their information in the cloud basically, because it could be stolen or, or utilized in a way that's negative." • "How do you build a portal that respects patient privacy? You also wonder if adding a caregiver affects the security of the health data."
Barriers to caregiver enrollment and use (n=46)	
	• "The number of people [caregivers] who don't have access to either a smartphone or the internet...I think is just the reflection of the income level and kind of resource constraints of our patient population." • "I'm guessing that there would be times when people are concerned by the results that look abnormal or don't understand them because no one explained it to them...." • "A lot of our patients just don't really use like electronic technology and they don't speak English, or they don't have an internet connection or a computer."

Positive Aspects of Portal Use by Caregivers

Table 2 displays themes and subthemes for the category describing positive aspects of portal use. In the theme, *caregiver designation in the electronic health record*, participants described various potential pathways for identifying caregivers via the electronic health record (EHR) and how this would be useful. Some participants also described the challenges of not having a workflow to readily integrate current caregiver information into the EHR and how having such information available would make it easier for physicians and other providers to engage the caregivers during and outside of the clinic encounter. This was important, as the participants in the study noted multiple *caregiver types*, which are present in different ways for patients within safety-net settings: relatives, friends, case managers, in-home supportive services, and social workers.

XSL•FO
RenderX

Participants also discussed *caregiver use and potential use of the portal*, detailing current uses of patient portals by caregivers and aspirational ideas about what the patient portals could potentially do for safety-net caregivers in terms of facilitating patient care. This theme was organized into 3 subthemes: for subtheme 1, *portal as a tool to assist caregivers with standard health care tasks*, participants mentioned ways that the current versions of their patient portal supported or eased the job of the caregiver by facilitating simple health care tasks for the patient via the portal (eg, obtaining medication refills or making appointments), making it easier for the caregiver to care for the patient (as they might have to spend less time on the phone making calls and possibly decreasing the need for face-to-face visits).

For subtheme 2, *portal as a tool to directly support the caregiver*, participants discussed how portals could provide well-being/resources specifically tailored for the caregiver. For example, they mentioned that patient portals could be a way to provide explicit support to caregivers (eg, preventing caregiver burnout) and connect caregivers directly to community resources (including information about self-care and programs that assist caregivers in the safety-net) that make the caregiving job more manageable. They also noted that such resources could include a tutorial on caregiver education/navigation. The portal could provide caregivers access to tools on how to care for patients and/or how to navigate a complicated safety-net health care and social service system.

For subtheme 3, *expanding portal functionality for caregiver use*, discussions centered on a wish list of what the portal could do in the future (Portal 2.0) for caregivers. Participants mentioned how technology could help paint a better picture of the patient and caregiver via telehealth (access to home context) and more frequent communication. Such modalities could facilitate the implementation of evidence-based initiatives such as advanced care planning in the safety-net.

Challenges to Portal Use by Caregivers

Table 3 displays themes and subthemes related to challenges to portal use by caregivers. The major theme, *portal privacy and security in the caregiver-provider-patient triad*, encompassed unresolved issues that these participants noted in communicating with the patient and caregiver through the portal. They raised several challenges that may have not been fully thought through yet as patient portals are being marketed to caregivers in the safety-net. These were summarized in 3 subthemes.

For subtheme 1, *portal triad relationship*, participants described the nuances of building or managing a relationship with the caregiver, as well as the patient, through the portal. For example, many times throughout these discussions, physicians cited the many nuanced ways in which they would validate caregiver reporting during clinic visits (talk to patient and caregiver separately), check for caregiver abuse of the patient by examining the patient or reading body language, and simply also assess how the caregiver was doing with the burden. They noted that this subtle art of relationship management and checking in with a caregiver, which are essential features of the triad relationship [24,36,37], would be very difficult to achieve via digital communication and expressed concern that some of these needed in-person checks and balances would be lost via the current version of our patient portals.

For subtheme 2, *control of shared access*, participants mentioned concerns about proxy access. For example, what if a patient allows a caregiver to log in as a proxy into their patient portal, but the patient does not want the caregiver to see their entire medical record? Or what if the proxy sends messages to the provider on the portal as the *patient*, but the provider is not aware that they are communicating with someone that is not their patient? This is an important concern in a safety-net population where patients may have sensitive diagnoses (such as sexually transmitted infections, illicit drug use, addiction, or mental illness) that have not been disclosed to the caregiver or other close friends or family because they carry some social stigma. Sensitive information on the portal could also reveal risks to family caregivers that they may not have been made aware of (genetic diseases or genetic risk factors). Participants also noted the simple workflow barriers to creating proxy login for caregivers. Although most systems have the potential to do this, most clinical settings do not outright offer it and/or facilitate the registration process.

For subtheme 3, *general portal security and technology concerns*, safety-net physicians made general comments about security or privacy issues with technology in health care. Providers worried that their patients and caregivers would not use a portal secondary to fears about their private data being accessed by outside parties (be *hacked*).

In the last major theme, *barriers to caregiver enrollment and use*, participants noted the challenges to enrolling caregivers as proxies into their patient's portal. Many of the barriers noted paralleled the general challenges to portal use among patients in the safety-net: decreased access to and familiarity with technology (low digital health literacy), lack of easy access to the internet at home, being LEP (when most of the portal is in English), fear of signing up because of immigration documentation status, lack of awareness about a patient portal in the safety-net, or lack of understanding of what a patient portal can do.

Discussion

Principal Findings

Patient portals and caregiver engagement in primary care both represent clear ways to improve our communication processes with vulnerable patients by making care more convenient and coordinated. However, even in large integrated delivery systems with established track records of portal use, there is very little research on the role that caregivers can play in terms of digital communication via the EHR. Our study echoes some of the prior literature recognizing the clear need for better functionality of portals for caregivers (specific content for them and better ways for patients to designate what types of information or access to share) [22-24]. We also shed light on some specific needs in this population, such as the needed support for caregivers with communication or language barriers (which is of high importance in safety-net settings).

XSL•FO
RenderX

Beyond the health care setting, there is already a *team* around many of our patients—family members and friends who serve as caregivers and trusted confidants when making health decisions. Health leaders must recognize the need to make it easier to connect with these trusted team members, in addition to the patient, when creating care plans together, especially as systems focus on the patient portal as a primary health management tool moving forward. The next version of portals can potentially improve the integration of telehealth options into the interface and provide resources that are specific to assisting caregivers, such as local resources, to reduce or prevent caregiver burnout. These innovations may be important for patients and caregivers in safety-net settings.

Participants brought up many important privacy and communication issues to resolve and improve to make portals work for safety-net caregivers and patients. Although most safety-net EHRs have the technical capability to create proxy account functionality [38], limited information exists regarding the use of shared access. Available data indicate that health system implementation of shared access functionality has been variable [39], but national estimates are that up to 30% of portal users have used the portal as a proxy for someone else (with about 50% of them as a formal caregiver/proxy login) [40,41]. Prior studies have evaluated patients' experiences with accessing and using their own patient portal account [21] and preferences for sharing their electronic health information with others [27,42,43]. These studies suggest that some patients would like the option of selecting the *level* at which a proxy caregiver has access to their complete medical record. As such, we found that addressing the privacy and security nuances of building a digital relationship with a caregiver in the safety-net was a major barrier to physician and provider participants being completely supportive of proxy patient portal use. In the safety-net, physicians have reported the need to provide limited portal views of *sensitive* information (eg, HIV results and prior intravenous drug use history) for patients who rely on family or friends as informal or formal caregivers. Many times, the option to use the portal is foregone by patients because of this current *all-or-nothing* access approach to patient information via a proxy login. One solution to this dilemma is to allow patients the ability to choose the level of proxy access a caregiver or family member will have. EHR vendors should be incentivized via health information policy to innovate, and safety-net health systems should advocate for more patient-centered options around *level of access* to health data via proxy relationships, especially when patient privacy and security remain a top level of concern among safety-net patients. This is a feature that should be available to all patients and will resonate with several groups, including adolescents, older adults, and patients with some sort of impairment or disability, in other health care settings.

Limitations and Strengths

Among the limitations of this formative study are the small sample size and generalizability to other safety-net health care systems, which are different from the California settings included in the study. However, the purpose of this qualitative study is to generate initial insights about provider buy-in regarding portal potential for caregivers in the safety-net.

This is one of the few studies to probe into provider perspectives around the role of caregivers in portal use within the safety-net. We found that providers recognized the potential for portals to improve information delivery and communication, benefiting caregivers who are assisting socially and medically complex patients in the safety-net. These safety-net health care providers and leaders focused on expanding the functionality of the portal, so that it does *more* for caregivers.

Conclusions

Overall, safety-net systems that seek to maximize the use of portals in their communities must develop specific strategies to uphold patient preferences and innovatively support caregivers while maintaining privacy and security. Further work is needed to assess and support diverse caregivers' access to portals by addressing pivotal barriers, such as health or digital literacy and English proficiency [44]. Safety-net health systems provide health care for our most medically and socially fragile patients: populations that include patients with multiple morbid conditions, limited English proficiency, cognitive impairment, physical and mental disabilities, low literacy, homelessness, substance use disorder, justice involvement, and immigrant and refugee status. As they serve the most vulnerable, safety-nets are the ideal places to develop and refine the next patient and caregiver-centered iterations of the EHR and the patient portal. However, to make this tool work for our most vulnerable patients, we must take intentional steps to ensure that the patient portal can be effectively and efficiently deployed by their caregivers in the systems that serve these high-risk patients.

Acknowledgments

The authors of this publication have conducted this research with support from the Agency for Healthcare Research and Quality (R00HS022408).

Conflicts of Interest

AC received funding from the Agency for Healthcare Research and Quality and the Patient-Centered Outcomes Research Institute (KL12HS026407). NR received funding from the Agency for Healthcare Research and Quality (K08HS022561). The remaining authors have no conflicts of interest.

References

XSL•FO
RenderX

1. Lyles CR, Aulakh V, Jameson W, Schillinger D, Yee H, Sarkar U. Innovation and transformation in California's safety net health care settings: an inside perspective. Am J Med Qual 2014;29(6):538-545 [FREE Full text] [doi: 10.1177/1062860613507474] [Medline: 24170938]
2. Househ MS, Borycki EM, Rohrer WM, Kushniruk AW. Developing a framework for meaningful use of personal health records (PHRs). Health Policy Technol 2014 Dec;3(4):272-280. [doi: 10.1016/j.hlpt.2014.08.009]
3. Kamo N, Bender AJ, Kalmady K, Blackmore CC. Meaningful use of the electronic patient portal - Virginia Mason's journey to create the perfect online patient experience. Healthc (Amst) 2017 Dec;5(4):221-226. [doi: 10.1016/j.hjdsi.2016.09.003] [Medline: 27727028]
4. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
5. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
6. Zickuhr K, Smith A. Digital Differences. Pew Research Center. 2018. URL: http://www.pewinternet.org/2012/04/13/digital-differences/ [accessed 2012-04-13]
7. Gonzalez M, Sanders-Jackson A, Emory J. Online health information-seeking behavior and confidence in filling out online forms among Latinos: a cross-sectional analysis of the California health interview survey, 2011-2012. J Med Internet Res 2016 Jul 4;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
8. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
9. Lyles CR, Tieu L, Sarkar U, Kiyoi S, Sadasivaiah S, Hoskote M, et al. A randomized trial to train vulnerable primary care patients to use a patient portal. J Am Board Fam Med 2019;32(2):248-258 [FREE Full text] [doi: 10.3122/jabfm.2019.02.180263] [Medline: 30850461]
10. Ochoa A, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. J Am Med Inform Assoc 2017 Nov 1;24(6):1160-1164. [doi: 10.1093/jamia/ocx040] [Medline: 28460130]
11. Dhanireddy S, Walker J, Reisch L, Oster N, Delbanco T, Elmore JG. The urban underserved: attitudes towards gaining full access to electronic medical records. Health Expect 2014 Oct;17(5):724-732 [FREE Full text] [doi: 10.1111/j.1369-7625.2012.00799.x] [Medline: 22738155]
12. Lobach DF, Willis JM, Macri JM, Simo J, Anstrom KJ. Perceptions of medicaid beneficiaries regarding the usefulness of accessing personal health information and services through a patient internet portal. AMIA Annu Symp Proc 2006:509-513 [FREE Full text] [Medline: 17238393]
13. Mirsky JB, Tieu L, Lyles C, Sarkar U. A mixed-methods study of patient-provider e-mail content in a safety-net setting. J Health Commun 2016;21(1):85-91 [FREE Full text] [doi: 10.1080/10810730.2015.1033118] [Medline: 26332306]
14. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
15. When Family Members Care for Aging Parents. The New York Times. 2018. URL: https://www.nytimes.com/2018/09/19/well/live/when-family-members-care-for-aging-parents.html [accessed 2019-09-01]
16. Caregiving in the US. The National Alliance for Caregiving (NAC) and the AARP Public Policy Institute. 2015. URL: https://www.aarp.org/content/dam/aarp/ppi/2015/caregiving-in-the-united-states-2015-report-revised.pdf [accessed 2020-06-04]
17. Wolff JL, Roter DL. Family presence in routine medical visits: a meta-analytical review. Soc Sci Med 2011 Mar;72(6):823-831 [FREE Full text] [doi: 10.1016/j.socscimed.2011.01.015] [Medline: 21353358]
18. Smith SK, Dixon A, Trevena L, Nutbeam D, McCaffery KJ. Exploring patient involvement in healthcare decision making across different education and functional health literacy groups. Soc Sci Med 2009 Dec;69(12):1805-1812. [doi: 10.1016/j.socscimed.2009.09.056] [Medline: 19846245]
19. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet personal health record pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
20. Mayberry LS, Kripalani S, Rothman RL, Osborn CY. Bridging the digital divide in diabetes: family support and implications for health literacy. Diabetes Technol Ther 2011 Oct;13(10):1005-1012 [FREE Full text] [doi: 10.1089/dia.2011.0055] [Medline: 21718098]
21. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
22. Colorafi K, Greenes RA, Kates M. Preferences of older adults and their families for meaningful use clinical summaries. Mhealth 2018;4:8 [FREE Full text] [doi: 10.21037/mhealth.2018.03.04] [Medline: 29682511]

XSL•FO
RenderX

23. Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller DP, Smith D, et al. Insights into older adult patient concerns around the caregiver proxy portal use: qualitative interview study. J Med Internet Res 2018 Nov 2;20(11):e10524 [FREE Full text] [doi: 10.2196/10524] [Medline: 30389654]
24. Reed ME, Huang J, Brand R, Ballard D, Yamin C, Hsu J, et al. Communicating through a patient portal to engage family care partners. JAMA Intern Med 2018 Jan 1;178(1):142-144 [FREE Full text] [doi: 10.1001/jamainternmed.2017.6325] [Medline: 29159402]
25. Portz JD, Fruhauf C, Bull S, Boxer RS, Bekelman DB, Casillas A, et al. 'Call a teenager… that's what I do!' - grandchildren help older adults use new technologies: qualitative study. JMIR Aging 2019 Jun 6;2(1):e13713 [FREE Full text] [doi: 10.2196/13713] [Medline: 31518274]
26. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
27. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
28. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
29. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. 'I want to keep the personal relationship with my doctor': understanding barriers to portal use among African Americans and Latinos. J Med Internet Res 2016 Oct 3;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
30. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
31. Hay JW, Katon WJ, Ell K, Lee P, Guterman JJ. Cost-effectiveness analysis of collaborative care management of major depression among low-income, predominantly Hispanics with diabetes. Value Health 2012;15(2):249-254 [FREE Full text] [doi: 10.1016/j.jval.2011.09.008] [Medline: 22433755]
32. Wu S, Ell K, Gross-Schulman SG, Sklaroff LM, Katon WJ, Nezu AM, et al. Technology-facilitated depression care management among predominantly Latino diabetes patients within a public safety net care system: comparative effectiveness trial design. Contemp Clin Trials 2014 Mar;37(2):342-354. [doi: 10.1016/j.cct.2013.11.002] [Medline: 24215775]
33. Huckfeldt PJ, Meeker D, Peters A, Guterman JJ, Diaz G, Goldman DP. Diabetes management for low-income patients in Los Angeles: two strategies improved disease control in the short term. Health Aff (Millwood) 2012 Jan;31(1):168-176 [FREE Full text] [doi: 10.1377/hlthaff.2011.0930] [Medline: 22232107]
34. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
35. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
36. Omole FS, Sow CM, Fresh E, Babalola D, Strothers H. Interacting with patients' family members during the office visit. Am Fam Physician 2011 Oct 1;84(7):780-784 [FREE Full text] [Medline: 22010616]
37. Mitnick S, Leffler C, Hood VL, American College of Physicians Ethics, Professionalism and Human Rights Committee. Family caregivers, patients and physicians: ethical guidance to optimize relationships. J Gen Intern Med 2010 Mar;25(3):255-260 [FREE Full text] [doi: 10.1007/s11606-009-1206-3] [Medline: 20063128]
38. Green-Shook S. Parental proxy access via web portals: ensuring compliance and quality documentation. J AHIMA 2009;80(7):60 [FREE Full text]
39. Policy: Privacy &Security Tiger Team. HealthIT. 2014. URL: http://www.healthit.gov/FACAS/calendar/2014/02/10/policy-privacy-security-tiger-team [accessed 2014-06-01]
40. Sarkar U, Bates DW. Care partners and online patient portals. J Am Med Assoc 2014;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
41. Osborn CY, Rosenbloom ST, Stenner SP, Anders S, Muse S, Johnson KB, et al. MyHealthAtVanderbilt: policies and procedures governing patient portal functionality. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i18-i23 [FREE Full text] [doi: 10.1136/amiajnl-2011-000184] [Medline: 21807648]
42. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 1;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]
43. Crotty BH, Walker J, Dierks M, Lipsitz L, O'Brien J, Fischer S, et al. Information sharing preferences of older patients and their families. JAMA Intern Med 2015 Sep;175(9):1492-1497. [doi: 10.1001/jamainternmed.2015.2903] [Medline: 26147401]
44. Grossman LV, Choi SW, Collins S, Dykes PC, O'Leary KJ, Rizer M, et al. Implementation of acute care patient portals: recommendations on utility and use from six early adopters. J Am Med Inform Assoc 2018 Apr 1;25(4):370-379. [doi: 10.1093/jamia/ocx074] [Medline: 29040634]

XSL•FO
RenderX

Abbreviations

AHS: Alameda Health System
EHR: electronic health record
IT: information technology
LAC DHS: Los Angeles County Department of Health Services
LEP: limited English proficient
SFHN: San Francisco Health Network

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by C Kruse, C Latulipe; comments to author 28.03.20; revised version received 13.05.20; accepted 14.05.20; published 20.07.20.

Please cite as:
Casillas A, Cemballi AG, Abhat A, Lemberg M, Portz JD, Sadasivaiah S, Ratanawongsa N, Semere W, Brown A, Lyles CR
An Untapped Potential in Primary Care: Semi-Structured Interviews with Clinicians on How Patient Portals Will Work for Caregivers in the Safety Net
J Med Internet Res 2020;22(7):e18466
URL: https://www.jmir.org/2020/7/e18466
doi:10.2196/18466
PMID:32706709

©Alejandra Casillas, Anupama Gunshekar Cemballi, Anshu Abhat, Miya Lemberg, Jennifer D Portz, Shobha Sadasivaiah, Neda Ratanawongsa, Wagahta Semere, Arleen Brown, Courtney Rees Lyles. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient Portal Implementation and Uptake: Qualitative Comparative Case Study

Melita Avdagovska[1], MA; Mark Ballermann[2], PhD; Karin Olson[3], RN, PhD; Timothy Graham[4], MD; Devidas Menon[1], PhD; Tania Stafinski[1], PhD

[1]University of Alberta, School of Public Health, Edmonton, AB, Canada

[2]University of Alberta, Faculty of Medicine and Dentistry, Edmonton, AB, Canada

[3]University of Alberta, Faculty of Nursing, Edmonton, AB, Canada

[4]Alberta Health Services, Edmonton, AB, Canada

Corresponding Author:
Melita Avdagovska, MA
University of Alberta
School of Public Health
3021, Research Transition Facility
8308 114 Street
Edmonton, AB, T6G 2V2
Canada
Phone: 1 7809083334
Email: avdagovs@ualberta.ca

Abstract

Background: Giving patients access to their health information is a provincial and national goal, and it is critical to the delivery of patient-centered care. With this shift, patient portals have become more prevalent. In Alberta, the Alberta Health Services piloted a portal (MyChart). There was a need to identify factors that promoted the use of this portal. Furthermore, it was imperative to understand why there was variability in uptake within the various clinics that participated in the pilot.

Objective: This study aims to identify potential factors that could improve the uptake of MyChart from the perspectives of both users and nonusers at pilot sites. We focused on factors that promoted the use of MyChart along with related benefits and barriers to its use, with the intention that this information could be incorporated into the plan for its province-wide implementation.

Methods: A qualitative comparative case study was conducted to determine the feasibility, acceptability, and initial perceptions of users and to identify ways to increase uptake. Semistructured interviews were conducted with 56 participants (27 patients, 21 providers, 4 nonmedical staff, and 4 clinic managers) from 5 clinics. Patients were asked about the impact of MyChart on their health and health care. Providers were asked about the impact on the patient-provider relationship and workflow. Managers were asked about barriers to implementation. The interviews were recorded, transcribed verbatim, and entered into NVivo. A thematic analysis was used to analyze the data.

Results: Results from a comparison of factors related to uptake of MyChart in 5 clinics (2 clinics with high uptake, 1 with moderate uptake, 1 with low uptake, and 1 with no uptake) are reported. Some theoretical constructs in our study, such as intention to use, perceived value, similarity (novelty) of the technology, and patient health needs, were similar to findings published by other research teams. We also identified some new factors associated with uptake, including satisfaction or dissatisfaction with the current status quo, performance expectancy, facilitating conditions, behavioral intentions, and use behavior. All these factors had an impact on the level of uptake in each setting and created different opportunities for end users.

Conclusions: There is limited research on factors that influence the uptake of patient portals. We identified some factors that were consistent with those reported by others but also several new factors that were associated with the update of MyChart, a new patient portal, in the clinics we studied. On the basis of our results, we posit that a shared understanding of the technology among patients, clinicians, and managers, along with dissatisfaction with nonportal-based communications, is foundational and must be addressed for patient portals to support improvements in care.

(J Med Internet Res 2020;22(7):e18973) doi:10.2196/18973

XSL•FO
RenderX

KEYWORDS

patient portals; information technology; adoption

Introduction

Background

The rising costs of health care services and the increased prevalence of chronic conditions necessitate transformative changes in health care delivery. Deliberations on these changes have focused on access to information and the interconnectedness of health information systems [1]. Providing patients secure access to their health information is a provincial [2-5] and national goal in Canada [6,7], and it is considered critical to the delivery of patient- and family-centered care [2-4]. Surprisingly, many portals have underperformed expectations in Canada [8] and in other jurisdictions [9,10]. Although patients today have greater access to their medical information than ever before, there is a need to identify the best way to provide this information and to understand the impact of the provision of medical information to patients on the health care providers.

To address the issues mentioned above, the Alberta Health Services (AHS), a province-wide health delivery organization, piloted a patient portal called eCLINICIAN MyChart (an AHS branding of EpicCare Ambulatory from Epic Systems, 2014 version) between 2016 and 2019 [11]. The goals of the study were to determine whether MyChart would help patients and their families participate actively in the maintenance and monitoring of their health information and to provide information that could be used to scale up the use of MyChart by documenting factors that both hindered and promoted MyChart uptake.

Objectives

We aimed to accomplish this goal by collecting data from both users and nonusers at the pilot sites. We focused on factors that promoted the use of MyChart along with the related benefits and barriers to its use, with the intention that this information could be incorporated into the plan for its province-wide implementation. To the best of our knowledge, this is the first study that has considered the perspectives not only of patients and health care providers but also of clinic managers and nonmedical staff, such as receptionists, on the impact of patient portal uptake. The timing of data collection began before the introduction of the portal to identify expectations held by health care providers' practitioners and managers. This approach to data collection was used to identify previously unknown or unanticipated factors that could limit the potential of patient portals for health care systems, including previously obscure power dynamics among clinical leadership. Portals are likely to have systemic groups of nonusers for reasons that have not been previously articulated. In this study, we sought to clarify why portals can, at times, fail to achieve their transformative potential. In parallel with this work, a quantitative survey was conducted on patient views of the portal [12], and the views of the clinicians and managers presented here complement those findings.

Methods

Research Question

The research question was as follows: *What factors impacted the differences in uptake of the patient portal MyChart and what elements were deemed crucial for uptake and access for all end users*?

This question was developed by incorporating the principles of *diffusion of innovation* by Rogers [13] and the *theory of technology acceptance* by Davis [14]. Roger's theory focuses on explaining how and why new ideas and technologies spread, whereas Davis' s theory focuses on the perceived usefulness of a technology to enhance performance [13,14]. By incorporating these 2 frameworks, we were able to consider the various theoretical approaches relevant when considering the uptake of a patient portal within a publicly funded health care system.

Ethics Approval

This study received ethics approval from the Research Ethics Board at the University of Alberta (Study ID: MS6_Pro00084135) and operational approval for the clinics included in this study from AHS (OA43157, OA43158, OA43159, and OA43160).

Design

Using a qualitative comparative case study design [15-17], we performed 56 in-depth semistructured interviews with clinic managers, health care providers, nonmedical providers (receptionists), and patients from 5 clinics that participated in the pilot. Comparative case studies involve the analysis and synthesis of similarities, differences, and patterns across 2 or more cases that share a common focus or goal [15]. In this study, each of the 5 clinics constituted a case. A comparison of cases provided an opportunity to learn more about factors that played a role in the uptake during a real-time patient portal implementation from the standpoint of study participants rather than from existing frameworks.

Study Setting

Alberta, Canada, has a publicly funded and managed health care system based on the principles of universal access to medically necessary services. AHS [18] is the largest provider of health care services in Alberta, delivering care to more than 4 million people. In 2016, AHS introduced a patient portal called MyChart that was connected to the central electronic medical record system called eCLINICIAN. This was carried out through a pilot where several self-selected clinics decided to participate; however, individual health care providers within the clinics were free to participate or not. MyChart allowed patients to view appointments, medical test results, and medication therapies and to communicate with their health care providers through a computer or an app on a phone or tablet. It was developed by the Epic Systems Corporation (Epic) and customized to fit the needs of AHS. The pilot was led by the eCLINICIAN MyChart Working Group, comprising physicians

XSL•FO
RenderX

from both family medicine and various specialties, technology experts, policy makers, and health care administrators.

Following ethics approval, 5 clinics in the pilot were selected for this study based on the number of patients enrolled in MyChart per clinic and the length of time each clinic had been using it. These clinics were considered minimal and safe users of the nonportal functionalities of the Epic product (eg, eCLINICIAN). Three of the clinics were specialist clinics, 1 clinic was a community-based family practice clinic, and 1 clinic was a family practice clinic hosted within an AHS facility.The AHS family practice clinic was selected because although in principle they agreed to participate, they never enrolled any patients. The key features of the cases are outlined in Table 1.

Table 1. Case study characteristics.

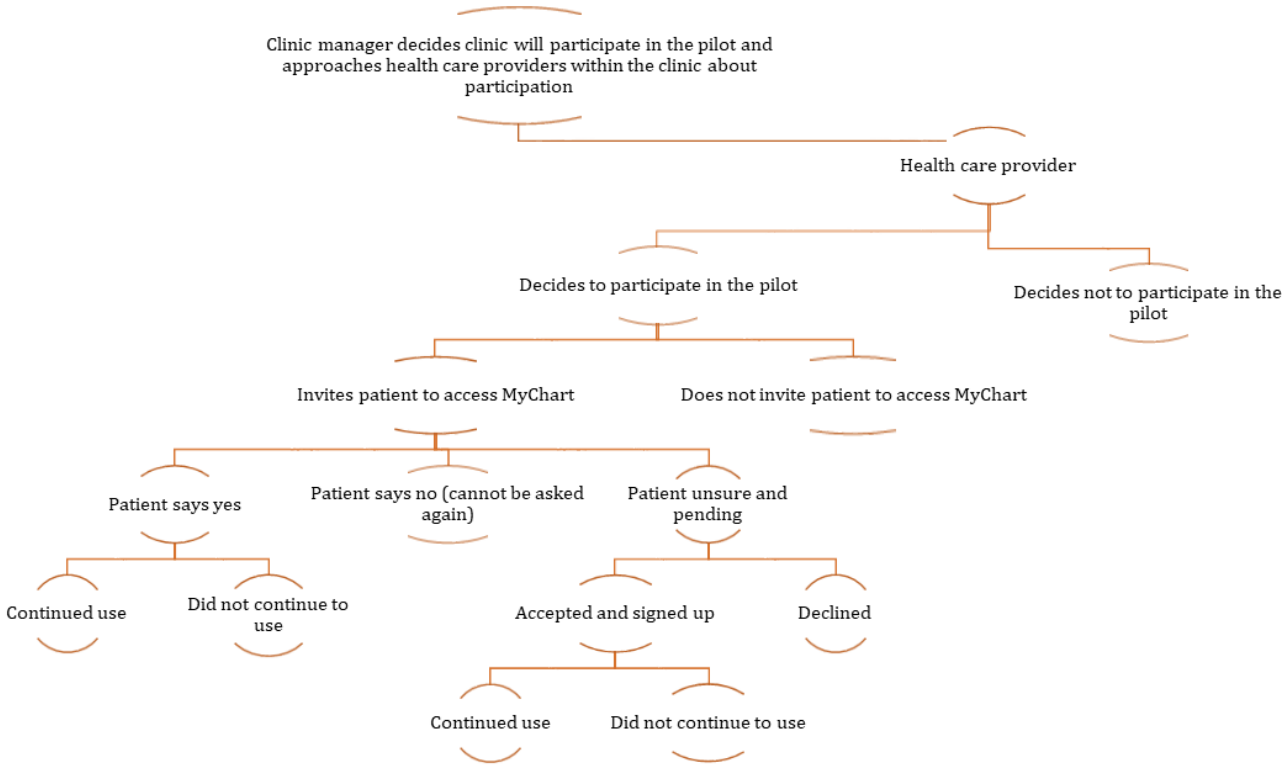
Details of the clinic	Type of setting	MyChart patient enrollment, n[a]
Clinic 1: medium user of MyChart	Specialty clinic in academic setting	357
Clinic 2: high user of MyChart	Specialty clinic in academic setting	875
Clinic 3: low user of MyChart	Specialty clinic in academic setting	172
Clinic 4: high user of MyChart	General practice clinic in community setting	965
Clinic 5: no users of MyChart	General practice clinic in academic setting	0

[a]As reported in April 2019.

Sample and Recruitment Strategy

Before deciding on the required study sample, the project team had discussions with the MyChart project team and the staff of several clinics in the pilot to understand how its use evolved in each setting. On the basis of those conversations, MA developed a participant recruitment framework (Figure 1), which was reviewed by KO, TS, DM, and MB and then finalized.

Figure 1. Participant recruitment framework.



On the basis of this framework, we decided to recruit clinic managers and health care providers who did or did not participate in the pilot and patients who used or did not use MyChart, including proxy users. MyChart permits proxy access to family members who provide care to their loved ones.

Clinic managers and health care providers were recruited using a purposive sampling [19] approach. TG contacted each identified clinic, requested an introductory meeting with the clinic manager, and responded to any follow-up concerns. The clinic managers invited health care providers (specialists, general practitioners, and/or nurses) to participate in these introductory meetings. MA developed the presentation and recruitment materials. DM and TS reviewed the materials. MA and MB attended these meetings, described the study, and discussed recruitment strategies for health care providers and patients within each clinic. After answering any questions, MA obtained

XSL•FO
RenderX

written consent from health care providers and other clinic staff who were willing to participate in the study. MA set up interviews with clinic managers and health care providers. During the interviews, we identified that only 1 clinic implemented the MyChart web-based booking and canceling appointments function, so all 4 receptionists (nonmedical providers) in that clinic were interviewed.

Patients were recruited using a recruitment card (Multimedia Appendix 1), which was handed out by health care providers at each clinic over a 2-week period. They had provided the card to all patients they saw during those 2 weeks. Patients contacted MA directly if they were interested in participating.

Data Collection

MA scheduled interviews and explained the study, answered questions, and obtained written consent from those who agreed to participate. The participants then completed a demographic information form and were assigned a deidentifying code used to maintain anonymity and track data. All recruited participants took part in an individual semistructured interview (conducted by MA face-to-face or by telephone) lasting between 30 and 45 min. Notes and reflections were written by MA immediately after each interview. The semistructured interview guides (Multimedia Appendix 2) were developed by MA, based on the research question, and then were reviewed by the other authors and revised based on feedback. To determine whether the interview guides required adjustments, MA evaluated the recordings of the first 2 interviews per group. No changes were made. All interviews were recorded digitally, transcribed verbatim, checked for accuracy, and then uploaded to NVivo 11 (QSR International) to facilitate analysis. Data collection and analysis occurred concurrently until no new dimensions of the participants' experiences were identified [20].

Data Analysis

The data collected guided the structure of the subsequent analyses, following the principles of inductive thematic analysis [21-23]. MA cleaned all transcripts and then read each transcript several times, coded transcripts for key ideas related to the research question, and grouped codes into preliminary categories. KO reviewed all codes and categories and discussed them with MA. Preliminary categories across cases and participant groups were organized according to recurring and emerging themes by MA and reviewed by KO and TS. As is common in qualitative research, the findings were built from the data [24]. The findings were then compared with those of other studies.

Several processes were undertaken to maintain the trustworthiness of the data [23], including establishing credibility by recruiting a diverse sample with different and sometimes opposing perspectives and maintaining a comprehensive audit trail in NVivo 11. To ensure transferability of the data, detailed descriptions of the sample and thick descriptions of findings were developed. Analytic rigor was enhanced through regular meetings of the research team to discuss emerging findings. As each interview was completed, all available interviews were reviewed collectively. Data saturation was reached after 25 patient and 20 health care provider interviews, but 2 more patients and 1 more provider were interviewed to ensure that important information was not missed. The demographic data were analyzed descriptively and added as case nodes in NVivo to explore factors associated with the adoption of MyChart.

Results

Demographic Characteristics

Our plan was to recruit patients who would fit one of the following categories: user (has access to MyChart and has used it at least once in the past 6 months), declined (offered but declined to sign up), did not continue to use (signed up but did not use), and proxy access user. We were able to recruit only 1 patient who declined and none who did not continue to use, as we had no means of obtaining information about discontinued usage.

We also recruited 1 caregiver who used MyChart under the proxy provisions and several caregivers who used the family member accounts. Table 2 shows the patient respondents' demographics and reported conditions.

We were able to recruit several health care providers who were nonusers (1 provider elected not to participate in the pilot, 1 medical resident who was not aware that MyChart was part of the clinic where he or she was working, and 3 medical office assistants who were familiar with MyChart; however, as a result of the implementation process for MyChart in that particular clinic, they were not involved directly with MyChart). The specialty physicians were endocrinologists (n=3), neurologists (n=5), and gastroenterologists (n=5). Table 3 shows the health care providers' and clinic managers' demographic information.

Overall, 56 participants were recruited (Table 4; to see interview codes and the demographic characteristics per participant group, see Multimedia Appendices 3 and 4).

XSL•FO
RenderX

Table 2. Demographics of patient respondents and the reported conditions (n=27).

Characteristics	Value, n (%)
Gender	
Female	17 (62)
Male	10 (37)
Age (years)	
18-25	1 (3)
26-35	1 (3)
36-45	0 (0)
46-60	16 (59)
≥61	9 (33)
Marital status	
Never legally married	2 (7)
Legally married and not separated	18 (66)
Separated, but still legally married	0 (0)
Divorced	2 (7)
Common law	4 (14)
Widowed	1 (3)
Education level	
Less than high school degree	0 (0)
High school degree or equivalent	5 (18)
Some postsecondary education but no degree	7 (25)
Registered apprenticeship or other trades certificate or diploma	5 (18)
Associate degree	1 (3)
Bachelor's degree	7 (25)
Graduate degree	0 (0)
Postgraduate degree	2 (7)
Employment status	
Employed, working ≥40 hours per week	9 (33)
Employed, working 1-39 hours per week	7 (25)
Not employed, looking for work	0 (0)
Not employed, not looking for work	2 (7)
Retired	4 (14)
Unable to work	3 (11)
Self-employed	2 (7)
Family income (Can $)	
<20,000 (US $14,723)	0 (0)
20,000-34,999 (US $14,723-25,765)	1 (3)
35,000-49,999 (US $25,766-34,806)	2 (7)
50,000-74,999 (US $36,807-55,210)	1 (3)
75,000-99,999 (US $55,211-73,614)	4 (14)
100,000-149,999 (US $73,615-110,422)	8 (29)
≥150,000 (US $110,423)	11 (40)
Chronic condition	

XSL•FO
RenderX

Characteristics	Value, n (%)
Yes[a]	23 (85)
No	4 (14)
MyChart users	
Yes	25 (92)
No	2 (7)
And/or proxy	5 (18)

[a]Chronic conditions reported by the participants included arthritis, Barrett syndrome, bleeding disorder, breast cancer, celiac disease, chronic obstructive pulmonary disease, Crohn disease, depression or anxiety, diabetes type 1, diabetes type 2, high blood pressure, hypothyroid, inflammatory bowel syndrome, irregular heartbeat, Langerhans cell histiocytosis, multiple sclerosis, osteoporosis, prediabetes, psoriasis, relapsing-remitting multiple sclerosis, sleep apnea, systemic mastocytosis, thyroid disease, and ulcerative colitis.

Table 3. Demographic information of the health care providers (n=21) and clinic managers (n=4).

Characteristics	Health care providers' demographics, n (%)	Clinic managers' demographics, n (%)
Gender		
Female	16 (76)	3 (75)
Male	5 (23)	1 (25)
Age (years)		
<18	0 (0)	0 (0)
18-29	0 (0)	0 (0)
30-39	2 (9)	0 (0)
40-49	8 (38)	2 (50)
50-59	8 (38)	1 (25)
60-64	2 (9)	1 (25)
≥65	1 (4)	0 (0)
Type of		
Family physician	2 (9)	1 (25)
Specialty physician	5 (23)	2 (50)
Registered nurse	8 (38)	1 (25)
Nurse practitioner	1 (4)	0 (0)
Medical office assistant	4 (19)	0 (0)
Resident	1 (4)	0 (0)
Working at		
Academic-based care	14 (66)	3 (75)
Community-based care	7 (33)	0 (0)
Both	0 (0)	0 (0)
In which type of setting do you work		
Community clinic or health center	8 (38)	1 (25)
Specialty clinic	13 (61)	3 (75)

XSL•FO
RenderX

Table 4. Interviewed participants per case.

Interviewees	Clinic manager, n	Health care providers, n	Nonmedical staff, n	Patients, n	Caregiver, n
Clinic 1					
MyChart user	1 (female)	3 (2 female and 1 male)	0	4 (female)	1 (male)
MyChart nonuser	0	0	0	0	0
Clinic 2					
MyChart user	1 (female)	5 (female)	0	5 (female)	N/A[a]
MyChart nonuser	0	0	0	0	0
Clinic 3					
MyChart user	1 (female)	4 (female)	0	4 (3 female and 1 male)	N/A
MyChart nonuser	0	1 (male)	0	1 (male)	N/A
Clinic 4					
MyChart user	1 (male)	2 (1 female and 1 male)	4 (female)	11 (5 female and 6 male)	N/A
MyChart nonuser	0	5 (4 female and 1 male)	0	1 (male)	N/A
Clinic 5					
MyChart user	0	0	0	0	N/A
MyChart nonuser	0	1 (male)	0	0	N/A

[a]N/A: not applicable.

Themes

We coded each group of interviewees (clinic managers, health care providers, nonmedical staff, and patients) separately. As we were reviewing the transcripts and coding them, it became evident that the concerns raised by nonusers were also raised by users of MyChart. Therefore, information collected from users and nonusers was aggregated. Textbox 1 shows the identified themes.

Textbox 1. Themes per sample group.

Patients

- Theme 1: My health, my responsibility, but I need the information to do that
- Theme 2: Convenience, convenience, convenience
- Theme 3: A few tweaks will do it
- Theme 4: Do not take it away

Health care providers

- Theme 1: Of course we need it
- Theme 2: Yes, it did some good things
- Theme 3: Not all was great
- Theme 4: Consider my needs

Clinic managers

- Theme 1: Time for some efficiency in communication
- Theme 2: Heard great things about it
- Theme 3: Heard about some issues as well
- Theme 4: Change is needed

Nonmedical providers

- Theme 1: Great service for the patient
- Theme 2: Needs some education
- Theme 3: Need to maintain and expand

XSL•FO
RenderX

The themes from each sample group are described in more detail below, with direct quotations used to support the analysis.

Patients

A total of 27 patients from 4 clinics were interviewed. Four themes emerged from the interviews: (1) my health, my responsibility, but I need the information to do that; (2) convenience, convenience, convenience; (3) a few tweaks will do it; and (4) do not take it away.

Theme 1: My Health, My Responsibility, but I Need the Information To Do That

When initially offered, patients were unsure what to expect from a system that would give them access to their own health care information—information that had been difficult to obtain previously. Furthermore, participants noted that the introduction to MyChart was not as convincing as it might have been, thus they were unsure if there was any benefit for them. Patients decided to enroll and use MyChart based on their interest in being more proactive in their own health:

> *First, I am convinced that my health is MY responsibility, not the physician or treatment team. They are here to advise and instruct and carry out detailed procedure that I accept. I must be convinced of the efficacy of treatment options. I am not a blind follower of absolute directions from the all-knowing care-teams. I always try to enter into a meaningful dialogue about treatment options and expected results, and I expect the treatment teams to deal with me professionally and personally.* [1PAT3]

Patients noted that seeing their medical results did not cause them to be anxious or call the doctor's office more, as had been expected by their health care providers. Patients were genuinely surprised that the system actually maintained records of their health information. Some were surprised to learn things that they were uninformed about previously.

This theme centered on the importance of having access to credible information that "helps me look after me" (3PAT1). Having access to the information allowed patients to make more informed choices. Empowerment and comfort were also considered important:

> *It has been empowering for me and that also feel that because of it, I am burdening the system a lot less. Fewer appointments, fewer phone calls...and also I feel not just burdening the system but I'm better able to take care of myself.* [3PAT1]

Patients no longer accepted the modus operandi that *no news, is good news*. They wanted to know whether their results were *good* or *bad*. Having access to their medical records was seen as a necessity, as patients were already trying to collect and maintain it by requesting copies from their health care providers. Patients expressed frustration with the fact that patient portals exist elsewhere and regarded the process of implementation in Alberta to be slow. Overall, patients felt that they needed this information to be responsible for their own health.

Theme 2: Convenience, Convenience, Convenience

The term *convenience* was used to describe a missing element in health care and the contribution of MyChart to solving this problem. Regardless of where the patient was, they liked being able to connect with their health care team through MyChart. Just knowing that they could reach out to their health care provider was considered a safety net. Being able to contact their health care provider directly through an email system and receive prompt replies was appreciated. Patients indicated that, from their point of view, they did not abuse this system and they only sent messages when needed. Although patients did not understand all their test results, they saw the benefits of having the information and used it to frame more specific questions for their health care providers and to discuss treatment plans. Most of the time, the provided ranges of results were sufficient for patients to understand their results. Furthermore, some patients had been performing the same tests over the years, thus they were already aware of how to interpret their test results.:

> *I may not understand everything, in fact I understand very little but this information enables me to ask a smarter question.* [4PAT2]

Patients wanted the same convenience in interactions with their health care provider as for other services such as banking and shopping. The ability to make and cancel appointments web-based was described as extremely convenient and valuable. One patient explained that "you should not have human resources doing the work that can be done with clicks" (4PAT2). Another patient described the ability to make appointments web-based as *freedom*. In addition, having the ability to fill out questionnaires before going to the appointment was seen as beneficial, as patients had the ability to take the time and carefully think about the answers. Caregivers used the information in MyChart to monitor the health of their family members.

Patients correlated the convenience of MyChart with cost saving, as patients were able to see their results without making unnecessary follow-up appointments, did not have to travel for a short appointment, did not have to take time off from work, and did not have to pay for parking and/or pay to get their results printed. They saw this not only as cost savings to them but also as cost savings to the health care system.

Although proxy access was one of the most challenging elements to set up within MyChart, patients noted that it was one of the most beneficial connections, especially when caring for a family member:

> *Now my father who is 90 is a different story and one of the very interesting and most important features of this is the availability we have to see dad's test results. So, I can access because you know my dad and I work together, and this is maybe kind of not how they thought this was going to work, but as a primary caregiver with my dad, I've got his user ID and password, which I set up for him anyway.* [4PAT3]

XSL•FO
RenderX

Patients who had provided proxy access to their accounts were cognizant of the possible issues, but they found that the benefits outweighed the risks. Patients in our study did not seem to be greatly concerned about privacy and security. They felt that this system was as secure as any other system that had their personal information:

> *Well there's a concern every time I pull out my debit card or my visa. There is a concern every time I go on Instagram. There's a concern with every single thing we do and I could choose to be paralyzed by that and not have a credit card, and not go online, not do anything. But I choose to do all those things to enjoy the benefits, I guess. It's I suppose convenience and you use reasonable care, I think. And you kind of recognize that things can be hacked into and data can be lost.* [4PAT4]

Theme 3: A Few Tweaks Will Do It

Although patients found MyChart very beneficial, they also described some of the barriers that they experienced with the system. Some patients expressed concerns about their ability to understand the results, as they did not find the descriptions in MyChart valuable at all times. Diagnostic reports were not accompanied with explanations; thus it was difficult for patients to interpret the findings. Some patients did not appreciate the fact that health care providers did not see them as sufficiently competent to understand the information:

> *But you know the medical professionals are God and everybody else bows before them and does whatever they say. You know they're the only ones that can interpret this stuff properly.* [4PAT4]

Patients expressed dissatisfaction about the inability to see the entire medical record, including the doctor's notes.

No education or training on how to use MyChart was provided. As they did not know what their use of MyChart would entail, some patients decided not to sign up, noting that it might take "too much time to learn the things" (4PAT11 [NU]). Patients were frustrated about the lack of access to most MyChart functions. Patients in the clinics without the option to book web-based appointments defined this gap as a barrier:

> *I can book my flights, I can book my hotels, I can do it all on an app on my smartphone. I can book my lab, why can't I book my doctor's appointment.* [4PAT4]

Patients also did not understand why all their family members could not be on the system.

The sign-up and the sign-in process were barriers to some patients. One person said, "I mean I understand that they need a certain security level, I totally understand that, but it was so delicate and so picky that it often did'nt work" (4PAT1). Although MyChart was seen as a beneficial tool that enhanced the patient-provider relationship, some expressed concerns about health care being computerized and what that would mean for patients.

Theme 4: Do Not Take It Away

Many of the interviewed patients expressed concerns about the possibility that MyChart might not be continued after the pilot and said it would be like going "back to the Stone Ages" (4PAT4). They said they were not being informed on regular bases about the future of MyChart. Furthermore, patients explained that if MyChart was removed, they would have to go back to playing phone tag with their providers:

> *In this world where we have access to so much information and can make choices and view and control things online. My expectation is that I should be able to do that with my healthcare. And my experience is it's probably giving me better healthcare more efficiently than the old way.* [4PAT4]

Participants wanted to inform decision makers regarding the value of the patient portal and their wish for it to be maintained, and they had some suggestions for improvements (see Multimedia Appendix 4 for additional quotes).

Health Care Providers

A total of 21 health care providers from 5 clinics were interviewed. Four themes emerged from the interviews: (1) of course we need it; (2) yes, it did some good things; (3) not all was great; and (4) consider my needs.

Theme 1: Of Course We Need It

We commenced the interviews with health care providers by soliciting opinions about their perceptions of MyChart and their experience of its implementation. Many expressed positive views as they saw it as a good vehicle for delivering information to patients:

> *I think it's the patient's information. And unless there's a really clear medical reason, like maybe they have severe anxiety, mental health issues, or something else that maybe they shouldn't have it. I don't see any reason why they shouldn't have that information based on pretty much real time unless there are a certain subset of things that possibly might qualify for that maybe it's a cancer diagnosis.* [4HCP1]

One described MyChart as "a piece of our healthcare system that's been missing" (1HCP2). Furthermore, many providers agreed that patients had the right to their information and said portals were a great tool for that purpose. Providers thought MyChart had a positive impact on both themselves and their patients, "I can say fairly from my patients that experience is always been positive and been positive to me" (1HCP2).

Theme 2: Yes, It Did Some Good Things

Health care providers described several benefits from the MyChart for patients, "by learning about their health even and that sort of thing that they can click on into a health portal and learn about their condition or what comes up, or surgeries or whatever" (1HCP1). Providing patients with information improves patient visits. One provider said, "when you are armed with the results and armed with some of those values, it will make the patient interviews and patient visits a bit more meaningful" (4HCP2). MyChart was seen as a secure system where everything was recorded, as even a nonuser provider stated, "everything's documented so you don't have to worry" (4HCP4 [NU]).

XSL•FO
RenderX

Providers stated that MyChart was desirable as it reduced phone calls and provided an alternative means of communication. One of its novel functions was 2-way secure communication for nonemergency issues. This function required many modifications and decisions because it had not been tested before the pilot, and participating clinicians had expressed skepticism. It was also one of the main reasons many health care providers within the participating clinics decided not to participate in the pilot. Providers assumed that they would be inundated with a high number of messages, and there was no payment plan in place to remunerate them for the time required to answer these messages. As patients on MyChart started calling the clinic less frequently to get results, health care providers recognized the improvement in workflow, "instead of them calling us, they're just sending us a message through MyChart, which is easier for us on our end" (2HCP4).

Although only 1 clinic during the pilot decided to implement the appointments function for their patients, it proved to be extremely valuable as patients could "do it at their convenience to save some calling back and forth" (4HCP1). Providing proxy access was seen as a much-needed innovation in delivering health care. It was a new function that was never offered in Alberta before the MyChart pilot. Providers were positive about the approach. The most common request for proxy was for a spouse or adults caring for their elderly parents. One provider said, "the wife is the one that like keeps track of all the health information or like the different results and medications and whatnot so they're usually the ones that email about their husband's symptoms or questions or whatnot" (2HCP1).

MyChart did not impact workflow negatively, as was originally expected. The major benefits in terms of workflow was that it improved communication between patients and providers and that it provided a means of accessing information quickly. The work itself did not change, but there was a change in the way the work was performed. The system allowed other team members to be informed about communication and decisions.

Theme 3: Not All Was Great

Although there were many benefits of having MyChart available to patients, providers expressed several concerns. Not all patients were deemed fit to have access to their health care information, as "too much information sometimes is not a great thing" (1HCP1). Another provider stated that "a lot of patients wouldn't really want to know that; they'll just want to know if they're living or dying" (5HCP1 [NU]). Providers made decisions about which patients would be offered MyChart and stated that sometimes they did not offer it because there was no time to do so:

> *I offer it to all although I can tell you I'm not consistent. So, I mean to offer it all but there's plenty of times when I'm rushing around in a clinic, and/or if a person has a lot of other concerns it doesn't pop into your mind and so you're more focused on other things during your clinic visit.* [2HCP3]

Health care providers did not want the additional work of educating and helping patients or providing technical support. One provider stated, "more physicians are annoyed with that; they'll just stop offering that service" (1HCP3). This reluctance was based on previous experiences that technologies tend to be implemented without proper technical support or education. Providers felt that they had to *wing it* in using the system. MyChart was perceived as potentially taking time away from medical consultations if providers had to do promotion and education. As another provider stated, "I do not have time in that time to introduce a concept that has to be done by somebody else" (3HCP5). Some providers described that they just forgot about MyChart and did not bother with enrolling patients:

> *It's kind of forgotten sometimes and it could be the nature of the clinic. It's a busy clinic and maybe it's just forgotten as a tool. There is one physician that does use it more so than others.* [3HCP3 NU]

There were no clear guidelines and expectations, thus there was hesitancy about what was the right thing to do. Health care providers expressed apprehension because they were unclear what MyChart meant in terms of changes in charting practices.

Theme 4: Consider My Needs

When asked about the upcoming provincial-wide implementation, health care providers stated that if MyChart were continued, then it needed to reflect their needs and be established in a manner that would work within their busy schedules. It was stated that the system would need to be developed around some kind of incentives. No incentives would mean that "people won't use it" (1HCP3) and patients would not receive the opportunity to be introduced to the portal. Policies and guidelines should be implemented about charting, proxy access, and 2-way communication. MyChart could be incorporated as long as it did not contribute to additional workload (see Multimedia Appendix 4 for additional quotes).

Clinic Managers

A total of 4 clinic managers were interviewed, and 4 themes emerged from the interviews: (1) time for some efficiency, (2) heard great things about it, (3) heard about some issues as well, and (4) culture change is needed.

Theme 1: Time for Some Efficiency in Communication

When managers were asked why they decided to be part of the pilot, they stated that it was to eliminate paper and phone calls, try an innovation, and provide new ways for patients to manage complex chronic conditions. They had positive views about portals and how portals could impact practice in a positive way. This was an important finding, as clinic managers were involved in decisions about whether to have their clinic involved in the pilot:

> *I think that you know this is going to be the future. The question is how we get there.* [4CM1]

They viewed MyChart as a way to connect team members and assist patients and providers. They thought that the system had the capacity to improve the communication processes currently broken in the health care system, as patients continue to complain that they are unable to reach their providers when needed.

XSL•FO
RenderX

Theme 2: Heard Great Things About It

Clinic managers stated that the impact of patients seeing their medical records was seen as a facilitator to better care. Patients already search for things on the web, and MyChart provided access to correct and legitimate information:

> *Our patients are different patients than what they were 20 or 30 years ago because they have Internet access already. They already research all.* [2CM21]

It is important to note that clinic managers agreed that not all patients would use MyChart and that the system needed to accommodate this possibility. The MyChart 2-way communication function capacity was seen as a means of establishing new ways of communication and reducing reliance on phone calls.

Theme 3: Heard About Some Issues As Well

When the pilot started, only physicians were allowed to sign up patients to MyChart. This led to low enrollment numbers because physicians did not have enough time to explain MyChart to each patient. Furthermore, clinic managers described that many physicians did not participate in the pilot even though the clinic was part of the pilot. This was because of a lack of buy-in as physicians did not see MyChart as something beneficial. One clinic manager identified the issue of control as a drawback to the use of MyChart:

> *Physicians traditionally are so accustomed to kind of controlling everything that happens with that patient. And when the patient starts seeing the results of what they're trying to achieve. They're worried about it derailing the treatment progress and the relationship between the provider and the patient, when in fact I think the actual opposite occurs.* [2CM21]

Clinic managers acknowledged that they heard from the health care providers about patients having issues with MyChart. The sign-up process was challenging and created grievances. Technical support was lacking, and health care providers had to figure out ways to help patients with the sign-up process. One of the barriers was that physicians did not take full responsibility for the 2-way communication.

Theme 4: Change Is Needed

Clinic managers stated that change is needed in the current provider-patient relationship. They agreed that all clinics should be able to provide it, so all patients would have the same opportunity to access their health care information:

> *It needs to be a normal part of the life or not, it needs to be a normal part of, more the norm than the exception.* [1CM1]

Furthermore, clinic managers stated that physicians need to learn how to work with innovations such as MyChart and that these changes might not be easy in the beginning but are needed (see Multimedia Appendix 4 for additional quotes).

Nonmedical Staff

Only 1 clinic within the pilot opened the web-based appointment booking and canceling appointments function. For the scheduling function to be accessible, participating health care providers had to enter their availability in the system, which would allow patients to select from the available slots. Once a patient selects a slot, the clinic receives a message. If the booking was done incorrectly or the slot was changed, the office staff would call the patient and modify the booking. From these interviews, we identified 3 themes: (1) great service for the patients, (2) needs some education, and (3) need to maintain and expand.

Theme 1: Great Service for the Patient

The nonmedical staff described that participants "loved booking their own appointments especially because they know (if) something happens in the middle of the night, they can go on and book for the next day if there's openings" (4NON2). If there were any issues with the appointment, then the staff would call the patient and get it all sorted out; therefore, MyChart lightened the workload.

Theme 2: Needs Some Education

As not all health care providers participated in the pilot, some patients in a given clinic were able to make web-based appointments, whereas others could not do so. In addition, a challenge was with patients who would book many appointments, as "they can book a whole bunch of them on MyChart in one day" (04MANON00213DEC2018).

Theme 3: Need to Maintain and Expand

Staff indicated that they would like more information about MyChart so that they could assist patients better and said that they needed "more awareness and if they ask questions I can help as I can't help them much now" (4NON3; see Multimedia Appendix 4 for additional quotes).

Summary of the Results

We report the uptake and adoption of an initial phase of a patient portal introduction among 5 health care clinics that were, at the outset of the introduction, judged to be clinics where high uptake would be found. In contrast, 2 clinics showed high uptake, 1 clinic showed moderate uptake, 1 clinic showed low uptake, and the last clinic was categorized as a clinic without use. We identified several factors related to the uptake of the patient portal, including satisfaction and/or dissatisfaction with the current communication tools, performance expectancy, facilitating conditions, behavioral intentions, and use behavior. All these factors had an impact on the level of uptake in each setting and created different opportunities for the end users. Table 5 summarizes our findings about each clinic by considering our findings.

XSL•FO
RenderX

Table 5. Impact of the MyChart pilot on each case study.

Case study	Level of usage	Satisfaction or dissatisfaction with the current communication tools	Performance expectancy	Facilitating conditions	Behavioral intention	Use behavior
Clinic 4	High user; 965 enrolled patients	Clinic considered as an innovation hub clinic; first clinic to go live within the PoC[a]; 2 family doctors and a registered nurse participated; 3 family doctors did not participate	MyChart described as the *most rewarding* initiative; offered the web-based booking and canceling appointment function. Only clinic to do so	The registered nurse acted as the MyChart support person	High enrollment as they offered it to almost all patients	The clinic maintained high enrollment and incorporated the frontline staff (office assistants) to be part of the process
Clinic 2	High user; 875 enrolled patients	It was described that the clinic had the *old way* and the *new way* processes	The participating providers used the technology to share medical results with their patients and decrease the number of phone calls to the clinic	Two specialists maintained the *old way* who did not participate in the PoC because of the assumption that it will increase their workload and patients will contact them all the time. Two specialists implemented the *new way*	This process was maintained for a while, until the nurses from the nonparticipating specialists decide to start enrolling patients on MyChart	After the nurses took on the initiative to sign up patients to MyChart, the enrollment increased
Clinic 1	Medium user; 357 enrolled patients	MyChart was brought in because of the potential for uploading forms and questionnaires that patients tend to do	The technology did not produce the identified needs as the forms on MyChart were found to be inappropriate by the clinic	Nurses took on the active role in enrolling patients	Only offered to certain patients (medium enrollment)	The implementation process was observed by 2 other specialty clinics that enrolled in the PoC and achieved a high enrollment (1047 and 667 patients)
Clinic 3	Low user; 172 enrolled patients	The participating specialist is the clinic manager that brought MyChart to the clinic with the intent to decrease the number of phone calls or unnecessary follow-up appointments	The portal was deemed needed because of the 2-way communication function. However, patient messages sent through MyChart were often unanswered because of staff changes	The nurses did not actively participate in the enrollment process; other staff did not see the benefit and were unwilling to participate because of the 2-way communications function	Low enrollment, although patients expressed interest (interviewed 1 patient from this clinic who did not have access but wanted it)	MyChart was described as an *afterthought*
Clinic 5	Nonuser; 0 enrolled patients	Although offered to participate, the staff at this academic setting family clinic decided not to	The technology was deemed difficult to incorporate in the clinic's flow	The clinic has patients with MyChart because of access provided at other clinics, but no provider from this clinic wanted to use or assist patients with the portal	The portal was not seen as something necessary in the delivery of health care services	No perceived value and thus no enrolment or usage

[a]PoC: proof of concept.

Discussion

Principal Findings

In our study, dissatisfaction with current health care communication tools appeared to be the primary factor that drove patients' and providers' decisions to consider the use of MyChart. Once they had made this decision, uptake was dependent on performance expectancy, which included understanding how to use the technology and valuing the additional functionality that it offered. Furthermore, although the themes emerged from the interview data, they aligned with the *theory of technology acceptance* by Davis [14] and included *perceived usefulness*, *perceived ease of use*, *barriers to ease of use*, and *future usage* [14]. For example, theme 1 from the patient interviews was *my health, my responsibility, but I need the information to do that*, which correlated with *perceived usefulness* [14], as shown in Textbox 2.

XSL•FO
RenderX

Textbox 2. Themes per sample group correlating with the theory of technology acceptance.

Patients

- Theme 1: My health, my responsibility, but I need the information to do that (perceived usefulness)
- Theme 2: Convenience, convenience, convenience (perceived ease of use)
- Theme 3: A few tweaks will do it (barriers to ease of use)
- Theme 4: Do not take it away (future usage)

Health care providers

- Theme 1: Of course we need it (perceived usefulness)
- Theme 2: Yes, it did some good things (perceived ease of use)
- Theme 3: Not all was great (barriers to ease of use)
- Theme 4: Consider my needs (future usage)

Clinic managers

- Theme 1: Time for some efficiency in communication (perceived usefulness)
- Theme 2: Heard great things about it (perceived ease of use)
- Theme 3: Heard about some issues as well (barriers to ease of use)
- Theme 4: Change is needed (future usage)

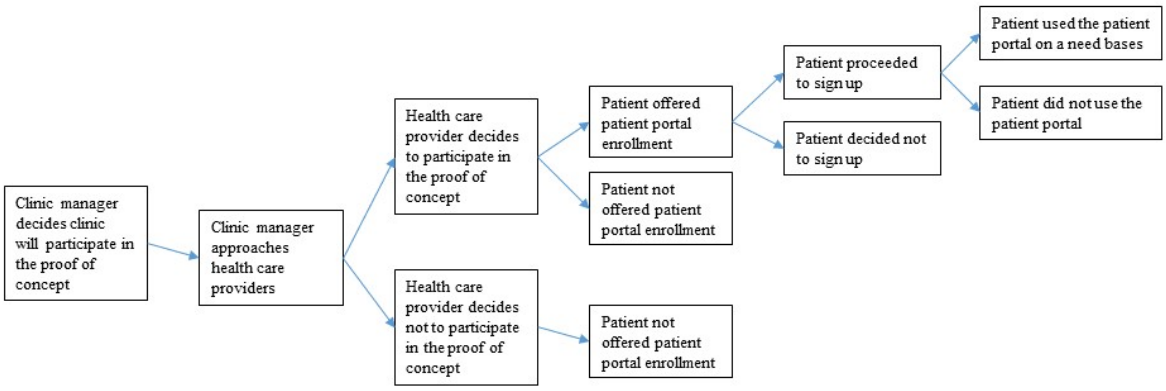
Nonmedical providers

- Theme 1: Great service for the patient (perceived usefulness)
- Theme 2: Needs some education (barriers to ease of use)
- Theme 3: Need to maintain and expand (future usage)

The conditions that facilitated understanding of the technology varied across clinics. Some clinics formally designated assistance with MyChart use to staff in their clinic. All clinics had high or medium uptake of MyChart. In other clinics, staff decided not to offer MyChart to some patients, which had an impact on the uptake of MyChart. Reasons are not entirely clear and require further investigation. Some patients who understood how to use the technology and valued it became *users*, whereas others did not. The main reason for discontinuing use appeared to be the lack of an immediate health care concern [25].

Clinic managers were foundational to the uptake of MyChart (Figure 2), a finding that has not been discussed in the literature to date. They made decisions about whether to present the technology to providers. Our data suggest that this decision was based on an informal assessment of satisfaction with the current communication tools. Clinic managers who declined participation in the pilot indicated that their clinic patients and providers would not be interested in trying the technology. Figure 2 summarizes the impact of the clinic managers on the uptake of MyChart.

Figure 2. Clinic managers as gatekeepers to uptake of MyChart.



The actions of the clinic managers could be observed through Roger's theory of diffusion, as they served as early adopters by assessing the advantages or disadvantages of the innovation. They determined the *perceived efficiencies*, its *compatibility* with the existing system, the *complexity*, and, ultimately, the *benefits* and unintended *consequences* of the innovation [13].

Usage and enrollment depended on how patients were informed about the portal. Patients in this study reported variable

XSL•FO
RenderX

understanding of the portal following the introduction by clinic staff. Similar to the study by Kim et al [25], a quick introduction did not appear to be valuable to patients. In this study, participating health care providers reported not having sufficient time to explain the portal to patients. In clinics with higher uptake, a short introduction became incorporated into the clinic visit workflow, which allowed for more sufficient means in enticing patients to sign up.

Knowledge about the technology was another important component that led to increased or decreased usage. Ryan et al [26] found that patients and health care providers had positive views about patient portals but that older patients had less knowledge about patient portals. In this study, older participants reported that they took the time to learn and understand the system once they decided to enroll. We found that patients had to learn the system by *trial and error*. When patients were unable to navigate the sign up and authentication, the use of MyChart became impossible. When patients accessed medical information in MyChart, they reported becoming more active in the management of their health, which allowed them to see their health trends and identify opportunities for self-care. Patients came prepared with specific health questions because they could spend less time going over less valuable health information (eg, lab results) and making better use of scarce time with their care provider. This finding was similar to the results reported by Dendere et al [27].

Some research teams have reported a relationship between concerns regarding privacy and technology uptake [28-36]. In our study, participants did not report perceiving significant risks to having the confidentiality of their health information violated. Participants reported a sense that the health system would maintain appropriate controls and procedures to protect the security of their health information.

Although not all health care providers saw the *value* of patients having access to the same information, participants *championing* use of MyChart experienced more efficient ways to connect and communicate with their patients. This finding is in contrast to the findings reported by others who have studied patient portals [27,37-40] and reported that portals increased workload. For example, Koivunen et al [41] described the negative views of nurses about patient portals. Our findings are consistent with the work of other research teams who reported that nurses saw patient portals as an opportunity to improve communication with patients [42-44].

Furthermore, theories of how patient portals support health care system transformation may not draw upon crucial empirical data. The current findings emphasize that without a shared understanding of the purpose of patient portals and some dissatisfaction with existing communication methods, uptake lags. We posit that these aspects will be crucial as patient portals introduced in real-life care settings often accompany other technologies for clinicians and managers to navigate [44].

We agree with Wiljer et al [45], who described a need for a culture shift, where access to medical records is a fundamental right of every patient. A shift of this magnitude could only be accomplished with broad support from policy makers, health care providers, health administrators, and patients. Without adequate support for patients as complete partners in the management of their health, patients would be left to take actions based on incomplete and possibly inaccurate information from untrustworthy sources. Clinically useful uptake of patient portals is an essential enabler for patients as partners in managing their health and transforming the wider health system.

Limitations

There were several limitations to this study. Patient interviews were only conducted with patients who volunteered to participate and were attending an appointment with a health care provider who was also participating in the study. The information provided by participants may not have been recalled accurately by them. In addition, although all efforts were made, we were unable to identify and recruit participants who did not continue to use MyChart as we had no means to obtain information about discontinued usage. Finally, the findings should be generalized with caution to other health care systems and to settings that use other types of patient portals.

Conclusions

This work exposes a number of factors that have an impact on the uptake of MyChart. We found that dissatisfaction with nonportal-based communications was a foundational element that likely needs to be addressed before more advanced goals can be reached. We also found that clinical managers were key gatekeepers in the uptake process. Once a clinic manager decided that the clinic should use the portal, the designation of a staff member to help patients become more familiar with the portal appeared to increase uptake.

Acknowledgments

The authors would like to acknowledge and thank the clinic managers and clinic staff for their interest and participation in this study. Without their help, this research would not have been possible. The authors would also like to thank all patients who participated in this study and guided the research team toward a better understanding of their experiences when using the patient portal. Furthermore, the authors would like to thank the Canadian Institutes of Health Research's (CIHR's) Health System Impact Fellowship and the Alberta Innovates Graduate Student Scholarship programs for providing funding for this work. Financial support for this work was provided to MA through the CIHR's Health System Impact Fellowship and the Alberta Innovates Graduate Student Scholarship.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment card.
[DOCX File , 243 KB - jmir_v22i7e18973_app1.docx]

Multimedia Appendix 2
Comparative case interview guide questions.
[DOCX File , 34 KB - jmir_v22i7e18973_app2.docx]

Multimedia Appendix 3
Interview codes for all interviewed participants in each case setting.
[DOCX File , 34 KB - jmir_v22i7e18973_app3.docx]

Multimedia Appendix 4
Additional quotes supporting thematic analysis.
[DOCX File , 28 KB - jmir_v22i7e18973_app4.docx]

References

1. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the 'triple aim' of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
2. Engaging the Patient in Healthcare: An Overview of Personal Health Record Systems and Implications for Alberta. Alberta Health Services (AHS). 2009. URL: https://www.albertahealthservices.ca/org/ahs-org-ehr.pdf [accessed 2019-02-10]
3. Information Management and Information Technology Strategic Plan. Alberta Health Services (AHS). 2015. URL: http://www.albertahealthservices.ca/assets/about/msd/ahs-msd-ahs-imit-strategy.pdf [accessed 2019-02-10]
4. The Patient First Strategy. Alberta Health Services (AHS). 2015. URL: http://www.albertahealthservices.ca/assets/info/pf/first/if-pf-1-pf-strategy.pdf [accessed 2019-03-15]
5. Better Healthcare for Albertans. Office of the Auditor General of Alberta. 2017. URL: https://www.oag.ab.ca/webfiles/reports/Better_Healthcare_Report.pdf [accessed 2019-03-15]
6. Opportunities for Action: A Pan-Canadian Digital Health Strategic Plan. Colleaga: Harnessing the Power of Collective Intelligence. 2013. URL: https://www.colleaga.org/sites/default/files/attachments/chi_strategic_plan_english_web_new.pdf [accessed 2019-04-10]
7. Electronic Health Records in Canada: An Overview of Federal and Provincial Audit Reports. Provincial Auditor of Saskatchewan. 2010. URL: https://auditor.sk.ca/pub/publications/special/2010EHealthRecordsinCanada.pdf [accessed 2019-04-10]
8. Basky G. Some provinces still delay access to health records via patient portals. Can Med Assoc J 2019 Dec 2;191(48):E1341 [FREE Full text] [doi: 10.1503/cmaj.1095829] [Medline: 31791973]
9. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
10. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. Br Med J 2008 Jun 7;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
11. ImagineNation Challenges. Canada Health Infoway. 2016. URL: http://imaginenationchallenge.ca/wp-content/uploads/2017/02/eCLINICIAN-MyChart-Working-Group.pdf [accessed 2019-03-12]
12. Graham T, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
13. Rogers EM. Diffusion of Innovations. New York, USA: Free Press; 2003.
14. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
15. Yin RK. Enhancing the quality of case studies in health services research. Health Serv Res 1999 Dec;34(5 Pt 2):1209-1224 [FREE Full text] [Medline: 10591280]
16. Yin R. Case Study Research: Design and Methods. Fifth Edition. Thousand Oaks, CA: Sage Publications; 1994.
17. Stake R. Case studies. In: Denzin NK, Lincoln YS, editors. Handbook of Qualitative Research. Thousand Oaks, CA: Sage Publications; 1994:236-247.

XSL•FO
RenderX

18. About AHS: Who are We. Alberta Health Services (AHS). URL: https://www.albertahealthservices.ca/about/about.aspx [accessed 2019-02-05]
19. Patton M. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
20. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
21. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92 [FREE Full text] [doi: 10.1177/160940690600500107]
22. Nowell L, Norris J, White D, Moules N. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 2;16(1):160940691773384 [FREE Full text] [doi: 10.1177/1609406917733847]
23. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Dir Program Eval 1986;1986(30):73-84 [FREE Full text] [doi: 10.1002/ev.1427]
24. Morse JM. Analytic strategies and sample size. Qual Health Res 2015 Oct;25(10):1317-1318. [doi: 10.1177/1049732315602867] [Medline: 26355022]
25. Kim J, Mathews H, Cortright LM, Zeng X, Newton E. Factors affecting patient portal use among low-income pregnant women: mixed-methods pilot study. JMIR Form Res 2018 Mar 22;2(1):e6 [FREE Full text] [doi: 10.2196/formative.5322] [Medline: 30684413]
26. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 4;23(2):848 [FREE Full text] [doi: 10.14236/jhi.v23i2.848] [Medline: 27869582]
27. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
28. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
29. Luque AE, van Keken A, Winters P, Keefer MC, Sanders M, Fiscella K. Barriers and facilitators of online patient portals to personal health records among persons living with HIV: formative research. JMIR Res Protoc 2013 Jan 22;2(1):e8 [FREE Full text] [doi: 10.2196/resprot.2302] [Medline: 23612564]
30. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
31. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new Europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]
32. Lockhart S, Wallace I, Nugent A, Black N, Quinn M, Johnston PC. A survey of patient's perceptions and proposed provision of a 'patient portal' in endocrine outpatients. Ulster Med J 2019 Sep;88(3):157-161 [FREE Full text] [Medline: 31619849]
33. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013 Aug 26;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
34. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
35. Vreugdenhil MM, Ranke S, de Man Y, Haan MM, Kool RB. Patient and health care provider experiences with a recently introduced patient portal in an academic hospital in the Netherlands: mixed methods study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
36. Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A. 'Nothing about me without me': an interpretative review of patient accessible electronic health records. J Med Internet Res 2015 Jun 29;17(6):e161 [FREE Full text] [doi: 10.2196/jmir.4446] [Medline: 26123476]
37. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. J Am Med Assoc 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]
38. Hobbs J, Wald J, Jagannath YS, Kittler A, Pizziferri L, Volk LA, et al. Opportunities to enhance patient and physician e-mail contact. Int J Med Inform 2003 Apr;70(1):1-9. [doi: 10.1016/s1386-5056(03)00007-8] [Medline: 12706177]
39. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
40. Kelly MM, Dean SM, Carayon P, Wetterneck TB, Hoonakker PL. Healthcare team perceptions of a portal for parents of hospitalized children before and after implementation. Appl Clin Inform 2017 Mar 15;8(1):265-278 [FREE Full text] [doi: 10.4338/ACI-2016-11-RA-0194] [Medline: 28293685]

XSL•FO
RenderX

41. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive internet-portal application for patient education in psychiatric hospitals. Patient Educ Couns 2008 Mar;70(3):412-419. [doi: 10.1016/j.pec.2007.11.002] [Medline: 18079085]
42. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
43. Hazara AM, Bhandari S. Barriers to patient participation in a self-management and education website renal patientView: a questionnaire-based study of inactive users. Int J Med Inform 2016 Mar;87:10-14. [doi: 10.1016/j.ijmedinf.2015.12.004] [Medline: 26806707]
44. Koppel R, Lehmann CU. Implications of an emerging EHR monoculture for hospitals and healthcare systems. J Am Med Inform Assoc 2015 Mar;22(2):465-471. [doi: 10.1136/amiajnl-2014-003023] [Medline: 25342181]
45. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]

Abbreviations

AHS: Alberta Health Services
CIHR: Canadian Institutes of Health Research

Edited by G Eysenbach; submitted 30.03.20; peer-reviewed by N Shaw, W van Harten; comments to author 20.04.20; revised version received 12.05.20; accepted 04.06.20; published 27.07.20.

Please cite as:
Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T
Patient Portal Implementation and Uptake: Qualitative Comparative Case Study
J Med Internet Res 2020;22(7):e18973
URL: http://www.jmir.org/2020/7/e18973/
doi:10.2196/18973
PMID:32716308

©Melita Avdagovska, Mark Ballermann, Karin Olson, Timothy Graham, Devidas Menon, Tania Stafinski. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Assessment of Supportive Accountability in Adults Seeking Obesity Treatment: Psychometric Validation Study

Karishma Chhabria[1], MPH, PhD; Kathryn M Ross[2], MPH, PhD; Shane J Sacco[3], MA; Tricia M Leahey[3], MA, PhD

[1]The Center for Health Care Data, Department of Management, Policy and Community Health, University of Texas Health Science Center at Houston, Houston, TX, United States

[2]Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, United States

[3]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

Corresponding Author:
Karishma Chhabria, MPH, PhD
The Center for Health Care Data
Department of Management, Policy and Community Health
University of Texas Health Science Center at Houston
1200 Pressler St
Houston, TX, 77030
United States
Phone: 1 713 500 9428
Email: karishma.s.chhabria@uth.tmc.edu

Abstract

Background: Technology-mediated obesity treatments are commonly affected by poor long-term adherence. Supportive Accountability Theory suggests that the provision of social support and oversight toward goals may help to maintain adherence in technology-mediated treatments. However, no tool exists to measure the construct of supportive accountability.

Objective: This study aimed to develop and psychometrically validate a supportive accountability measure (SAM) by examining its performance in technology-mediated obesity treatment.

Methods: Secondary data analyses were conducted in 2 obesity treatment studies to validate the SAM (20 items). Study 1 examined reliability, criterion validity, and construct validity using an exploratory factor analysis in individuals seeking obesity treatment. Study 2 examined the construct validity of SAM in technology-mediated interventions involving different self-monitoring tools and varying amounts of phone-based interventionist support. Participants received traditional self-monitoring tools (standard, in-home self-monitoring scale [SC group]), newer, technology-based self-monitoring tools (TECH group), or these newer technology tools plus additional phone-based support (TECH+PHONE group). Given that the TECH+PHONE group involves more interventionist support, we hypothesized that this group would have greater supportive accountability than the other 2 arms.

Results: In Study 1 (n=353), the SAM showed strong reliability (Cronbach α=.92). A factor analysis revealed a 3-factor solution (representing Support for Healthy Eating Habits, Support for Exercise Habits, and Perceptions of Accountability) that explained 69% of the variance. Convergent validity was established using items from the motivation for weight loss scale, specifically the social regulation subscale (r=0.33; *P*<.001) and social pressure for weight loss subscale (r=0.23; *P*<.001). In Study 2 (n=80), the TECH+PHONE group reported significantly higher SAM scores at 6 months compared with the SC and TECH groups (r^2=0.45; *P*<.001). Higher SAM scores were associated with higher adherence to weight management behaviors, including higher scores on subscales representing healthy dietary choices, the use of self-monitoring strategies, and positive psychological coping with weight management challenges. The association between total SAM scores and percent weight change was in the expected direction but not statistically significant (r=−0.26; *P*=.06).

Conclusions: The SAM has strong reliability and validity across the 2 studies. Future studies may consider using the SAM in technology-mediated weight loss treatment to better understand whether support and accountability are adequately represented and how supportive accountability impacts treatment adherence and outcomes.

Trial Registration: ClinicalTrials.gov NCT01999244; https://clinicaltrials.gov/ct2/show/NCT01999244

(J Med Internet Res 2020;22(7):e17967) doi:10.2196/17967

XSL•FO
RenderX

KEYWORDS

supportive accountability; social support; SALLIS; factor analysis; obesity; weight loss; technology; mobile phone

Introduction

Background

Technology-mediated health behavior change interventions have become ubiquitous [1-4]. Adherence tends to drop off dramatically over time, such that few participants meet the program goals related to the use of technology-based intervention tools (eg, self-monitoring platforms) by the end of treatment [5-7]. Importantly, this poor adherence may lead to suboptimal treatment outcomes. Although technology-based interventions can lead to clinically significant weight loss in some participants [8], most of these programs demonstrate subpar outcomes in relation to in-person treatments [9,10]. For example, electronically delivered behavioral weight loss programs commonly achieve a weight loss of 3 kg to 5 kg within 3 to 6 months of treatment [11,12], whereas traditional in-person programs often produce a weight loss of 8 kg to 10 kg within the same time frame [11,13].

Given this challenge, research has focused on methods to improve adherence to technology-based behavioral interventions. To date, one of the most effective strategies has been the provision of additional support (eg, via telephone, email, text messages or via smartphone apps) from interventionists or lifestyle coaches [14-16]. Indeed, existing guidelines for the treatment of overweight and obesity specifically note that programs delivered electronically should provide personalized feedback from an interventionist and that primarily knowledge-based internet programs may be ineffective [11].

Consistent with these empirical findings, Mohr et al [17] recently proposed Supportive Accountability Theory, which posits that adherence to technology-based interventions may be improved through the provision of both social support (whether delivered in person or electronically and either synchronously or asynchronously) and accountability (the expectation that an individual would regularly have to explain their progress toward program goals). Despite the utility of Supportive Accountability Theory for identifying important potential mechanisms for improving adherence and, ultimately, program outcomes in technology-based interventions, there is no validated measure to assess supportive accountability.

Objectives

The objective of this study was to describe the development and validation of a measure for assessing supportive accountability within the context of technology-mediated programs for the treatment of adult obesity. Building off of the Social Support for Diet and Exercise Behaviors Scale by Sallis et al [18], a widely used measure of social support for healthy eating and activity, new items were added to represent the construct of accountability as described by Mohr et al [17]. The psychometric characteristics of this new measure (Supportive Accountability Measure [SAM]) were assessed using data collected from 2 weight loss trials. Baseline data from the first trial (Study 1) aimed to examine reliability, conduct an exploratory factor analysis (EFA), and assess convergent validity. Data from the second trial (Study 2) aimed to further investigate the criterion validity of the new measure. Specifically, Study 2 examined the impact of newer self-monitoring technology and phone coaching on weight loss; thus, it was hypothesized that participants who were randomized to receive additional phone coaching, in addition to technology-based self-monitoring tools, would report higher supportive accountability compared with participants not provided with this additional coaching.

Methods

Supportive Accountability Measure Development

The SAM was developed by adding novel items representing the construct of accountability to items from the existing Social Support for Diet and Exercise Behaviors Scale developed by Sallis et al [18], a widely used social support measure in the weight management literature [15]. First, questions assessing accountability were developed by a senior author (TL), drawing on Mohr's description of Supportive Accountability Theory [17] and her expertise in the field of obesity treatment. Items were developed to assess accountability for typical behavioral intervention components, including goals related to weight loss, dietary intake, and physical activity. Participants respond to items on this scale using a 7-point Likert scale, with response options ranging from 1 (*not at all*) to 7 (*very much*). A total of 10 items were judged to fully capture all facets of this construct (supporting content validity) and were thus retained. Multimedia Appendix 1 shows the final accountability items selected for the SAM.

To ensure balance in the number of items, 10 items were selected from the Sallis Social Support measure [17]. The original Sallis measure includes 43 items; some assess social support, whereas others assess sabotage, negative comments, and punishment. Overall, these subscales have demonstrated good internal consistency, criterion validity, and construct validity and acceptable test-retest reliability [18]; however, less support has been found for the utility of the negative subscales (Negative Comments, Sabotage, and Punishment) [18,19]. Given the lack of psychometric support for the negative subscales and that the questions focused on negative comments and sabotage did not align with the theoretical construct of supportive accountability, the 10 items that were retained from the Sallis measure were those that assessed positive support for healthy eating and exercise [18]. Participants were asked to rate support received from family, friends, and their weight loss coach on the 10 items using a 5-point Likert scale identical to that used in the original Sallis measure (1=*Never*, 2=*Rarely*, 3=*A few times*, 4=*Often*, 5=*Very often*, and 6=*Does not apply*, with scores of *6* being recoded as *1* before scoring).

A total score for the SAM was created using the 10 support items and 10 accountability items. The Social Support subscale involves summing all social support items. Since the

XSL•FO
RenderX

accountability items were measured on a 7-point scale and the social support items on a 5-point scale, the scores for the accountability measure involved summing all accountability items, multiplying the sum by 5, and then dividing by 7 to yield identical maximum scores for the 2 subscales. The total scores for social support and accountability were then summed to create an overall SAM score, with higher scores indicating higher levels of supportive accountability (possible range of total SAM scores 17.14-100). We then examined the psychometric characteristics of the SAM across 2 weight management trials (the study details are provided below under the Methods section for Study 1 and Study 2). Study procedures for Study 1 and Study 2 were approved by The Miriam Hospital Institutional Review Board, and secondary analyses of Study 2 data were approved by the University of Florida Institutional Review Board.

Study 1

Study 1 evaluated the psychometric characteristics of the SAM in adults with obesity enrolled in a behavioral weight management trial. Only baseline data from this trial were used, allowing for the evaluation of the SAM in the relevant population before any intervention or support.

Participants

Study 1 included 353 adults (aged 40-60 years) with obesity (BMI between 30 kg/m^2 and 40 kg/m^2). Full eligibility criteria and recruitment procedures have been described elsewhere [20]. In brief, participants were recruited through mass mailings, newspaper/listserv ads, and direct referrals to an obesity treatment research center and were screened over the phone. Exclusion criteria were participation in another weight loss program, use of weight loss medication, history of bariatric surgery, weight loss of ≥5% in the last 6 months, pregnancy, lactation, less than 6 months postpartum, or plans to become pregnant during the study period, any health conditions that would contraindicate weight loss (eg, cancer, reported uncontrolled heart condition, inability to walk 2 blocks without stopping, and unexplained loss of consciousness), or inability to participate in an in-person, group program (eg, scheduling constraints). For individuals reporting joint problems, medication use, or other medical conditions that may limit the ability to exercise or need adjustment with weight loss, physicians' consent was obtained. Study 1 participants were excluded from the current analyses if they did not complete the SAM measure at baseline (n=3).

Measures

All measures were collected at baseline.

Demographics

Standard demographics (eg, sex, age, race/ethnicity) were collected via a survey.

Anthropometrics

Height and weight were objectively assessed, with participants wearing one layer of light indoor clothing and shoes removed.

Supportive Accountability

The SAM was used to assess supportive accountability (see the Supportive Accountability Measure Development section for details).

Social Motivation for Weight Loss

The Treatment Self-Regulation Questionnaire (TSRQ) was administered to assess its convergent validity with the SAM. Given that the SAM is expected to measure social accountability, we examined whether the social external regulation subscale of the TSRQ [21] was associated with the SAM; 6 items were included, ranging from 1=*Not true at all* to 7=*Very true* (eg, "I want to lose weight or control my weight because…my spouse, family, friends or doctor would be upset if I didn't"). Mean scores were calculated for this scale.

Social Pressure for Weight Loss

Social pressure was assessed using an item from the Motivating Factors for Weight Loss Scale, developed to assess motivation for weight loss among participants in the National Weight Control Registry [22]. Participants were asked how important social pressure was in their decision to lose weight or join a weight loss program. Response options ranged from 1=*Not at all important* to 5=*Extremely important*.

Statistical Analyses

Analyses were conducted using SPSS Statistics for Windows, version 25 (IBM Corp). The reliability of SAM items was assessed via the Cronbach alpha [23], and construct validity was assessed using an EFA using principal component analysis with Oblimin rotation (delta=0). A scree plot was used to assess the number of factors using the standard eigenvalue of *1* or greater. Convergent validity of the SAM was assessed by correlating total SAM scores with the total Social Motivation score and the Social Pressure item.

Study 2

Study 2 was a randomized weight management trial investigating the impact of newer self-monitoring technology (ie, a Bluetooth-enabled activity monitor, a smart scale, and a website/smartphone app that synced with both of these devices and allowed individuals to self-monitor caloric intake) and phone coaching on weight loss. It was hypothesized that participants who received interventionist support through phone coaching would report significantly higher supportive accountability, as assessed by the SAM, at a 6-month posttest because of the presence of additional support. Moreover, we hypothesized that higher supportive accountability at intervention posttest would be associated with greater intervention adherence. As an exploratory aim, we investigated whether higher supportive accountability at the posttest was associated with greater weight loss from baseline to posttest.

Participants

Study 2 participants were 80 adults (aged between 18 and 70 years) with overweight or obesity (BMI between 27 kg/m^2 and 40 kg/m^2) who reported having access to a computer and Wi-Fi at home [24]. Exclusion criteria were similar to Study 1; full eligibility criteria and details regarding participant recruitment

XSL•FO
RenderX

and screening have been published previously [24]. Study 2 participants were excluded from the current analyses if they did not complete the SAM measure at the 6-month follow-up.

Intervention

Study 2 was a randomized trial that examined the impact of a 6-month weight loss intervention in which participants were randomized to 1 of 3 treatment groups, using traditional self-monitoring tools (a paper food record, a printed calorie reference book, a standard pedometer, and a standard in-home scale—SC group), newer, technology-based self-monitoring tools (Fitbit Zip, Fitbit Aria smart scale, and use of the Fitbit app/website to track dietary intake—TECH group), or these newer technology tools plus phone-based interventionist support (TECH+PHONE group). All participants received a one-time, group-based Weight Loss 101 session that provided information on how to accurately monitor calories, weight, and physical activity and weight management goals for calories, exercise, and weight loss. Participants were also taught how to use their assigned self-monitoring tools. Participants randomized to SC (n=26) and TECH (n=27) received self-monitoring tools only; they did not receive any interventionist support. Participants randomized to TECH+PHONE (n=27) received the additional phone-based intervention (8 weekly, 4 biweekly, and 2 monthly contacts; each lasted 10-15 min), delivered by trained interventionists (either a clinical psychologist or dietitian, both experienced in delivering behavioral weight management programs), using a manualized protocol.

Measures

Demographics

Standard demographics (eg, sex, age, race/ethnicity) were collected via a survey at baseline.

Anthropometrics

Height and weight were measured with participants wearing light indoor clothing and with shoes removed. Height was measured at baseline and weight was measured at baseline and at the 6-month posttest. Weight change was operationalized as percent weight loss from baseline to the posttest visit.

Supportive Accountability

Supportive accountability was assessed at the 6-month posttest using the SAM.

Use of Weight Management Strategies

Weight management strategies were assessed at the 6-month posttest using the Weight Control Strategies Scale (WCSS) [25]. This self-report measure assesses behaviors across 4 subscales: dietary choices, self-monitoring strategies, physical activity, and psychological coping. WCSS dietary choices and physical activity subscales have been shown to be associated with changes in caloric intake and physical activity during weight management interventions, and WCSS scores have been shown to correlate with weight loss during these programs.

Statistical Analyses

Analyses were conducted using SPSS. Reliability was reassessed in this sample using Cronbach alpha. Construct validity of the SAM was assessed using a one-way analysis of variance, investigating differences in SAM scores by treatment group. We hypothesized that, at the 6-month posttest, the SAM score would be significantly higher in the TECH+PHONE condition than in the SC and TECH conditions, given that this condition was provided with additional interventionist support.

Results

Study 1

Study 1 included a total of 350 participants (Table 1). The mean SAM score was 48.63 (SD 16.33); mean support and accountability subscale scores were 22.71 (SD 7.80) and 36.28 (SD 17.28), respectively. There was an association between total SAM scores and baseline BMI (r=0.13; P=.02). In examining the subscales, accountability scores were positively associated with baseline BMI (r=0.17; P=.002), whereas support scores were not (P=.96). There were no associations between SAM scores and participant demographic characteristics (P>.06).

The overall reliability of the SAM was excellent, as demonstrated by the internal consistency (Cronbach α=.92). The lowest item to total SAM correlation (r=0.41) was for the item, “My friends/family encouraged me to not eat unhealthy foods when I’m tempted to do so.” However, the total Cronbach alpha was not improved with this item removed; thus, the item was retained. Cronbach alpha for the support and accountability subscales were also strong (.88 and .95, respectively). Inter-item correlations for the SAM are provided in Multimedia Appendix 2. As expected, items reflecting social support (items 1-10) correlated more closely with each other and had lower correlations with accountability items (items 11-20).

Results from the EFA demonstrated that a three-factor solution provided the best fit (Figure 1), explaining 69.2% of the variance. Item loadings for the three-factor solution, representing Support for Healthy Eating Habits, Support for Exercise Habits, and Perceptions of Accountability, are presented in Table 2. Assumptions of EFA were verified using the Kaiser-Meyer-Olkin (KMO) statistic, which demonstrated sampling adequacy (KMO=0.88). Furthermore, the Bartlett test of sphericity demonstrated that items correlated satisfactorily (χ^2_{190}=6573.91; P<.001).

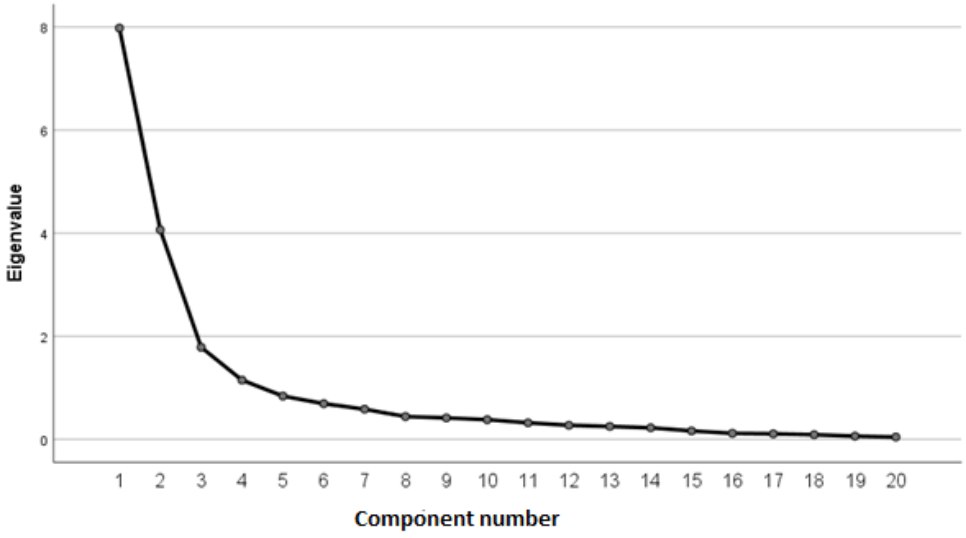
Convergent and divergent validity analyses revealed statistically significant correlations between the SAM total score and TSRQ items representing external motivation for weight loss (r=0.34; P<.001) and the social pressure item (r=0.23; P<.001). In examining the subscales of SAM, there were significant correlations between the Support subscale and Social Motivation items (r=0.19; P=.001) and the Social Pressure item (r=0.17; P=.002). Similarly, there were significant correlations between the Perceptions of Accountability subscale and Social Motivation items (r=0.33; P<.001) and the Social Pressure item (r=0.20; P<.001).

XSL•FO
RenderX

Table 1. Baseline and demographic characteristics of participants in study 1

Characteristic	Study 1 (n=350)
Age (years), mean (SD)	51.7 (5.6)
BMI (kg/m^2), mean (SD)	34.8 (3.3)
Gender, n (%)	
Male	80 (22.9)
Female	270 (77.1)
Education, n (%)	
High school or less	38 (10.9)
Vocational training	24 (6.9)
Some college	83 (23.7)
College degree	108 (30.9)
Graduate degree	97 (27.7)
Race, n (%)	
American Indian or Alaska Native	6 (1.7)
Asian	1 (0.3)
Black or African American	36 (10.3)
White	262 (74.9)
Other	36 (10.3)
Ethnicity, n (%)	
Hispanic or Latino	45 (12.9)
Not Hispanic or Latino	304 (86.9)

Figure 1. Scree plot for exploratory factor analysis of the supportive accountability measure.



XSL•FO
RenderX

Table 2. Three-factor structure for Supportive Accountability Measure items.

Items	Support for Healthy Eating	Support for Exercise Habits	Perceptions of Accountability
My weight coach, friends, and/or family			
1. Encouraged me not to eat *unhealthy foods* (cake, potato chips) when I'm tempted to do so	0.81	N/A[a]	N/A
2. Discussed my eating habit changes with me (asked me how I'm doing with my eating changes)	0.83	N/A	N/A
3. Reminded me not to eat high fat, high calorie foods	0.85	N/A	N/A
4. Complimented me on changing my eating habits ("Keep it up. We are proud of you.")	0.68	N/A	N/A
5. Commented if I went back to my old eating habits	0.71	N/A	N/A
6. Exercised with me	N/A	0.88	N/A
7. Offered to exercise with me	N/A	0.87	N/A
8. Gave me helpful reminders to exercise ("Are you going to exercise tonight?")	N/A	0.85	N/A
9. Gave me encouragement to stick with my exercise program	N/A	0.78	N/A
10. Changed their schedule so we could exercise together	N/A	0.81	N/A
11. I feel accountable to others (eg, friends, family, or doctor) for meeting my weight goals.	N/A	N/A	0.77
12. I feel accountable to others (eg, friends, family, or doctor) for meeting my dietary goals.	N/A	N/A	0.79
13. I feel accountable to others (eg, friends, family, or doctor) for meeting my exercise goals.	N/A	N/A	0.79
14. I feel that I would let others down (eg, friends, family, or doctor) if I did not meet my weight goals.	N/A	N/A	0.93
15. I feel that I would let others down (eg, friends, family, or doctor) if I did not meet my dietary goals.	N/A	N/A	0.94
16. I feel that I would let others down (eg, friends, family, or doctor) if I did not meet my exercise goals.	N/A	N/A	0.92
17. Feeling accountable to others (eg, friends, family, or doctor) has helped me control my weight.	N/A	N/A	0.57
18. Feeling accountable to others (eg, friends, family, or doctor) has helped me stay on track with my diet.	N/A	N/A	0.55
19. Feeling accountable to others (eg, friends, family, or doctor) has helped me stay on track with my physical activity.	N/A	N/A	0.54
20. In general, I feel accountable to others to control my weight.	N/A	N/A	0.74

[a]N/A: not applicable.

Study 2

Of the 80 participants included in Study 2, 55 completed the SAM at the 6-month assessment and were thus included in these analyses. Table 3 provides participants' baseline characteristics by group. There were no significant differences between groups in terms of age, baseline BMI, sex, education, or race/ethnicity, all *P* values were >.05.

At the 6-month follow-up, the mean SAM score was 47.86 (SD 20.88); scores ranged from 17.14 to 94.57 (highest possible score of 100). Replicating the analyses conducted in Study 1, the reliability analyses for 6-month SAM scores demonstrated strong internal consistency (Cronbach α=.95).

XSL•FO
RenderX

Table 3. Baseline and demographic characteristics of participants by group in study 2.

Characteristic	Study 2 (n=55)		
	SC[a] (n=17)	TECH[b] (n=17)	TECH+PHONE[c] (n=21)
Age (years), mean (SD)	54.9 (9.7)	46.4 (12.7)	52.9 (11.28)
BMI (kg/m^2), mean (SD)	34.0 (4.0)	33.0 (3.4)	32.2 (3.3)
Gender, n (%)			
Male	4 (23)	2 (11)	4 (19)
Female	13 (76)	15 (88)	17 (81)
Education, n (%)			
High school or less	1 (5)	0 (0)	2 (9)
Vocational training	0 (0)	1 (5)	0 (0)
Some college	4 (23)	3 (17)	5 (23)
College degree	6 (35)	7 (41)	6 (28)
Graduate degree	6 (35)	6 (35)	8 (38)
Race, n (%)			
American Indian or Alaska Native	0 (0)	0 (0)	0 (0)
Asian	0 (0)	0 (0)	0 (0)
Black or African American	2 (11)	2 (11)	0 (0)
White	15 (88)	15 (88)	21 (100)
Other	0 (0)	0 (0)	0 (0)
Ethnicity, n (%)			
Hispanic or Latino	0 (0)	1 (6)	1 (4)
Not Hispanic or Latino	17 (100)	16 (94)	20 (95)

[a]SC: standard in-home scale group.

[b]TECH: technology-based self-monitoring tools group.

[c]TECH+PHONE: technology-based self-monitoring tools plus phone-based support group.

As hypothesized, there were significant differences in the 6-month total SAM scores by treatment group ($F_{2,52}$=20.9; r^2=0.45; P<.001; Table 4). Bonferroni-adjusted posthoc analyses demonstrated significantly higher SAM scores in the TECH+PHONE group than the TECH group (t_{54}=3.70, P<.001; SC: t_{54}=6.39, P<.001). Moreover, the TECH group demonstrated significantly higher SAM scores compared with SC (t_{54}=2.57; P=.04).

Between-group differences in SAM subscales were also examined (Table 4). The Support for Healthy Eating subscale demonstrated a similar pattern to the overall SAM scores. There was a significant difference in Support for Healthy Eating subscale scores by group ($F_{2,52}$=21.4; r^2=0.45; P<.001), with the TECH+PHONE group demonstrating higher subscale scores than both TECH (t_{54}=3.48; P<.001) and SC (t_{54}=6.51; P<.001) groups, and the TECH group demonstrating significantly higher scores compared with the SC group (t_{54}=2.88; P=.02). Although an overall significant difference between groups was demonstrated on the Support for Exercise Habits subscale ($F_{2,52}$=3.9; r^2=0.13; P=.03), there were no significant group differences when conducting Bonferroni-corrected posthoc comparisons. Finally, there was a significant difference between groups in the Perceptions of Accountability subscale ($F_{2,52}$=15.3; r^2=0.37; P<.001), such that the TECH+PHONE group had higher scores on this subscale compared with both TECH (t_{54}=3.96; P<.001) and SC (t_{54}=5.22; P<.001) groups. There were no significant differences in Perceptions of Accountability subscale scores between TECH and SC (P=.71) groups.

Table 5 presents correlations between SAM scores and adherence to weight management behaviors and percent weight change during the intervention. Higher SAM scores were associated with higher adherence to weight management behaviors, including higher scores on subscales representing healthy dietary choices, the use of self-monitoring strategies, and positive psychological coping with weight management challenges. The association between total SAM scores and percent weight change was in the expected direction but not statistically significant (r=−0.26; P=.06).

XSL•FO
RenderX

Table 4. Six-month scores on supportive accountability measure (SAM) and each of the three SAM subscales by intervention group.

Scale	Value, mean (SE)
Supportive accountability measure	
SC[a]	30.9 (3.8)
TECH[b]	44.9 (3.8)
TECH+PHONE[c]	64.0 (3.5)
Factor 1: Support for Healthy Eating	
SC	9.8 (1.2)
TECH	14.5 (1.2)
TECH+PHONE	20.0 (1.1)
Factor 2: Support for Exercise	
SC	9.2 (1.4)
TECH	14.0 (1.4)
TECH+PHONE	13.7 (1.2)
Factor 3: Perceptions of Accountability	
SC	11.9 (2.6)
TECH	16.3 (2.6)
TECH+PHONE	30.3 (2.3)

[a]SC: standard in-home scale group.

[b]TECH: technology-based self-monitoring tools group.

[c]TECH+PHONE: technology-based self-monitoring tools plus phone-based support group.

Table 5. Correlation between the supportive accountability measure (total scores and subscales) and adherence to weight control strategies and weight change from baseline to 6-month posttest.

Scale	Total SAM[a] score		Support for Healthy Eating Habits		Support for Exercise		Perceptions of Accountability	
	r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
Weight Control Strategies Scale total score	0.47	.003[b]	0.26	.06	0.23	.09	0.50	<.001[b]
Dietary choices	0.29	.03[b]	0.14	.30	0.15	.27	0.32	.01[b]
Self-monitoring strategies	0.47	.003[b]	0.28	.03[b]	0.26	.06	0.49	.002[b]
Physical activity	0.23	.08	−0.02	.90	0.15	.28	0.31	.02[b]
Psychological coping	0.45	.007[b]	0.36	.006[b]	0.16	.25	0.46	<.001[b]
Percent weight change from baseline to 6 months	−0.26	.06	−0.21	.11	−0.12	.39	−0.25	.06

[a]SAM: supportive accountability measure.

[b]Statistically significant ($P<.05$).

Discussion

Study 1

Study 1 evaluated the psychometric properties of the novel, theory-based SAM in a sample of adults interested in a behavioral weight management trial. EFA revealed a three-factor solution for the SAM, representing subscales for Support for Healthy Eating Habits, Support for Exercise Habits, and Perceptions of Accountability. All items were retained, leaving 10 items representing social support and 10 items representing accountability, which together form the theoretical basis for the construct of supportive accountability [17]. Overall, the 20-item SAM demonstrated excellent internal consistency. Moreover, SAM total scores were significantly associated with measures of Social Motivation [21] and Social Pressure [22]. However, these correlations were small to moderate in magnitude, consistent with the idea that supportive accountability is a construct distinct from motivation related to fears of upsetting others or feelings of social pressure [17]. Examining the convergent validity of the SAM with other measures of social support with weight loss may yield higher correlations. Overall, SAM showed strong reliability and validity in adults with obesity seeking weight loss.

XSL•FO
RenderX

Study 2

Study 2 evaluated the criterion validity of the SAM. Consistent with the hypotheses, participants provided with additional phone support (TECH+PHONE) reported higher SAM scores at the end of a 6-month weight loss program compared with participants who did not receive phone support (TECH only or SC only). Interestingly, participants who did not receive phone support but were provided with newer, technology-based self-monitoring tools reported higher supportive accountability at the end of the intervention compared with participants randomized to self-monitoring using traditional tools (a standard pedometer, bathroom scale, a calorie reference book, and paper self-monitoring logs used to track physical activity, weight, and caloric intake). Subscale analyses revealed that this was likely driven by increased feelings of social support but not accountability. It may be that the brief, automated feedback provided by these tools gave participants a sense of support. Specifically, these tools provided immediate feedback to participants who met short- and long-term goals (such as notifications on the activity monitor and pushed smartphone notifications and *badges* displayed on the app homepage that functioned as visual reminders of goals that had been met) [24]. Similarly, tailored feedback related to intervention goals, even when provided by automated systems, has been demonstrated to improve weight loss outcomes [10]. Thus, there may be some *digital support* inherent in this feedback. Additionally, the availability of social/community features on the Fitbit platforms (eg, the ability to share step counts with friends, competitions among friends to achieve certain step goals) might have promoted feelings of social support. The parent study for Study 2 was conducted between and 2013 and 2015 when there were only minimal social components to these tools. Future research should investigate whether newer, more comprehensive social/community features further lead to increased feelings of support.

Consistent with the Supportive Accountability Theory [10], the provision of additional phone-based support resulted in the highest SAM scores. Compared with the 2 groups that did not receive phone support (TECH and SC), participants provided with phone support demonstrated significantly higher SAM total scores and subscale scores for Support for Healthy Eating Habits and Perceptions of Accountability. This suggests that ongoing interventionist contact improves support for healthy eating and provides a sense of accountability toward weight management behaviors and weight loss goals. Although the overall test statistic demonstrated significant differences across groups in Support for Exercise Habits, there were no posthoc group differences between any of the groups after Bonferroni corrections. This may have been related to power (Bonferroni corrections represent the most conservative approach but can lead to type II errors, especially in small samples [25]). Moreover, it may be possible that the items on this subscale were influenced less by the type of intervention utilized in Study 2. It would not be expected that items asking about whether supportive individuals “exercised with me,” “offered to exercise with me,” and “changed their schedule so we could exercise together” would be rated higher in participants receiving additional phone-based interventionist support. Thus, future research may investigate whether alternative item wording for exercise may better reflect the social support provided by intervention staff.

Finally, consistent with Mohr’s theory that supportive accountability could increase intervention adherence [16], the results demonstrated that higher SAM scores were associated with better adherence to weight management strategies. Interestingly, these associations appear to be driven by ratings of accountability versus support. The correlations between the Perceptions of Accountability subscale and report of adherence to weight management behaviors on the WCSS (both the WCSS total score and all subscale scores) were larger than associations between weight management adherence and the social support subscales (either Support for Healthy Eating Habits or Support for Exercise Habits). Although there was no statistically significant association between total SAM score and percent weight change during the intervention, the association was in the expected direction and represented a medium effect size. Similarly, the association between the SAM subscales (Support for Healthy Eating, Support for Exercise, and Perceptions of Accountability) and percent weight change was not statistically significant; however, it was in the expected direction.

Overall, data from 2 studies were used to examine the psychometric properties of the SAM, a new survey to assess supportive accountability for weight management behaviors. Across both studies, the SAM demonstrated excellent internal consistency and construct validity. This study has some limitations. Participants were predominantly female and non-Hispanic white, which limits the generalizability of the study results. Future research is needed to examine whether SAM demonstrates a similar factor structure, internal consistency, and validity in more diverse samples. The small sample size in Study 2, combined with the fact that participants in only 2 of the 3 groups were asked to use technology-based self-monitoring tools, precluded the investigation into whether the SAM was associated with objective engagement with the technology-based self-monitoring tools. This study also did not assess test-retest reliability or sensitivity to change over time. Future work in these areas would strengthen confidence in this measure for assessing the construct of supportive accountability and further provide important results that could inform future theory development. Finally, the scoring on this measure was complicated by the different scales used for scoring the accountability items (which used a 7-point scale) and the social support items (which used a 5-point scale). The accountability items were developed before the selection of social support items from the Sallis questionnaire, and a 7-point scale was chosen to optimize variability in responses. Future research should investigate whether this scale performs similarly when all items use the same scale (eg, either 5 or 7 response points).

This research also has notable strengths. The development of SAM was theory-based, relying on the Supportive Accountability Theory by Mohr [17]. The SAM included the use of a widely used validated measure of social support as a basis [18] and built upon this measure by adding psychometrically sound accountability items. Furthermore, reducing the items from the original social support scale makes this new tool a concise measure of support and accountability

XSL•FO
RenderX

that can be used in many technology-mediated interventions that would avoid scale fatigue and respondent burden. In Study 1, the sample size was large, which allowed for initial validation and an EFA. Furthermore, all weight data were objectively assessed. Moreover, in 2 separate samples, the SAM demonstrated both internal consistency and construct validity. Finally, the SAM was included in a randomized trial that involved interventions with and without phone support, which allowed for the examination of criterion validity. These methodologies allowed for a robust examination of the psychometric properties of a new measure of supportive accountability.

Supportive accountability was developed as a construct within the context of technology-mediated *electronic health* interventions [16]; however, this construct may have broader applicability. For example, it is theoretically plausible that supportive accountability may play a role in promoting adherence even in traditional face-to-face behavioral interventions. With the SAM, future studies will be able to investigate whether supportive accountability mediates outcomes in face-to-face programs. This may be particularly important in interventions that rely heavily on social support.

Moreover, although the construct of supportive accountability suggests the importance of human support [16], the SAM would be useful for research focused on the development of technologically mediated support systems. For example, the SAM could be used to evaluate whether tools that integrate social components (eg, leaderboards, competitions among friend groups, or chat rooms/bulletin boards that allow contacts to provide support) increase feelings of supportive accountability and ultimately promote intervention adherence. The SAM could also be used to evaluate whether automated feedback provided via tailoring algorithms or artificial intelligence programs can impact feelings of supportive accountability.

Considering the ever-growing technological innovations, the SAM will help researchers better understand the factors that drive the effectiveness of technology-based treatments. This use of the SAM may thus guide the development of more effective interventions and help improve foundational knowledge regarding the mechanisms that drive treatment effects in technologically mediated treatment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The supportive accountability measure and scoring.
[DOC File , 68 KB - jmir_v22i7e17967_app1.DOC]

Multimedia Appendix 2
Inter-item correlations between supportive accountability measure items.
[PNG File , 41 KB - jmir_v22i7e17967_app2.png]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 94 KB - jmir_v22i7e17967_app3.pdf]

References

1. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite 2015 Dec;95:138-151. [doi: 10.1016/j.appet.2015.07.008] [Medline: 26165415]
2. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266 [FREE Full text] [doi: 10.1371/journal.pone.0104266] [Medline: 25100267]
3. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
4. Plotnikoff RC, Wilczynska M, Cohen KE, Smith JJ, Lubans DR. Integrating smartphone technology, social support and the outdoor physical environment to improve fitness among adults at risk of, or diagnosed with, Type 2 Diabetes: findings from the 'eCoFit' randomized controlled trial. Prev Med 2017 Dec;105:404-411. [doi: 10.1016/j.ypmed.2017.08.027] [Medline: 28887192]
5. Shuger SL, Barry VW, Sui X, McClain A, Hand GA, Wilcox S, et al. Electronic feedback in a diet- and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. Int J Behav Nutr Phys Act 2011 May 18;8:41 [FREE Full text] [doi: 10.1186/1479-5868-8-41] [Medline: 21592351]

XSL•FO
RenderX

6. Lynch SM, Stricker CT, Brown JC, Berardi JM, Vaughn D, Domchek S, et al. Evaluation of a web-based weight loss intervention in overweight cancer survivors aged 50 years and younger. Obes Sci Pract 2017 Mar;3(1):83-94 [FREE Full text] [doi: 10.1002/osp4.98] [Medline: 28392934]
7. Coons MJ, Demott A, Buscemi J, Duncan JM, Pellegrini CA, Steglitz J, et al. Technology interventions to curb obesity: a systematic review of the current literature. Curr Cardiovasc Risk Rep 2012 Apr;6(2):120-134 [FREE Full text] [doi: 10.1007/s12170-012-0222-8] [Medline: 23082235]
8. Leahey TM, Thomas G, Fava JL, Subak LL, Schembri M, Krupel K, et al. Adding evidence-based behavioral weight loss strategies to a statewide wellness campaign: a randomized clinical trial. Am J Public Health 2014 Jul;104(7):1300-1306. [doi: 10.2105/AJPH.2014.301870] [Medline: 24832424]
9. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
10. Kozak AT, Buscemi J, Hawkins MA, Wang ML, Breland JY, Ross KM, et al. Technology-based interventions for weight management: current randomized controlled trial evidence and future directions. J Behav Med 2017 Feb;40(1):99-111 [FREE Full text] [doi: 10.1007/s10865-016-9805-z] [Medline: 27783259]
11. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol 2014 Jul 1;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
12. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
13. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]
14. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
15. Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. Patient Prefer Adherence 2016;10:1547-1559 [FREE Full text] [doi: 10.2147/PPA.S103649] [Medline: 27574404]
16. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. J Am Med Assoc 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
17. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
18. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
19. Kiernan M, Moore SD, Schoffman DE, Lee K, King AC, Taylor CB, et al. Social support for healthy behaviors: scale psychometrics and prediction of weight loss among women in a behavioral program. Obesity (Silver Spring) 2012 Apr;20(4):756-764 [FREE Full text] [doi: 10.1038/oby.2011.293] [Medline: 21996661]
20. Leahey TM, Huedo-Medina T, Grenga A, Gay L, Fernandes Pierre D, Denmat Z, et al. Patient provided e-support for reduced intensity obesity treatment: The INSPIRE Trial. Health Psychology. Forthcoming.
21. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ. Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors. Health Educ Res 2007 Oct;22(5):691-702. [doi: 10.1093/her/cyl148] [Medline: 17138613]
22. LaRose JG, Leahey TM, Hill JO, Wing RR. Differences in motivations and weight loss behaviors in young adults and older adults in the National Weight Control Registry. Obesity (Silver Spring) 2013 Mar;21(3):449-453 [FREE Full text] [doi: 10.1002/oby.20053] [Medline: 23404944]
23. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
24. Ross KM, Qiu P, You L, Wing RR. Characterizing the pattern of weight loss and regain in adults enrolled in a 12-week internet-based weight management program. Obesity (Silver Spring) 2018 Feb;26(2):318-323 [FREE Full text] [doi: 10.1002/oby.22083] [Medline: 29239141]
25. Pinto AM, Fava JL, Raynor HA, LaRose JG, Wing RR. Development and validation of the weight control strategies scale. Obesity (Silver Spring) 2013 Dec;21(12):2429-2436 [FREE Full text] [doi: 10.1002/oby.20368] [Medline: 23512914]
26. Nakagawa S. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav Ecol 2004;15(6):1044-1045. [doi: 10.1093/beheco/arh107]

XSL•FO
RenderX

Abbreviations

EFA: exploratory factor analysis
KMO: Kaiser-Meyer-Olkin
SAM: Supportive Accountability Measure
TSRQ: Treatment Self-Regulation Questionnaire
WCSS: Weight Control Strategies Scale

Edited by G Eysenbach; submitted 24.01.20; peer-reviewed by S Hoffman Marchese, A Keath; comments to author 18.03.20; revised version received 01.04.20; accepted 10.04.20; published 28.07.20.

Please cite as:
Chhabria K, Ross KM, Sacco SJ, Leahey TM
The Assessment of Supportive Accountability in Adults Seeking Obesity Treatment: Psychometric Validation Study
J Med Internet Res 2020;22(7):e17967
URL: https://www.jmir.org/2020/7/e17967
doi:10.2196/17967
PMID:32720911

©Karishma Chhabria, Kathryn M Ross, Shane J Sacco, Tricia M Leahey. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Benefits of Massive Open Online Course Participation: Deductive Thematic Analysis

Elizabeth R Blum[1*], PhD; Terese Stenfors[1*], PhD; Per J Palmgren[1*], DC, PhD

Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[*]all authors contributed equally

Corresponding Author:
Per J Palmgren, DC, PhD
Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Tomtebodavägen 18a
Stockholm, 171 77
Sweden
Phone: 46 8 524 85 294
Email: per.palmgren@ki.se

Abstract

Background: Massive open online courses (MOOCs), as originally conceived, promised to provide educational access to anyone with an internet connection. However, the expansiveness of MOOC education has been found to be somewhat limited. Nonetheless, leading universities continue to offer MOOCs, including many in the health sciences, on a number of private platforms. Therefore, research on online education must include thorough understanding of the role of MOOCs. To date, studies on MOOC participants have focused mainly on learners' assessment of the course. It is known that MOOCs are not reaching the universal audiences that were predicted, and much knowledge has been gained about learners' perceptions of MOOCs. However, there is little scholarship on what learners themselves gain from participating in MOOCs.

Objective: As MOOC development persists and expands, scholars and developers should be made aware of the role of MOOCs in education by examining what these courses do offer their participants. The objective of this qualitative synthesis of a set of MOOC evaluation studies was to explore outcomes for MOOC learners, that is, how the learners themselves benefit from participating in MOOCs.

Methods: To explore MOOC learners' outcomes, we conducted a qualitative synthesis in the form of a deductive thematic analysis, aggregating findings from 17 individual studies selected from an existing systematic review of MOOC evaluation methods. We structured our inquiry using the Kirkpatrick model, considering Kirkpatrick levels 2, 3, and 4 as potential themes in our analysis.

Results: Our analysis identified six types of Kirkpatrick outcomes in 17 studies. Five of these outcomes (learning/general knowledge, skills, attitudes, confidence, and commitment) fit into Kirkpatrick Level 2, while Kirkpatrick Level 3 outcomes concerning behavior/application were seen in four studies. Two additional themes were identified outside of the Kirkpatrick framework: culture and identity outcomes and affective/emotional outcomes. Kirkpatrick Level 4 was not represented among the outcomes we examined.

Conclusions: Our findings point to some gains from MOOCs. While we can expect MOOCs to persist, how learners benefit from the experience of participating in MOOCs remains unclear.

(J Med Internet Res 2020;22(7):e17318) doi:10.2196/17318

KEYWORDS

MOOC; MOOC evaluation; qualitative; thematic analysis; learner; online education; Kirkpatrick framework; outcomes; MOOC outcomes

XSL•FO
RenderX

Introduction

When the first massive open online course (MOOC) was offered in 2008, the MOOC format—free, online, and open to anyone with an internet connection—was touted as revolutionary for its potential to democratize access to educational opportunities due to its theoretically universal availability [1-3]. The earliest MOOCs used a connectivist paradigm in which the course was built from networks of online resources and relied on openness and participation from learners. These so-called cMOOCs had the potential to allow learners to participate in their own education outside the traditional, face-to-face classroom setting and to connect with learners worldwide [4]. Extended MOOCs (xMOOCs) brought the MOOC format back to a more traditional structure, with instructors determining the content while still providing "open" availability to anyone with internet access. In practice, there are limits to what this expansive availability has accomplished [2,5]. However, as MOOCs persist, it is useful to explore their role in education by examining what they do offer their participants.

Learning is a complex phenomenon that can be described from different perspectives. Understanding learning is about understanding not only learning processes but also the conditions that influence—and are influenced by—the learning process [6]. In this paper, learning is understood from a constructivist and social-constructivist perspective in which reality and new understanding are constructed by learners on the basis of their previous knowledge, perceptions, and experiences. Learning thus consists of contextual aspects (ie, teachers present information in a way that enables learners to construct meaning on the basis of their own experiences, with a focus on situating learning in an authentic activity); cognitive aspects (ie, recognizing individuals' perception, memory, and meaning-making); and social aspects (ie, converging on learning as a social activity that occurs through interactions between the learner and others) [7,8]. This conception of learning thus reintegrates the artificial and no longer useful distinction between cMOOCs and xMOOCs [9].

A number of systematic reviews have examined MOOCs [4,10-17]. These reviews indicate that much research on MOOCs focuses on evaluating noncompletion rates and retention vs attrition; learner motivation and engagement as well as other behavioral elements, and how these relate to retention and achievement; implications of the latter for MOOC design; and learners' own assessments of the courses [3,4,10,12,16,18]. Research also points to a lack of studies on learners' own experiences and outcomes [3,4,10]; however, there are some exceptions [14,19]. For example, in their review, Pilli and Admiraal [19] investigated MOOC learner outcomes with the intention of informing MOOC course design. Joksimovic et al [14] argued that outcomes and learner engagement are commonly differentiated in the MOOC literature; however, their systematic review proposes an approach that reconnects the two, especially for MOOCs that do not include assessments (eg, cMOOCs as originally conceived). Joksimović et al [14] built on a model by Reschly and co-workers [20] that conceives of learning outcomes as "proximal" or "distal," with academic, social, and affective outcomes within each; they modified this model for the "nonformal, digital educational settings" of MOOCs [14]. Despite their work on outcomes, Joksimovic and colleagues reiterated the finding that attempts to measure or evaluate the benefits to learners of participating in MOOCs have been mostly limited to date.

Another systematic review by Rowe et al [17] investigated the utility of open online courses (OOCs, including MOOCs) in health professions education. They evaluated the available research with a framework that included five "outcome" categories, including effectiveness (increase in learner knowledge), learner experiences, feasibility, pedagogy, and economics; they concluded that the available evidence neither unequivocally supports nor refutes the use of such courses. Their review was limited to the health professions; however, it highlighted the absence of rigorous research on MOOCs and the concurrent persistence of these courses. Their "effectiveness" category further highlighted the absence of research on benefits to MOOC learners, specifically in the health professions. They argued that the application of MOOCs in health professions education should be limited until a great deal more quality research is performed [17].

In their recent systematic review, Alturkistani et al [21] also added to the discourse on MOOC evaluation methods. Alturkistani et al identified three "evaluation-focused categories" among the studies they reviewed: learner-focused, teaching-focused, and MOOC-focused [21]. We approached this review as a jumping-off point to further synthesize understanding of MOOC learner outcomes. Here, we unpack the learner-focused category in [21] and, more specifically, the "learning outcomes and experience" subcategory to investigate the learner outcomes for the included MOOCs. In our study, "learner outcomes" are direct statements that describe the knowledge, skills, and attitudes that learners have demonstrated or are expected to reliably demonstrate when successfully completing a course. Learner outcomes is an understudied area that warrants further investigation, as MOOCs are a learning environment distinct from traditional classrooms and even other forms of e-learning, and they continue to be embraced as an educational modality [22].

Thus, despite their persistence, MOOCs have not lived up to the early expectation that they would allow widespread, nearly universal access to education. For example, there is consistent evidence that learners who use MOOCs, and indeed those who are more likely to complete them, are generally more educated and affluent [1,23,24]. There is also insufficient evidence that MOOCs are useful in areas such as health professions education [17]. MOOC learners are heterogeneous along numerous dimensions, including native language, prior training, age, economic status, and geographic location [24]. The heterogeneity of the expectations and goals of MOOC learners has also undoubtedly contributed to the difficulty of evaluating MOOCs and characterizing their benefits, a difficulty that is illustrated below in the heterogeneity of the studies reviewed. Thus, if MOOCs are not, in practice, democratizing education, and they have not lived up to traditional learning settings for at least some professional fields, what are they offering? In this study, we focus our attention on what learners do gain from participating in MOOCs, including but not limited to

XSL•FO
RenderX

performance measures; that is, we explore how learners benefit from the experience of participating in MOOCs, including and beyond outcomes directly related to learning.

Methods

We conducted a qualitative synthesis in the form of a deductive thematic analysis, aggregating findings from individual studies, to explore MOOC learners' outcomes. The datasets used and analyzed during the current study are available from the first author on reasonable request. To structure our inquiry, we relied on a commonly used framework for evaluating learning with applications in multiple learning and training settings: the Kirkpatrick model [25]. This model frames training on four levels: (1) *reaction*, (2) *learning*, (3) *behavior*, and (4) *results*. A more recent version [26] updates and clarifies the model, proposing that *reaction* includes customer satisfaction, engagement, and relevance; *learning* includes knowledge, skills, attitude, confidence, and commitment; and *behavior* refers to how the learner applies the learning "on the job." The more recent version of the behavior level adds "processes and systems that reinforce, encourage, and reward performance of critical behaviors on the job" [26], which can be seen as catalysts for applying what has been learned. These processes and systems, which include job aids, coaching, work review, and incentive systems, are referred to in [26] as "required drivers" or factors that increase the likelihood that people will retain and apply what they have learned in a given setting, referred to as "required drivers". *Results* are the targeted outcomes of the training, such as whether the results of the training are seen within an organization; the more recent version adds "leading indicators" (short-term measures that can indicate whether the results are likely to occur) [26].

The studies in the current synthesis derive from Alturkistani et al's systematic review of MOOC evaluation methods [21]. Their review included studies from 2008 to 2018 that focused primarily on MOOC evaluation and studies that reviewed or applied MOOC evaluation methods. Both quantitative and qualitative studies were included, after a careful assessment of their methodological quality, as well as grey literature. During the last few years, the contribution of qualitative evidence has been acknowledged within research [27]. This is in line with the epistemological stance of this review. The complete search strategy and further details of the source review [21] can be found in [18]. Alturkistani *et al* [21] identified 3275 records; after a review procedure, the final review included 33 studies.

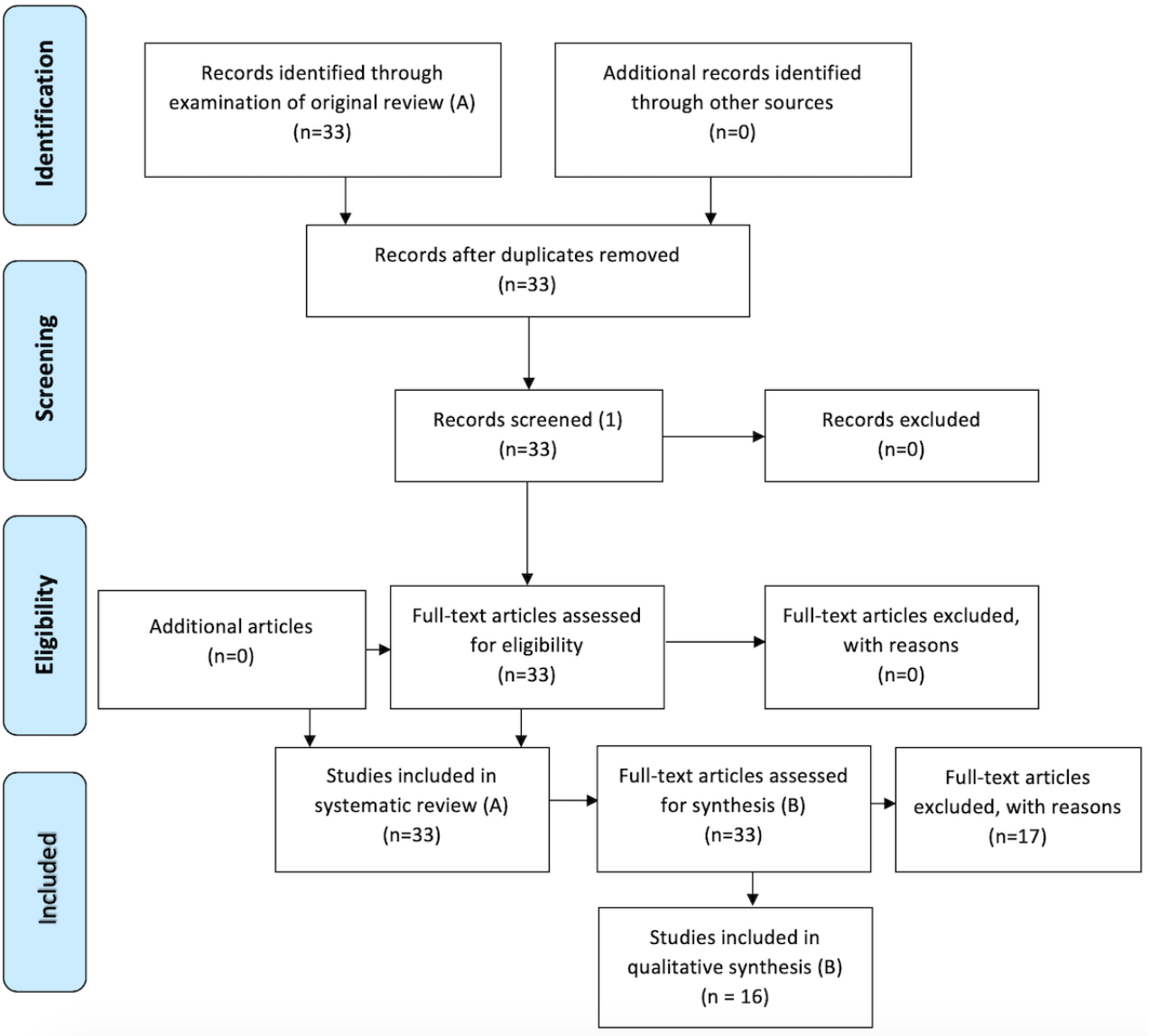
Specifically, Alturkistani et al's "learning outcomes and experiences" subcategory was the basis for the current synthesis, as we looked at what learners gain from the experience of participating in MOOCs. This subcategory included 21 studies. We reviewed each paper in this category for findings that included learners' outcomes. Each study was examined for outcomes specific to the learners themselves. We did not include measures of engagement, motivation, completion, or attrition in our analysis unless they were clearly tied to the outcomes for learners. In an additional step intended to capture all learner outcomes, we examined Multimedia Appendix 3 in Alturkistani et al's review [21], which included all 33 studies. As a result of this review, we excluded 12 studies that did not include clear outcomes for learners (Figure 1), which left 21 studies for our analysis. As the analysis proceeded, we determined that the outcomes in 6 of these 21 studies were not clear enough to include. Notably, we did include one study [28] that was not included in Alturkistani et al's "learner outcomes and experience" category. Of the resulting 16 studies for analysis, 4 had more than one outcome. Multimedia Appendix 1 describes this procedure in detail.

More specifically, in this qualitative synthesis, we performed a deductive thematic analysis [29] where the starting themes were the four Kirkpatrick levels. We extracted all outcomes from the 16 studies; we then placed these in Kirkpatrick level 2, 3, or 4. After this first coding, which was conducted by ERB, TS and PJP reviewed the results. Second, ERB further analyzed the findings in each category according to the subthemes within each Kirkpatrick level. Subsequently, the findings were discussed and subjected to adjustments until consensus among all investigators was reached. Although the aforementioned steps appear to be consecutively ordered, the process of analysis and search for patterns was in no way linear; rather, it was iterative and recursive. No software program was used to aid the analysis. The structure of our analysis allows for the possibility that the same study will have multiple outcomes and thus will appear under more than one level. Level 1 (*reaction*) in the Kirkpatrick model was not of interest to our investigation, as there is a great deal of existing research on learners' assessment of MOOCs.

Outcomes that could not be matched with the Kirkpatrick levels were set aside for a separate inductive thematic analysis, which is presented as "Outcomes beyond Kirkpatrick."

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram: systematic review (A) for a synthesis paper on MOOC learning outcomes (B). Modified from Alturkistani et al [21].



Results

Our analysis resulted in six types of outcomes. These are summarized in Table 1 as framed by the levels in the Kirkpatrick model.

Kirkpatrick Level 2: Learning

Our deductive analysis showed that 15/16 (94%) of the examined studies included one or more outcomes corresponding to Kirkpatrick Level 2. Thus, the Learning theme here incorporates concepts such as knowledge, skill, attitude, confidence, and commitment. Each subtheme is presented using the identified data and illustrated with supporting quotations.

Subtheme: Knowledge

Most of the Level 2 outcomes we identified were scores or survey items that assessed knowledge in some form. For example, in their MOOC on new media in teaching and learning, Chen et al [30] reported "learning performance" via quiz scores and a final paper, for which the participants could earn "Excellence Awards." Four studies in our sample took a longitudinal view of learning outcomes via a pretest/posttest design. For example, Konstan et al [31] used a longitudinal design to test knowledge of technology that predicts preferences based on previous behavior (recommender systems technology); based on precourse and postcourse test scores within their MOOC, the gains in knowledge measured at the end of the course persisted at a 5-month follow-up in most cases. Further, in a MOOC designed to prepare medical students for global health experiences, Jacquet and colleagues [32] found an increase in post-MOOC compared to pre-MOOC test scores. Next, using average quiz and homework scores, Liang et al [33] reported an increase in quiz and homework scores enhanced by participation in online activities. Cross [34] used preassessments and postassessments to track changes in knowledge on a scale from "novice" to "expert," while Colvin et al [35] reported improved scores on postcourse versus precourse tests in introductory physics, and Mackay et al [36] saw a postcourse increase in scores on their assessment of participants' knowledge about animal welfare.

XSL•FO
RenderX

Table 1. Outcomes of MOOC studies framed by Kirkpatrick Level 2 or Level 3.

Kirkpatrick level, subtheme, and study	Data collection	Data analysis	Outcome variables	Outcome findings
Level 2: Learning				
1. General learning/change in knowledge				
Chen et al (2015) [30]	Scores on quizzes and final paper	Inferential statistics	Possible "Excellent Paper," "Excellent Participation," and "Excellent Group Member" awards	Learners received these awards if they fulfilled the criteria
Konstan et al (2015) [31]	Three-part longitudinal design: precourse, postcourse, and 5-month follow-up "knowledge tests" and surveys	Inferential statistics; qualitative analysis	Assessed knowledge of recommender systems[a]	Gains in knowledge and 5-month retention of acquired knowledge
Jacquet et al (2018) [32]	LMS[b] data; pre-MOOC and post-MOOC knowledge tests	Inferential statistics	Score on knowledge test	Increased knowledge score from pretest to posttest
Liang et al (2014) [33]	Assessments: quizzes and homework	Inferential statistics	Average assessment score	Increase in assessment score related to degree of participation
Cross (2013) [34]	Precourse and postcourse surveys; LMS	Descriptive statistics	Knowledge: "novice" to "expert"[a]	Increase in knowledge
Colvin et al (2014) [35]	Normalized gain between pretests and posttests in introductory physics; "ability" based on test items attempted, analyzed with Item Response Theory (IRT)	Inferential statistics	Comparison of pre-MOOC and post-MOOC physics knowledge and "ability"	Learning (measured via posttest score) across several cohorts identified using IRT
MacKay et al (2016)[36]	Precourse and postcourse assessments of animal welfare knowledge	Inferential statistics	Scores on animal welfare knowledge assessment	Increased scores
2. Skill				
Brunton et al (2017)[37]	Weekly Likert scale quizzes during the MOOC: "individual digital readiness tools" and postcourse quiz	Descriptive statistics	Preparedness for online learning[a]	Self-assessed changes in preparedness for online learning
Rubio (2015)[38]	Precourse and postcourse comprehensibility ratings	Inferential statistics	Spanish comprehensibility (language pronunciation)	Increased comprehensibility in postcourse ratings
Stephens and Jones (2014) [39]	Precourse and postcourse surveys with mostly open-ended items	Content analysis	Skills discovery[a]	Technological skills
Liu et al (2014) [40]	End-of-course surveys (Likert scale and open-ended); email interviews	Descriptive and thematic analysis (focused coding)	Three things students learned[a]	Skills in data visualization, critiquing, and creating infographics
3. Commitment				
Alturkistani et al (2018) [41]	Case studies; interviews	Thematic analysis	Learning achievement; use of information in the workplace[a]	Intention to apply knowledge
4. Attitude				
MacKay et al (2016) [36]	Multiple-choice quizzes; confidence and attitude surveys (mostly Likert scale)	Inferential statistics	Change in attitudes; certificate of achievement for completion[a]	Change in attitude
5. Confidence				

XSL•FO
RenderX

Kirkpatrick level, subtheme, and study	Data collection	Data analysis	Outcome variables	Outcome findings
Hossain et al (2015) [28]	Ten-point scale; confidence-to-treat	Inferential statistics	Confidence to treat spinal cord injury[a]	Gains in confidence
Cross (2013) [34]	Precourse/postcourse survey; LMS	Descriptive statistics	Confidence to apply learning[a]	Gains in confidence
Mackness et al (2013) [42]	Interviews (face-to-face and email) and focus groups; assessment of microteaching	Qualitative case study approach	Confidence to participate in social learning environments[a]	Gains in confidence
Lei et al (2015) [43]	Pre-MOOC and post-MOOC surveys; forum threads	Sentiment analysis	Identity and confidence[a]	Confidence in work; confidence to inspire
Milligan and Littlejohn (2014) [44]	Interviews mid-MOOC	Qualitative analysis	Changes in practice[a]	Confidence about practices on the job
Level 3: Behavior				
Behavior/Application				
Milligan and Littlejohn (2014) [44]	Survey and interview	Qualitative analysis	Application of learning in professional practice[a]	Integrating new understanding in practice
Lei et al (2015) [43]	pre-MOOC and post-MOOC surveys; forum threads	Sentiment analysis	Effects on learners and community[a]	Bringing knowledge back to community
Cross (2013) [34]	Precourse/postcourse survey; LMS	Descriptive statistics	Changes in practice[a]	Implementation of tools in course design
Konstan et al (2015) [31]	Follow-up interview and survey	Inferential statistics	Application of new recommender system skills[a]	Application of systems at work, school, business

[a]Includes a self-report.

[b]LMS: learning management system.

Subtheme: Skill

We found several examples of skill outcomes, including self-assessed preparedness (readiness for online education) [37] and improvement in Spanish language pronunciation and comprehensibility measured by pre-MOOC and post-MOOC assessments [38]. Further, participants in a library and information science MOOC were asked "What did you gain most from taking part in the MOOC?" Their responses included "Students gained new technological skills through their learning experience" [39]. Liu et al [40] found that learners gained skill through learning to "visualize data and critique infographics (and) learning visualization concepts and…tool use"; these were the most frequently cited "three things [students] had learned" in a journalism MOOC.

Subthemes: Commitment, Attitude, and Confidence

Other Level 2 outcomes were commitment, as shown through intention to apply knowledge [41]; attitude about animal welfare [36]; and confidence to treat patients, as measured in a randomized control trial study comparing a MOOC with a self-directed online learning module [28]. Additionally, Cross [34] reported that learners gained confidence with regard to applying what they had learned, and Mackness et al [42] also reported confidence to participate in various interactive learning activities:

> *They also gain the confidence to attend and contribute to live synchronous sessions, to openly share their work and ideas, and to cooperate and/or collaborate in social networking environments. "They shift from being consumers to producers."*

In their MOOC on Asian vernacular architecture, Lei et al [43] used a case study design to investigate learners' postlearning experiences, asking, "How has the course influenced learners and their surrounding community?" This influence is reflected in the following learner's experience:

> *It is through learning that I have gained the most confidence, in my identity and in my work. And I hope that this course would be the one of many stepping stones towards me being able to help inspire and nurture future generations….*

Using a clinical trials MOOC, Milligan and Littlejohn [44] asked learners halfway through the course "to reflect on how their practice had changed as a result of the course." Some learners had already seen an effect on their confidence and perspective: "These respondents reported a range of general benefits: that the course had given them a new perspective, had made them assured, or had helped them bring a greater criticality to their practice." One participant stated, "I know why and why not…you have an overview, I cannot say I apply everything in my day to day work, but the fact that you feel more confident,

XSL•FO
RenderX

for me, it helps a lot." This outcome in turn intersects with Kirkpatrick Level 3, as discussed in the next section.

Kirkpatrick Level 3: Behavior: Application

Our analysis found 4/21 studies (19%) with evidence of Level 3 outcomes. Level 3 includes application via critical behaviors plus the presence of outcomes that make it more likely that people will retain and apply what they have learned in a given setting (the abovementioned catalysts for application or "required drivers").

In addition to effects on confidence (Level 2), Milligan and Littlejohn [44] found evidence of Level 3 outcomes from their clinical trials MOOC; in answer to the same question as above (how their practice changed as a result of the course), most learners reported having already incorporated their learning. For example, the respondent quoted above also reported immediate effects: "Well, it gives me a better understanding of why I do what I do…I understand why I have to submit my protocol or a complete or total submission to authorities, how a protocol has been developed." [44] Another respondent said, "It is much, much better, I could address all of the challenges much better and make better decisions, and actually I participate with this CRO in developing the protocol and the study documents and everything."

Lei and colleagues [43] described effects on how the learners brought their experience back to their communities, a behavioral application which reflects Kirkpatrick Level 3. For example, one participant from an area damaged by earthquakes reflected:

> *This course helped me to see the significance of the collapsed houses, temples, shrines, monuments and courtyards in a different angle which otherwise I would not have been able to see…I have already started contributing my knowledge with the local community as we come together to rebuild what has been destroyed.*

Cross [28] described learners' goals, including plans to implement tools from the MOOC in their course design; some learners reported having already done so, which is another example of application of the MOOC experience. Employing a longitudinal study design, Konstan et al [31] investigated MOOC learners' application of course content (recommender systems technology). Kirkpatrick Level 3 behaviors are evident in the participants' reports of incorporating the systems at work, school, or in entrepreneurial settings, and some also applied the underlying algorithms in other contexts.

Kirkpatrick Level 4

In this qualitative synthesis, we did not find any data congruent with Kirkpatrick Level 4, which includes outcomes and "leading indicators."

Outcomes Beyond Kirkpatrick

Not all of the outcomes described in the studies are congruent or align well with the Kirkpatrick framework; hence, we present these outcomes separately here. After our inductive thematic analysis, we identified two themes among these outcomes: "culture and identity" outcomes and "affective/emotional" outcomes. Culture and identity outcomes included "insights about themselves through personal reflection about their learning styles, professional practices, and the ways they view the world" [39], as well as connection to a community, whether of fellow educators [39,42] or those with a shared cultural heritage [43]. Affective outcomes such as "excitement" and "inspiration" are evident in [39], where learners gained "inspiration, energy, and excitement about the field."

Discussion

Principal Results

In this qualitative analysis, we explored the benefits that MOOCs in a broad range of subjects offer their participants. We synthesized the types of outcomes reported in a set of MOOC studies, including but not limited to outcomes that assess learning in some way. Using the Kirkpatrick model as a framework, the most prominent findings were that most of the MOOCs described in the included studies only had outcomes that could be categorized as Kirkpatrick Level 2. Kirkpatrick Level 3 outcomes were also represented, although these were not as common as Level 2 outcomes. We did not observe any Kirkpatrick Level 4 outcomes in the data we analyzed. If a MOOC were to aim for or result in Level 4 outcomes, we would expect to see changes at the organizational level. This might reveal itself in the form of implemented changes in policy in a health care setting after a group of managers participated in a policy MOOC, or in the case of higher education, a change in pedagogical training for educators after a MOOC was attended by several faculty members. Our complementary analysis of outcomes that did not align with Kirkpatrick yielded two additional themes.

Previous Research

Previous research has shown that students generally perform better in face-to-face courses than in online courses [45], and several of the studies in our review used comparisons between MOOC and non-MOOC learning contexts. The studies analyzed in this study did not report outcomes that were unique to MOOCs; however, they did provide insight into what MOOCs do and do not offer to participants. For example, in a randomized control trial by Hossain et al [28] comparing a self-paced online course with an online course with MOOC-based guidance and study tips, improvement in knowledge of spinal cord injury treatment as well as gains in confidence to treat were observed after both courses; however, there was no advantage in the MOOC group. Additionally, Chen and coworkers [30] found no difference in scores on assignments between an online and an onsite version of a digital media course. Colvin et al [35] compared learning gains measured in their MOOC with learning gains in traditional settings; they found evidence of learning in the MOOC, in which scores were slightly higher than typical for a comparable lecture-based course but significantly lower than those seen in other courses with an "interactive engagement" component. In a finding that appears counter to the above, Rubio [38] found that improvement in language comprehensibility was greater in a MOOC compared to a face-to-face course. Finally, in their review, Rowe et al [17] looked specifically at the effectiveness of MOOCs in health professions education; they concluded that it cannot be said that

XSL•FO
RenderX

MOOCs "enhance student learning" despite the proliferation of MOOCs and the "hype" about their potential. These contradictory findings suggest that when comparing MOOCs to other learning formats, the benefits of MOOCs remain unclear.

MOOCs were also expected to foster and build social networks. However, in reality, the amount of interaction among MOOC participants is often limited, and a small proportion of learners are usually responsible for most of this interaction. This finding was reinforced by the studies we examined [42,43,45]. However, there are social elements to MOOC participation, as discussed in the Outcomes Beyond Kirkpatrick section above. Joksimović and colleagues [14] proposed a model that may be a useful framework for illuminating some of the outcomes that do not readily fit with the Kirkpatrick framework. Their model considers social outcomes (along with academic and affective outcomes) in "immediate," "course-level," and "postcourse" settings. Since affective and social outcomes are evident in the studies critically analyzed here, it is worthwhile to consider them as benefits to MOOC participation, which may warrant additional research in its own right; the model proposed by Joksimović and colleagues [14] may be a useful starting point.

Methodological Considerations

Using a well-known model to frame and lens our findings, in this study, we explored one understudied aspect of MOOCs that provides a view of what learners can gain from MOOCs. The richness of data using an in-depth secondary analysis of a small number of studies from a systematic review with broad subject matter, combined with frequent debriefing sessions and investigator triangulation, enhanced the credibility of the findings. We argue that qualitatively synthesizing existing data in an attempt to make sense of contextually and methodologically diverse findings is an important contribution to the scholarly literature. There are also some limitations to this study. Synthesizing both quantitative and qualitative data is a daunting task, as these data derive from very different paradigms. Thus, an important factor limiting the applicability of our findings is the problem with extracting results from eclectic and dissimilar studies, including qualitative and quantitative methods and grey literature, and attempting to contrast and compare them. The findings should thus be interpreted with due caution in light of this fact. Further, as our work builds on a previous review, we included only studies that were included therein. This may leave out some relevant studies, despite the rigorous inclusion criteria of the previous review. Finally, despite the frequent scholarly use of the Kirkpatrick framework, there are some inherent limitations to the model that also have implications for this work. It has been argued that the four-level model depicts an oversimplified view of learning and training effectiveness that does not take individual or contextual influences into account in the evaluation of the learning that occurs [46]. Thus, using the Kirkpatrick framework deductively as in this study and assorting "contextual" data into predefined themes was challenging. Further, Kirkpatrick's model assumes that the four levels denote a causal chain in which positive reactions lead to greater learning and training, yielding greater transfer and, consequently, more positive results. While the Kirkpatrick model is vague about the causal relationships between level outcomes, it does imply that a simple causal relationship exists between the levels in the model [47]. Finally, in this study, we examined data that were not congruent with the framework but which are nonetheless important to the discussion of MOOC outcomes. For example, when considering the outcomes reported in the studies we reviewed, we chose not to include outcomes we viewed as belonging to Kirkpatrick Level 1, Reaction. This level is usually reserved for outcomes that reflect a participant's reaction to a particular program or training. Since this may include how the participants "feel" about the program in question, Level 1 outcomes can certainly include an affective state in relation to the training. We found some outcomes that we described as "affective," which included "feelings" such as excitement and inspiration. However, these feelings did not refer to the MOOC (training) itself. Instead, the "excitement" and "inspiration" were feelings about the subject of the MOOC as a result of the MOOC, which does not seem to us to fall clearly within Kirkpatrick Level 1. We believe that these feelings may even fall under Kirkpatrick Level 2 in the "Attitude" category; however, we made the conservative decision to separate them. Whether these feelings are part of a Kirkpatrick framework would be an interesting topic for further inquiry.

Conclusions

Our findings point to some gains from MOOCs, and while we can expect MOOCs to persist, how learners benefit from the experience of participating in these courses remains unclear. This is especially true when comparing MOOCs to other learning modes, as evidenced by the comparative studies included in our sample. In our study, we looked for gains or benefits to MOOC learners in all subject areas, and we used the Kirkpatrick framework to explore what learners might gain. From a diverse set of studies, we found outcomes that included changes in knowledge, skills, attitude, and confidence as well as changes in behavior, increased excitement about a subject, and effects on cultural identity as a result of MOOC participation. Thus, beyond outcomes that can be classified as "learning," such as increased knowledge or skill, it does appear that MOOCs provide some value for participants via the gains described above.

In contrast to systematic reviews of MOOC research, we carried out a deeper qualitative analysis of a set of studies from one systematic review that looked only at MOOC evaluation methods. Thus, as an extension of Alturkistani et al [21], we sought to identify MOOC outcomes that benefit the learner. With a qualitative investigation of a subset of studies on MOOC evaluation methods, we were able to apply the Kirkpatrick framework to identify a number of types of learner outcomes. However, as others have pointed out, the absence of systematic ways of measuring the benefits to learners is evident in our synthesis, and work remains to be done to determine the role of MOOCs and what they offer to participants and to the world.

XSL•FO
RenderX

Acknowledgments

We would like to thank Abrar Alturkistani, MPH and Edward Meinart, PhD for collaboration on the EIT Health project that inspired this paper. We also thank Hanna Augustsson, PhD for her expert review of our application of the Kirkpatrick model. All conclusions are the authors' own. This work was partially funded by EIT Health (Grant 18654).

Authors' Contributions

ERB, TS, and PJP contributed to the conception, study design, data collection, analysis and interpretation, and drafting and critical revision of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the analytic procedure for the qualitative synthesis.
[DOCX File , 18 KB - jmir_v22i7e17318_app1.docx]

References

1. Hollands FM, Tirthali D. Institute of Education Sciences. New York: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University; 2014. MOOCs: Expectations and Reality. Full Report URL: https://files.eric.ed.gov/fulltext/ED547237.pdf [accessed 2020-05-14]
2. de Freitas SI, Morgan J, Gibson D. Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. Br J Educ Technol 2015 Apr 08;46(3):455-471. [doi: 10.1111/bjet.12268]
3. Zhu M, Sari A, Lee MM. A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). Internet High Educ 2018 Apr;37:31-39. [doi: 10.1016/j.iheduc.2018.01.002]
4. Veletsianos G, Shepherdson P. A Systematic Analysis and Synthesis of the Empirical MOOC Literature Published in 2013–2015. IRRODL 2016 Mar 01;17(2):198-221. [doi: 10.19173/irrodl.v17i2.2448]
5. Blanco Á, García-Peñalvo F, Sein-Echaluce M. A methodology proposal for developing Adaptive cMOOC. 2013 Nov Presented at: Conference: Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality; November 2013; Salamanca, Spain p. 553-558 URL: https://www.researchgate.net/publication/262167521_A_methodology_proposal_for_developing_Adaptive_cMOOC [doi: 10.1145/2536536.2536621]
6. Illeris K. A comprehensive understanding of human learning. In: Illeris K, editor. Contemporary Theories of Learning. London: Routledge; 2009:7-20.
7. Mann K, Dornan T, Teunissen P. Perspectives on learning. In: Dornan T, Mann K, Scherpbier A, Spencer J, editors. Medical Education: Theory and Practice. Edinburgh: Churchill Livingstone Elsevier; 2011:17-38.
8. Simons RJ, van der Linden J, Duffy T, editors. New Learning. Netherlands: Springer; 2000.
9. Bayne S, Ross J. The Higher Education Academy. 2014. The pedagogy of the Massive Open Online Course (MOOC): the UK view URL: https://www.advance-he.ac.uk/knowledge-hub/pedagogy-massive-open-online-course-mooc-uk-view [accessed 2020-05-14]
10. Liyanagunawardena TR, Adams AA, Williams SA. MOOCs: A systematic study of the published literature 2008-2012. IRRODL 2013 Jul 05;14(3):202. [doi: 10.19173/irrodl.v14i3.1455]
11. Liyanagunawardena TR, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014 Aug 14;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
12. Zhu M, Sari A, Lee MM. A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). Internet High Educ 2018 Apr;37:31-39 [FREE Full text] [doi: 10.1016/J.IHEDUC.2018.01.002]
13. Deng R, Benckendorff P, Gannaway D. Progress and new directions for teaching and learning in MOOCs. Comput Educ 2019 Feb;129:48-60. [doi: 10.1016/j.compedu.2018.10.019]
14. Joksimović S, Poquet O, Kovanović V, Dowell N, Mills C, Gašević D, et al. How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs. Rev Educ Res 2017 Nov 14;88(1):43-86. [doi: 10.3102/0034654317740335]
15. Bozkurt A, Akgün-Özbek E, Zawacki-Richter O. Trends and Patterns in Massive Open Online Courses: Review and Content Analysis of Research on MOOCs (2008-2015). IRRODL 2017 Aug 15;18(5):1-30. [doi: 10.19173/irrodl.v18i5.3080]
16. Deng R, Benckendorff P. A Contemporary Review of Research Methods Adopted to Understand Students' and Instructors' Use of Massive Open Online Courses (MOOCs). Int J Inf Educ 2017;7(8):601-607. [doi: 10.18178/ijiet.2017.7.8.939]
17. Rowe M, Osadnik CR, Pritchard S, Maloney S. These may not be the courses you are seeking: a systematic review of open online courses in health professions education. BMC Med Educ 2019 Sep 14;19(1):356 [FREE Full text] [doi: 10.1186/s12909-019-1774-9] [Medline: 31521150]
18. Foley K, Alturkistani A, Carter A, Stenfors T, Blum E, Car J, et al. Massive Open Online Courses (MOOC) Evaluation Methods: Protocol for a Systematic Review. JMIR Res Protoc 2019 Mar 07;8(3):e12087. [doi: 10.2196/12087]

XSL•FO
RenderX

19. Pilli O, Admiraal W. Students' Learning Outcomes in Massive Open Online Courses (MOOCs): Some Suggestions for Course Design. Yuksekogretim Derg 2017 Apr 29;7(1):46-71. [doi: 10.2399/yod.17.001]
20. Reschly AL, Christenson SL. Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In: Christenson SL, Reschly AL, Wylie C, editors. Handbook of Research on Student Engagement. New York: Springer Science + Business Media; 2012:3-19.
21. Alturkistani A, Lam C, Foley K, Stenfors T, Blum E, Van Velthoven MH, et al. Massive Open Online Course Evaluation Methods: Systematic Review. J Med Internet Res 2020 Apr 27;22(4):e13851 [FREE Full text] [doi: 10.2196/13851] [Medline: 32338618]
22. Literat I. Implications of massive open online courses for higher education: mitigating or reifying educational inequities? High Educ Res Dev 2015 May 05;34(6):1164-1177. [doi: 10.1080/07294360.2015.1024624]
23. Hansen JD, Reich J. Democratizing education? Examining access and usage patterns in massive open online courses. Science 2015 Dec 04;350(6265):1245-1248. [doi: 10.1126/science.aab3782] [Medline: 26785488]
24. Reich J, Ruipérez-Valiente JA. The MOOC pivot. Science 2019 Jan 11;363(6423):130-131. [doi: 10.1126/science.aav7958] [Medline: 30630920]
25. Kirkpatrick DL, Kirkpatrick JD. Evaluating training programs: The Four Levels. Oakland: Berret-Koehler; 2006.
26. Kirkpatrick Partners. 2019. The New World Kirkpatrick Model URL: https://www.kirkpatrickpartners.com/Our-Philosophy/The-New-World-Kirkpatrick-Model [accessed 2020-05-15]
27. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Syst Rev 2016 May 04;5:74 [FREE Full text] [doi: 10.1186/s13643-016-0249-x] [Medline: 27145932]
28. Hossain MS, Shofiqul Islam M, Glinsky JV, Lowe R, Lowe T, Harvey LA. A massive open online course (MOOC) can be used to teach physiotherapy students about spinal cord injuries: a randomised trial. J Physiother 2015 Jan;61(1):21-27 [FREE Full text] [doi: 10.1016/j.jphys.2014.09.008] [Medline: 25498151]
29. Clarke V, Braun V. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. Psychologist 2013 Feb;26(2):120-123 [FREE Full text]
30. Chen W, Jia J, Miao J, Wu X, Wang A, Yang B. Assessing Students' Learning Experience and Achievements in a Medium-Sized Massively Open Online Course. 2015 Presented at: IEEE 15th International Conference on Advanced Learning Technologies; 2015; Hualien. [doi: 10.1109/ICALT35929.2015]
31. Konstan JA, Walker JD, Brooks DC, Brown K, Ekstrand MD. Teaching Recommender Systems at Large Scale: Evaluation and Lessons Learned from a Hybrid Mooc. ACM Trans. Comput.-Hum. Interact 2015 Apr 15;22(2):1-23. [doi: 10.1145/2728171]
32. Jacquet GA, Umoren RA, Hayward AS, Myers JG, Modi P, Dunlop SJ, et al. The Practitioner's Guide to Global Health: an interactive, online, open-access curriculum preparing medical learners for global health experiences. Medical Education Online 2018 Aug 07;23(1):1-8. [doi: 10.1080/10872981.2018.1503914]
33. Liang D, Jia J, Wu X, Miao J, Wang A. Analysis of learners' behaviors and learning outcomes in a massive open online course. Knowledge Management & E-Learning 2014 Sep;6(3):281-298. [doi: 10.34105/j.kmel.2014.06.019]
34. Cross S. Evaluation of the OLDS MOOC curriculum design course: participant perspectives, expectations and experiences. Milton Keynes: Open University; 2013. URL: http://oro.open.ac.uk/37836/1/EvaluationReport_OLDSMOOC_v1.0.pdf [accessed 2020-05-15]
35. Colvin KF, Champaign J, Liu A, Zhou Q, Fredericks C, Pritchard DE. Learning in an introductory physics MOOC: All cohorts learn equally, including an on-campus class. IRRODL 2014 Aug 15;15(4):1-21. [doi: 10.19173/irrodl.v15i4.1902]
36. MacKay JR, Langford F, Waran N. Massive Open Online Courses as a Tool for Global Animal Welfare Education. J Vet Med Educ 2016 Jan;43(3):287-301. [doi: 10.3138/jvme.0415-054r2]
37. Brunton J, Brown M, Costello E, Farrell O, Mahon C. Giving Flexible Learners a Head Start on Higher Education: Designing and Implementing a Pre-induction Socialisation MOOC. 2017 Presented at: European Conference on Massive Open Online Courses; 2017; Madrid. [doi: 10.1007/978-3-319-59044-8_2]
38. Rubio F. Teaching pronunciation and comprehensibility in a language MOOC. In: Martín-Monje E, Bárcena E, editors. Language MOOCs: Providing Learning, Transcending Boundaries. Berlin: De Gruyter; 2015:143-160.
39. Stephens M, Jones K. MOOCs as LIS Professional Development Platforms: Evaluating and Refining SJSU's First Not-for-Credit MOOC. Journal of Education for Library and Information Science 2014;55(4):345-361.
40. Liu M, Kang J, Cao M, Lim M, Ko Y, Myers R, et al. Understanding MOOCs as an Emerging Online Learning Tool: Perspectives From the Students. Am J Distance Educ 2014 Aug 26;28(3):147-159. [doi: 10.1080/08923647.2014.926145]
41. Alturkistani A, Car J, Majeed A, Brindley D, Wells G, Meinert E. Determining The Effectiveness Of A Massive Open Online Course In Data Science For Health. 2018 Presented at: International Association for Development of the Information Society (IADIS) International Conference on e-Learning; 2018; Madrid.
42. Mackness J, Waite M, Roberts G, Lovegrove E. Learning in a small, task–oriented, connectivist MOOC: Pedagogical issues and implications for higher education. IRRODL 2013 Sep 30;14(4):1-20. [doi: 10.19173/irrodl.v14i4.1548]
43. Lei C, Hou X, Kwok T, Chan T, Lee J, Oh E. Advancing MOOC and SPOC development via a learner decision journey analytic framework. 2015 Presented at: IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE); 2015; Zhuhai. [doi: 10.1109/tale.2015.7386034]

XSL•FO
RenderX

44. Milligan C, Littlejohn A. Supporting professional learning in a massive open online course. IRRODL 2014 Oct 03;15(5):197-213. [doi: 10.19173/irrodl.v15i5.1855]
45. Tawfik AA, Reeves TD, Stich AE, Gill A, Hong C, McDade J, et al. The nature and level of learner–learner interaction in a chemistry massive open online course (MOOC). J Comput High Educ 2017 Mar 8;29(3):411-431. [doi: 10.1007/s12528-017-9135-3]
46. Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Eval Program Plann 2004 Aug;27(3):341-347. [doi: 10.1016/j.evalprogplan.2004.04.011]
47. Holton EF. The flawed four-level evaluation model. Hum Resour Dev Q 1996;7(1):5-21. [doi: 10.1002/hrdq.3920070103]

Abbreviations

cMOOC: connectivist massive open online course
LMS: learning management system
MOOC: massive open online course
OOC: open online course
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
xMOOC: extended massive open online course

Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by L Blakemore, A Giordano; comments to author 03.02.20; revised version received 21.02.20; accepted 23.03.20; published 08.07.20.

Please cite as:
Blum ER, Stenfors T, Palmgren PJ
Benefits of Massive Open Online Course Participation: Deductive Thematic Analysis
J Med Internet Res 2020;22(7):e17318
URL: https://www.jmir.org/2020/7/e17318
doi:10.2196/17318
PMID:32672680

©Elizabeth R Blum, Terese Stenfors, Per J Palmgren. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of a Low-Cost Portable 3D Virtual Reality Gesture-Mediated Simulator for Training and Learning Basic Psychomotor Skills in Minimally Invasive Surgery: Development and Content Validity Study

Fernando Alvarez-Lopez[1*], MD; Marcelo Fabián Maina[2*], PhD; Francesc Saigí-Rubió[3*], PhD

[1]Faculty of Health Sciences, Universidad de Manizales, Caldas, Colombia

[2]Faculty of Psychology and Education Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[3]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[*]all authors contributed equally

Corresponding Author:
Francesc Saigí-Rubió, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Avinguda Tibidabo, 39-43
Barcelona, 08035
Spain
Phone: 34 933263622
Email: fsaigi@uoc.edu

Abstract

Background: Simulation in virtual environments has become a new paradigm for surgeon training in minimally invasive surgery (MIS). However, this technology is expensive and difficult to access.

Objective: This study aims first to describe the development of a new gesture-based simulator for learning skills in MIS and, second, to establish its fidelity to the criterion and sources of content-related validity evidence.

Methods: For the development of the gesture-mediated simulator for MIS using virtual reality (SIMISGEST-VR), a design-based research (DBR) paradigm was adopted. For the second objective, 30 participants completed a questionnaire, with responses scored on a 5-point Likert scale. A literature review on the validity of the MIS training-VR (MIST-VR) was conducted. The study of fidelity to the criterion was rated using a 10-item questionnaire, while the sources of content-related validity evidence were assessed using 10 questions about the simulator training capacity and 6 questions about MIS tasks, and an iterative process of instrument pilot testing was performed.

Results: A *good enough* prototype of a gesture-based simulator was developed with metrics and feedback for learning psychomotor skills in MIS. As per the survey conducted to assess the fidelity to the criterion, all 30 participants felt that most aspects of the simulator were adequately realistic and that it could be used as a tool for teaching basic psychomotor skills in laparoscopic surgery (Likert score: 4.07-4.73). The sources of content-related validity evidence showed that this study's simulator is a reliable training tool and that the exercises enable learning of the basic psychomotor skills required in MIS (Likert score: 4.28-4.67).

Conclusions: The development of gesture-based 3D virtual environments for training and learning basic psychomotor skills in MIS opens up a new approach to low-cost, portable simulation that allows ubiquitous learning and preoperative warm-up. Fidelity to the criterion was duly evaluated, which allowed a good enough prototype to be achieved. Content-related validity evidence for SIMISGEST-VR was also obtained.

(J Med Internet Res 2020;22(7):e17491) doi:10.2196/17491

KEYWORDS

simulation training; minimally invasive surgery; user-computer interface; operating rooms; medical education; computer-assisted surgery

XSL•FO
RenderX

Introduction

Background

The emergence of minimally invasive surgery (MIS) in the mid-1980s [1] led to an increase in the number of iatrogenic bile duct injuries, as many surgeons worldwide switched from the paradigm of open surgery to these procedures with no previous training [2,3]. In the wake of these developments, simulation became valuable as a tool for learning psychomotor skills in MIS and numerous studies have demonstrated its usefulness [4,5].

Simulators for skill learning in MIS can be classified into 3 large groups: (1) traditional box trainers, (2) augmented reality simulators (hybrids), and (3) virtual reality (VR) simulators [6,7]. The last two are expensive and are unavailable in most universities or hospitals in developing countries [8]. The first VR simulator for MIS training was MIS training–VR (MIST-VR) [9]. In 1998, evidence for the construct validity of the device was established [10]. Later, in 2002, the evidence for prediction validity was added [4,11]. Finally, from 2002 onward, the evidence for concurrent validity was also demonstrated [12,13]. Recent years have seen the development of low-cost, gesture-based touchless devices that can interact with 3D virtual environments, such as the Microsoft Kinect (MS Kinect, Microsoft Corp), Leap Motion Controller (LMC; Leap Motion Inc), and the Myo armband (Thalmic Labs, Kitchener) [14].

For the development of the simulator used in this study, the researchers adopted the design-based research (DBR) paradigm, also known as *design research*. DBR seeks the creation and validation of useful artifacts that do not exist in nature [15] and is described by Manson [16] as "a process of using knowledge to design and create useful artefacts, and then using various rigorous methods to analyze why, or why not, a particular artefact is effective. The understanding gained during the analysis phase feeds back into and builds the body of knowledge of the discipline." DBR is a solution-oriented process that focuses on solving practical and complex real-world problems [17]. The artifacts created can be constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), *instantiations* (implemented and prototype systems), and better design theories [18,19]. To develop the simulator, the study followed the method proposed by Manson [16,20,21], in which using processes of abduction and deduction that detect errors in the design or function of the prototype, supports the development of improved versions until a sufficiently *good enough* functional product is obtained that can be subjected to validation studies [16,20-22]. These *good enough* devices are rarely complete and are functional systems ready to be used in practice; rather, they are innovations that define the ideas, practices, technical capabilities, and products using which systems analysis, design, implementation, and use are achieved effectively and efficiently [17].

Objectives

The first aim of this study was to describe the development of a web-based 3D VR simulator mediated by a gesture interface device (LMC) for learning basic psychomotor skills in MIS, called gesture-mediated simulator for MIS–VR (SIMISGEST-VR). The device is characterized by its portability and low cost, as well as the possibility of learning and training at any time and place (ubiquitous learning). The second aim of this study was to evaluate fidelity to the criterion and to find sources of content-related validity evidence for SIMISGEST-VR.

Methods

Overview

This is a descriptive report of the development, using a DBR paradigm, of a gesture-mediated simulator for learning basic psychomotor skills and of the prospective evaluation of the data obtained from Likert scale surveys to evaluate fidelity to the criterion and the sources of content-related evidence. To this end, the study participants rated fidelity to the criterion using a 10-item questionnaire about its ease of use, relevance as a tool for simulation in MIS, degree of correspondence between the movements of the forceps and their representation in the virtual space, and feedback. The sources of content-related validity evidence were (1) a literature review on a previously validated tool, the MIST-VR, and (2) an expert panel that answered 10 questions about the training capacity and 6 questions about each proposed task, with responses scored on a 5-point Likert scale that rated the extent to which the test content represented the domain evaluated. An iterative process of simulator development was performed using pilot testing by surgeons, engineers, and education experts until a *good enough* prototype was achieved.

The hypotheses were as follows:

- It is possible to develop a portable, low-cost, gesture-mediated simulator using the LMC for training and learning basic psychomotor skills in MIS.
- The 3D virtual environment and the proposed tasks showed fidelity to the criterion.
- It is possible to demonstrate sources of evidence for the content validity of the test items.

The first step of the validation process was to define the construct and proposed interpretation. In this study, the general construct is psychomotor skills in surgery, specifically basic psychomotor skills in MIS. The assumptions and proposed interpretations are that the 3D virtual environment is faithful to the criterion and the tasks adapted from the MIST-VR represent the construct that is intended to be measured. The instrument under investigation is a contactless, gesture-mediated simulator that uses the LMC (construct context). To determine the current use of gesture-mediated interfaces in surgery, especially in the field of surgical simulation, a systematic literature review was conducted [14]. Finally, as content-related validity evidence was collected, the goal was to identify whether there were any areas of construct underrepresentation or construct irrelevance.

Phase 1: Initial Development of the Gesture-Mediated Simulator for Minimally Invasive Surgery-Virtual Reality

To develop a new type of web-based 3D VR simulator mediated by a gesture interface device (LMC) for learning basic

XSL•FO
RenderX

psychomotor skills in MIS, a group consisting of a pediatric surgeon, systems engineer, industrial designer, and specialists in education was formed. The following technical elements were assembled: an electronic device (LMC), a computer program for the development of the 3D environment, a computer, hardware devices with no electronic components, and a database administrator.

Electronic Device: Leap Motion Controller

In May 2012, a sensor was launched based on the principle of infrared optical tracking, which detects the positions of fine objects such as fingertips or pen tips in a Cartesian plane. Its interaction zone is an inverted cone of approximately 0.23 m^3, and it has a motion detection range that fluctuates between 20 mm and 600 mm [23,24]. This sensor measures 76 mm × 30 mm × 13 mm and weighs 45 g. It has 3 infrared emitters and 2 infrared cameras that capture the movements generated within the interaction zone [25,26]. The manufacturer reports an accuracy of 0.01 mm for fingertip detection, although one independent study showed an accuracy of 0.7 mm [27]. Although the LMC is designed mainly to detect the motion of the hands, it can track objects such as pencils and laparoscopic surgical forceps [28,29].

The LMC has been used as a tool to manipulate medical images in the fields of interventional radiology and image-guided surgery or when there is a risk of contamination through contact (eg, autopsy rooms). It has also been used for touchless control of operating lights and tables and simulation in MIS and robotic surgery using physical or VR simulators [14,28].

Unity3D and Development of the Web-Based Virtual Environment Based on Minimally Invasive Surgery Training–Virtual Reality Tasks

The 3D virtual environment with MIS tasks was created using a tool for developing games, Unity3D, which allows apps to be developed that are independent of the operating system or device [30].

The basis for the development of this environment was the MIST-VR, presented in 1997. This device is a low-cost, nonprocedural simulator that provides a large variety of metric data for analysis [31] and generates simple and abstract images that allow the training and learning of basic psychomotor maneuvers that cross many surgical disciplines [9,32,33]. The simple images allow novice learners to progress rapidly in the early phase of the basic psychomotor skills learning curve [34-36], although detailed performance analysis and feedback allow them to train alone, with no need for specialized instructors [37].

The basic psychomotor skills in MIS that can be learned using the MIST-VR are navigation-coordination, touching, grasping, stretching-traction, translocation, and electrocautery [38].

Computer

The computer displays the 3D virtual environment, records the metrics, stores them on a database, and provides feedback using graphs that show the score obtained after each exercise. The virtual environment developed runs on both PC and iOS operating systems.

Hardware Devices

The mechanical devices are represented by 2 MIS forceps that do not need to be functional, 2 support devices for the forceps with an entry trocar simulator, 1 support device for the LMC, and 1 pad for mounting the support devices.

During the development of the virtual environment, the types of specificity recommended by Bowman et al [39] were applied:

- *Application:* To design a web-based 3D virtual environment for basic psychomotor skills training in MIS
- *Domain:* Basic psychomotor skills in MIS
- *Tasks:* 6 tasks described in the MIST-VR were adopted
- *Device:* LMC, LEAP
- *Users:* Surgeons in training for learning basic psychomotor skills in MIS

Phase 2: Evaluation of Fidelity to Criterion, Content-Related Validity Evidence

Subjects

The study was performed over a period of 3 months at different locations: XXXIV Brazilian Congress of Paediatric Surgery (Campo Grande, Brazil); Hospital Vall d'Hebron (Barcelona, Spain); and Hospital Infantil de la Cruz Roja (Manizales, Colombia). A total of 22 experienced surgeons (performed more than 100 MIS procedures) and 8 pediatric and general surgery residents (referent group, performed less than 100 MIS procedures) assisted in an informative session on the characteristics of the project, watched a demonstration video of the different tasks supported by the simulator, and had 2 opportunities to perform each of the tasks on the simulator. The performance metrics were not taken into account during this study, as the emphasis was placed on the assessment of the tool by those surveyed.

Content-Related Validity Evidence for a Previously Validated Tool

The first source of content validity for the SIMISGEST-VR sought to identify the main sources of validity evidence for the MIST-VR, as well as the studies that have demonstrated such validity.

Questionnaire

First, a demographic survey was administered that included questions on the level of training as a surgeon and level of experience in MIS, as well as experience with video games. The different factors in the evaluation of fidelity to the criterion and content validity study were assessed using a Likert scale, where 1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=strongly agree [40].

The questionnaire to assess fidelity to the criterion evaluated 10 aspects, while the content validity rated the training capacity and the tasks. In terms of the training capacity, 6 aspects were evaluated, and each of the 6 tasks (Table 1) was assessed based on whether or not it represented a specific surgical maneuver (Multimedia Appendix 1).

XSL•FO
RenderX

Simulator, Hardware, and Software

This study used SIMISGEST-VR with 6 tasks and their respective metrics and feedback. The hardware and software components of the simulator are described in phase 1: Development of SIMISGEST-VR of this paper.

Statistics

Normality was tested using the Shapiro-Wilk test. The distribution of the variables was not normal. The Likert scale median and interquartile range differences between the levels of education and experience were compared using the Kruskal-Wallis test. A statistically significant level <0.05 was established. The analysis was performed using Stata version 15.0 (StataCorp).

Results

Phase 1: Development of Gesture-Mediated Simulator for Minimally Invasive Surgery–Virtual Reality

The Virtual Environment

The virtual environment consists of the following modules:

- *Registration:* Collects the user's demographic information and stores it in the database
- *Tutorial:* Presents demonstration videos of the exercises
- *Test (tasks):* Supports 6 tasks, each of which corresponds to a surgical equivalent (Table 1) [9,41]
- *Performance graphs:* When an exercise is completed, the platform displays the results of the metrics in terms of the time taken to perform the exercise, precision of movement, and presence or absence of errors (immediate feedback; Figure 1). In this module, the student can look up the score obtained after each exercise and check whether or not their performance has improved (terminal feedback; Figure 2).

Except for Task 3, all tasks have the option of configuring the dominant hand during the exercise. Task 3 requires the simultaneous use of both hands and therefore both play a dominant function.

The web-based virtual environment runs on PC and iOS platforms.

These exercises are based on the instructional strategy known as *drill and practice*, which promotes the acquisition of knowledge or skill through repetitive practice [42].

Table 1. Description of the tasks and their surgical equivalents.

Task[a]	Description	Surgical equivalent
Task 1: Grip and placement	Take the sphere with one hand and move it to a new location within the workspace	Gripping and retraction of a tissue to a given position, placement of clips and hemostasis, and use of extractor bags
Task 2: Transfer and placement of an object	Take the sphere, transfer it to another instrument, and place it inside a hollow cylinder	Transfer of a needle between a clamp and a needle holder
Task 3: Cross	Instruments travel along a surface in a 3D cylinder	Small intestine exploration
Task 4: Removal and reinsertion of instruments	Removal of the instruments from the operative site and reinsertion	One instrument stabilizes one organ while the other is removed from the field and reintroduced
Task 5: Diathermy	Cauterize a series of targets located in a fixed sphere	Cauterize a bleeding blood vessel
Task 6: Target manipulation and diathermy	Take the sphere with the instrument and place it inside a virtual space represented by a cube. Cauterize a series of targets with the other hand	Present and set a target to cauterize

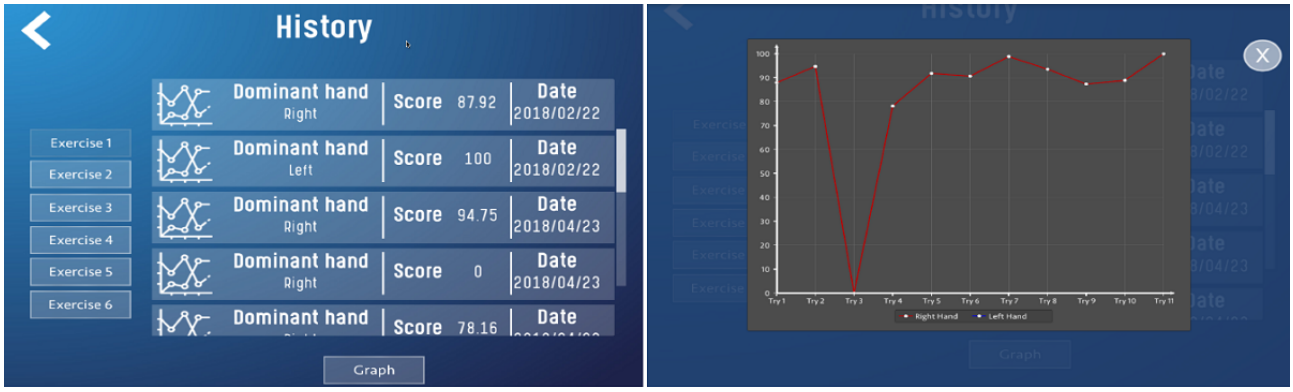
[a]Adapted from [9].

XSL•FO
RenderX

Figure 1. Immediate feedback.



Figure 2. Performance history and terminal feedback curve.



Metrics

The metrics were established using 5 parameters:

1. *Time:* Time taken from starting the task until completion. The time is measured only for each individual task.
2. *Efficiency of movement for the right and left hand:* This is the time during which the tip of the forceps is outside the ideal path, that is, the difference between the actual and the ideal path length [10,43].
3. *Economy of diathermy:* If contact occurs with the target for more than 2 sec during the diathermy, it is considered excess burn time and is penalized as a specific error.
4. *Error:* The following were defined as errors [43,44]: contact of the target with a part of the forceps other than the tip: all exercises; contact of the instrument with the limits of the virtual working space: all exercises; number of contacts of the instrument with the target sphere: exercises 1 and 2; number of contacts of the instrument or the sphere with the container margins: exercises 1 and 2; number of times that the instruments made undue contact between them: exercises 3 and 4; number of times that the instruments exceeded the number of contacts permitted with the oval: exercise 4; time during which the tip of the instrument remained outside the ideal path for the exercise: exercises 1, 2, 4, and 6; diathermy of the sphere outside the stated objectives: exercises 5 and 6; excess burn time: exercises 5 and 6.
5. *Final score:* The final score is obtained from the sum of the results of the efficiency of movement for each hand plus the economy of diathermy and from the errors generated in each exercise. Each exercise generates different types of errors, and each error is assigned a value between 1 and 5, where 1 is the least important and 5 is the most important. For example, diathermy outside the assigned targets means an error with an assigned score of 5. The number of errors committed in each of the exercises is then counted, and this figure is multiplied by the value assigned to each error. Finally, all the figures obtained in each of the errors are added, and the final score results from subtracting the sum obtained from 100. This is expressed in the following formula: $100–\sum(\text{error}\times\text{value})$. Thus, a higher score indicates better performance.

XSL•FO
RenderX

Feedback

The haptic sensation and the concurrent feedback are simulated using sound signals, color changes in the objects, and movement of the object when an undue collision occurs between the different components of the environment or when an error occurs during the exercise. At the end of each task, the system provides information on the presence or absence of errors, the efficacy and efficiency, and the time required (immediate feedback). At the end of each training session, the system provides a series of graphs and tables that show the performance over time; this is the terminal feedback (Figures 1 and 2).

SQLite Database Engine

The data generated by the program were initially stored on an independent Structured Query Language database engine. However, during the development, this database was integrated into the virtual environment, which facilitated the acquisition of the users' demographic data, registration of all the data provided by the metrics, and generation of reports of the users' demographic and performance data. This information is stored on the computer on which the tests are performed.

Hardware

Two laparoscopic forceps were used. These MIS forceps did not need to be functional.

In the initial phase of development, the researchers used a prototype that did not have support devices (Figure 3), but it soon became evident that the fulcrum effect was not being reproduced. For this reason, they designed support devices for the forceps, which simulate the entry portal to the abdomen (Figure 4). These devices, while generating friction when inserting and removing the forceps, limit the moment of the arms, as occurs in real surgical procedures. During the process of designing these devices, principles were prioritized, such as noninterference with the forceps reading by the LMC, portability, and low cost. A pad for mounting the support devices and the LMC was also designed, which had a 45-degree tilt on a horizontal plane.

The final artifact with all its components assembled is shown in Figure 5. It shows the fixing pad (1) for the LMC and the mounting support devices (3) for the MIS laparoscopic forceps (2), which allow simulation of the fulcrum effect; the LMC (4), responsible for detecting the movements of the instruments; and the computer, which using the software programs administers the virtual environment and the metrics and provides feedback and the final performance score on the screen (5), where the 3D virtual environment is displayed.

In Figure 4, the LMC has a 45-degree tilt toward the screen with respect to the horizontal plane. This arrangement was the result of a process of trial and error, which showed that setting the LMC at this angle with respect to the horizontal plane ostensibly improved the detection of the forceps. Another significant change during the design was that the original black color of the shaft of the forceps did not facilitate reading by the LMC [45]; therefore, they were painted white in the final prototype (Figure 4).

Figures 6-8 show various stages in the development of the prototype for the 3D virtual environment. As in the development of the hardware elements, the 3D virtual environment design process was iterative, so that each new version of the 3D virtual environment became increasingly closer to the version considered *good enough* in terms of the design and function.

Figure 6 shows the initial attempt at the interaction between the forceps and the basic 3D virtual environment. At this stage of the design, the researchers achieved *capture* of the virtual objects by the tip of the instruments and their transfer to a virtual container (Figure 6). The second stage of development accomplished the development of the 5 tasks in a 3D virtual environment characterized by rectangular geometric shapes (Figure 7). Although the researchers did have concurrent feedback based on sounds, color changes, and a sensation of collision, at that time, the metrics had not been developed. Figure 8 shows the final *good enough* result of the 3D virtual environment. On the basis of the feedback provided by the expert surgeons, the environment was redesigned without rectangular geometric shapes, although with abstract circular shapes that were closer to the view of the body cavities during the MIS procedure.

The changes shown in Table 2 reflect the steps in the process described by Manson [16,20], where during the development of the artifact, through iterative processes of deduction and circumscription, errors were recognized in the design or function of the prototype that required further versions to be developed until the study achieved one that was considered *good enough* [16,17] and functional.

XSL•FO
RenderX

Figure 3. Initial version of the prototype without support devices for the forceps.

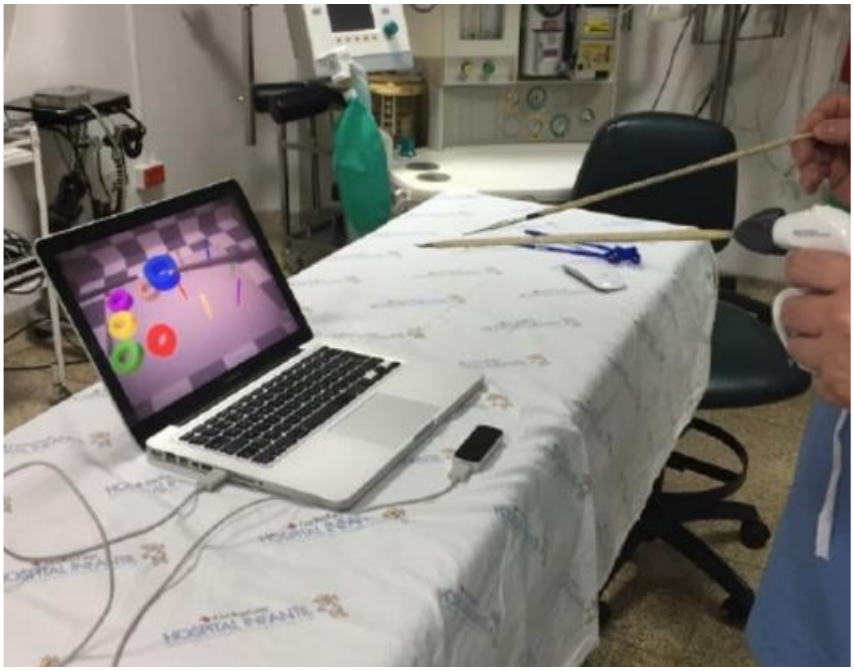
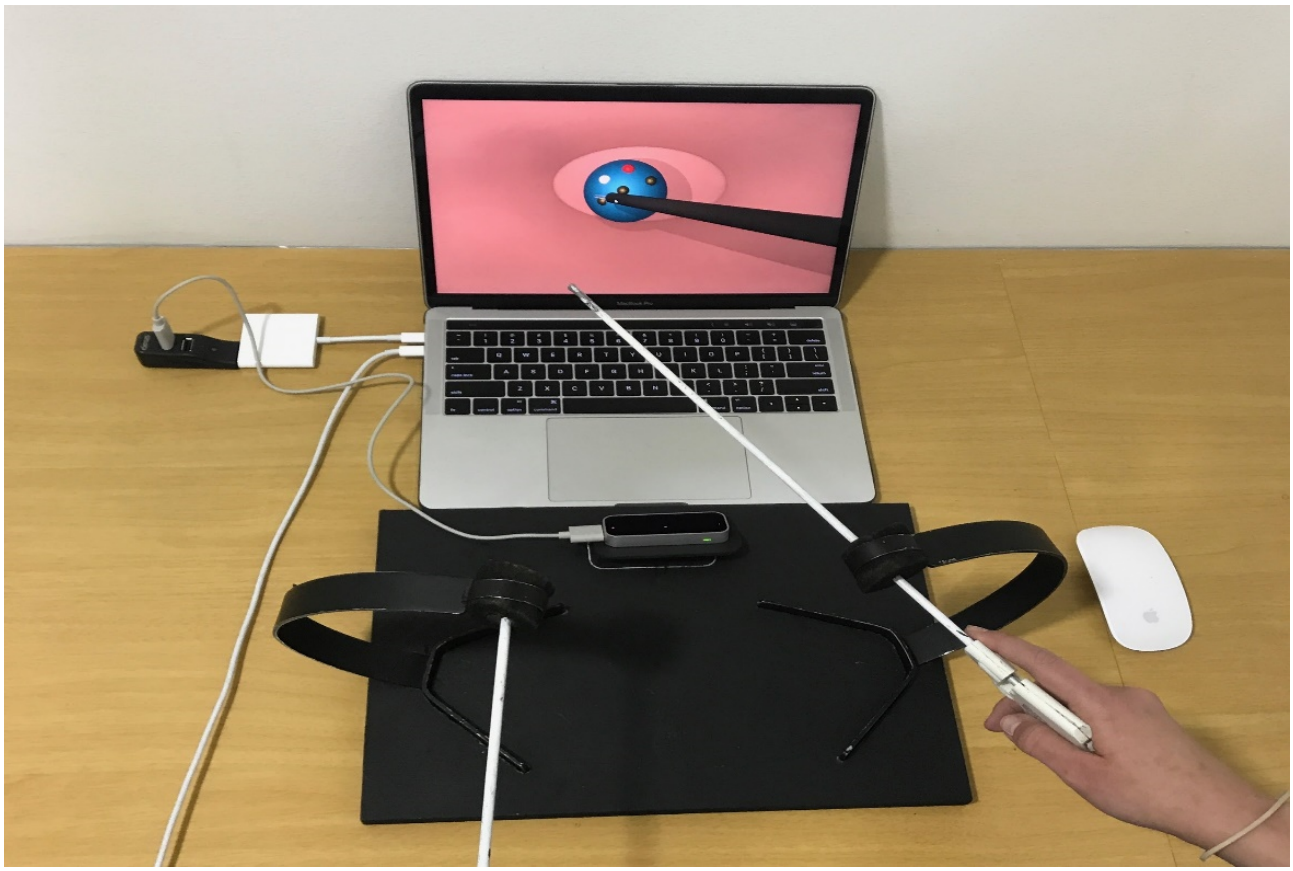


Figure 4. The final version of the simulator once the nonelectronic hardware devices had been added: the pad and support devices for the forceps and the Leap Motion Controller.



XSL•FO
RenderX

Figure 5. Diagram of the artefact.

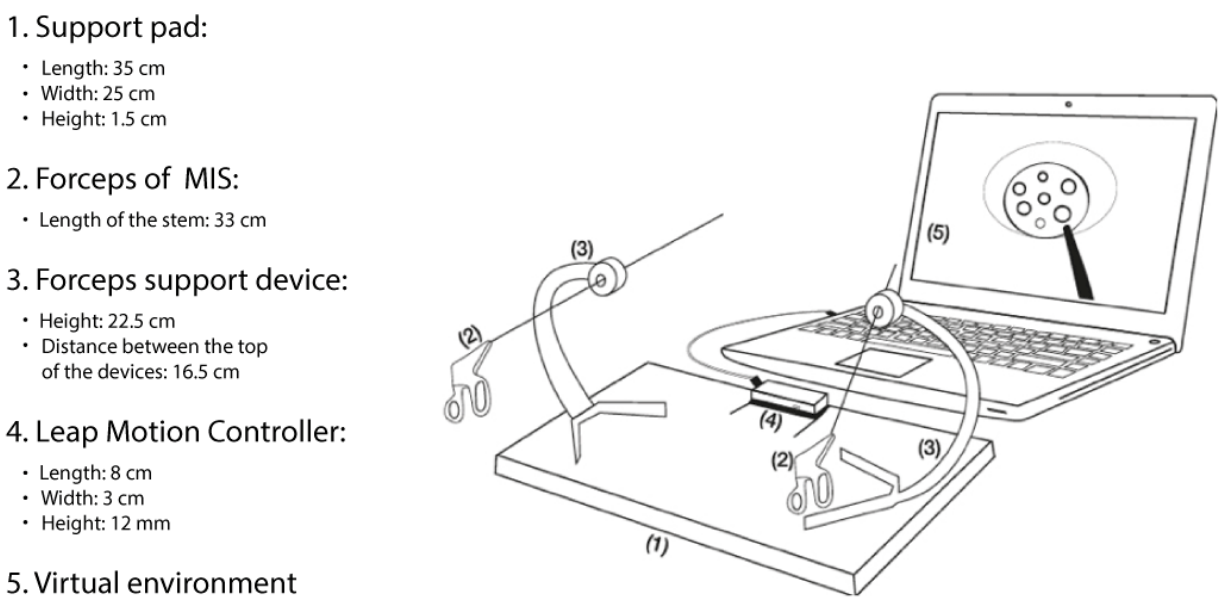
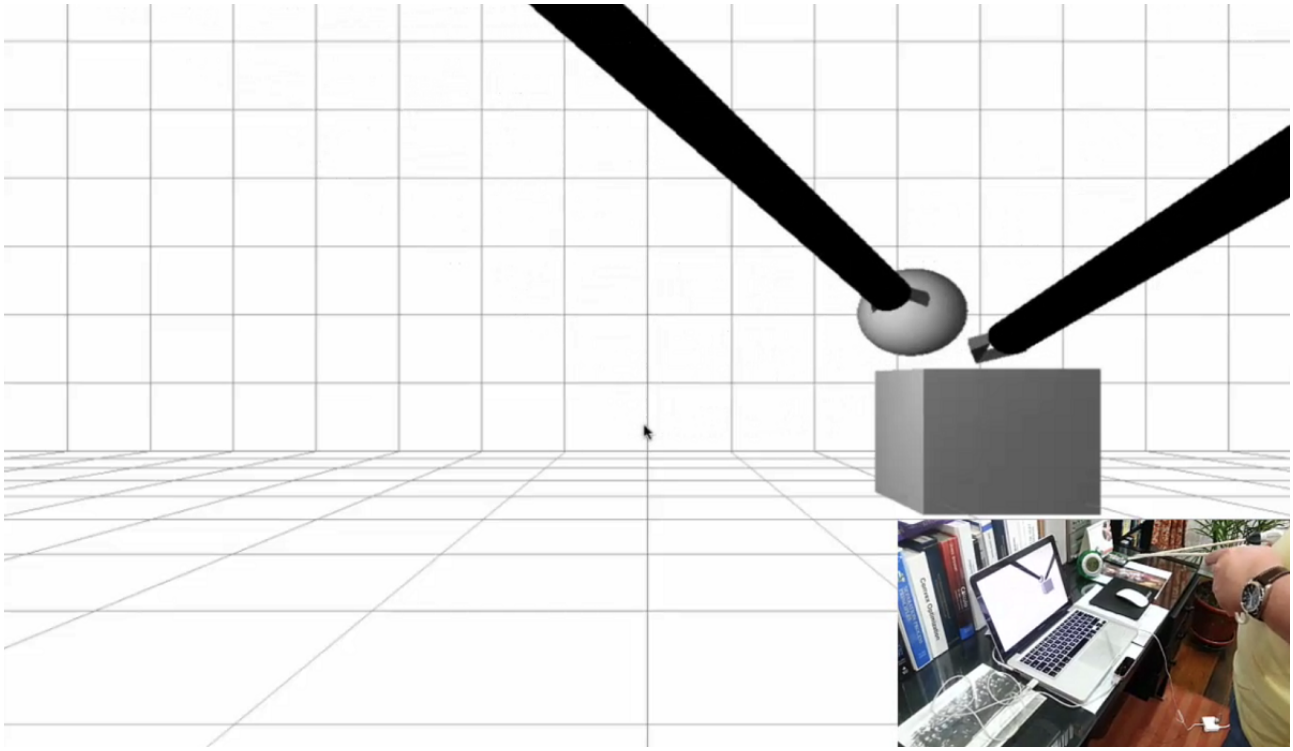


Figure 6. Initial attempts at interaction between minimally invasive surgery forceps and Leap Motion Controller within a basic 3D virtual environment.



XSL•FO
RenderX

Figure 7. The first functional version of the virtual environment before the feedback given by surgeons with expertise in minimally invasive surgery.

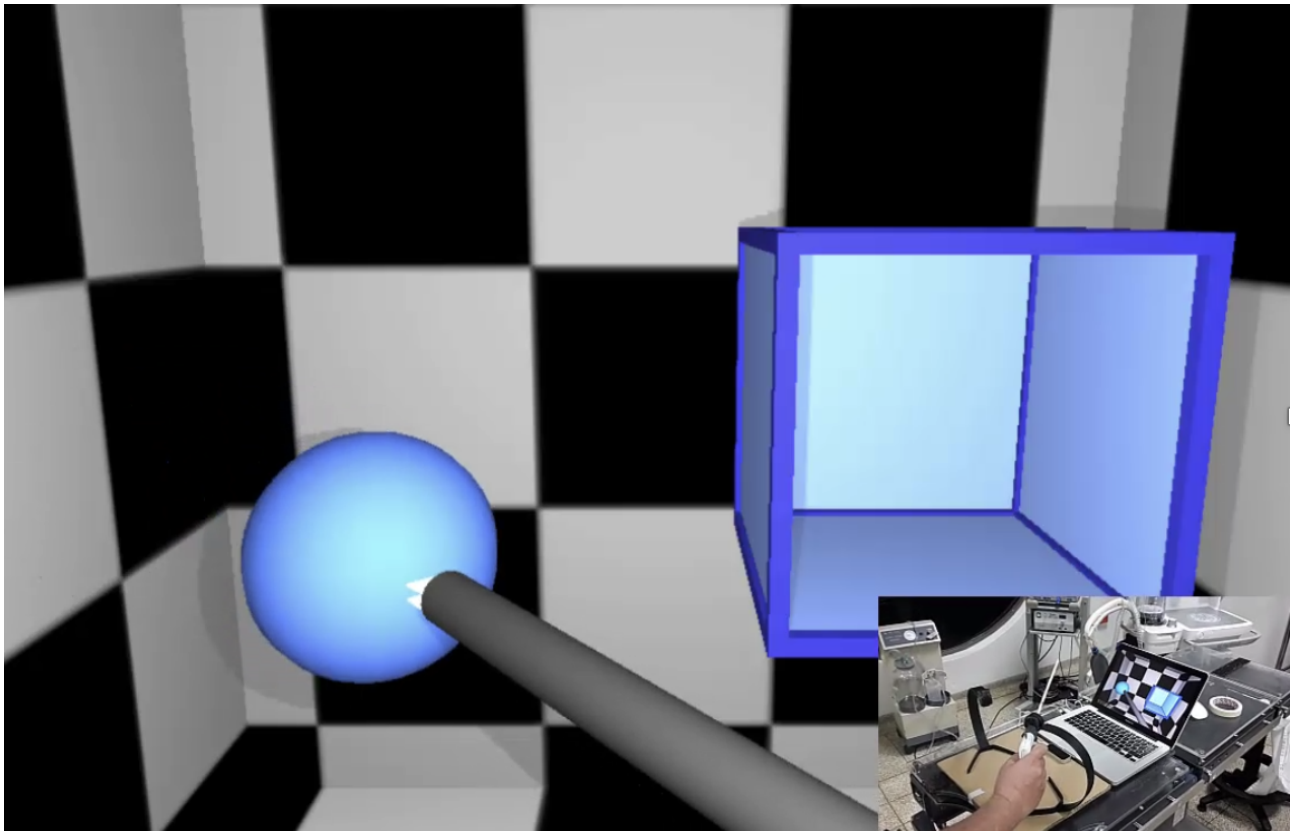
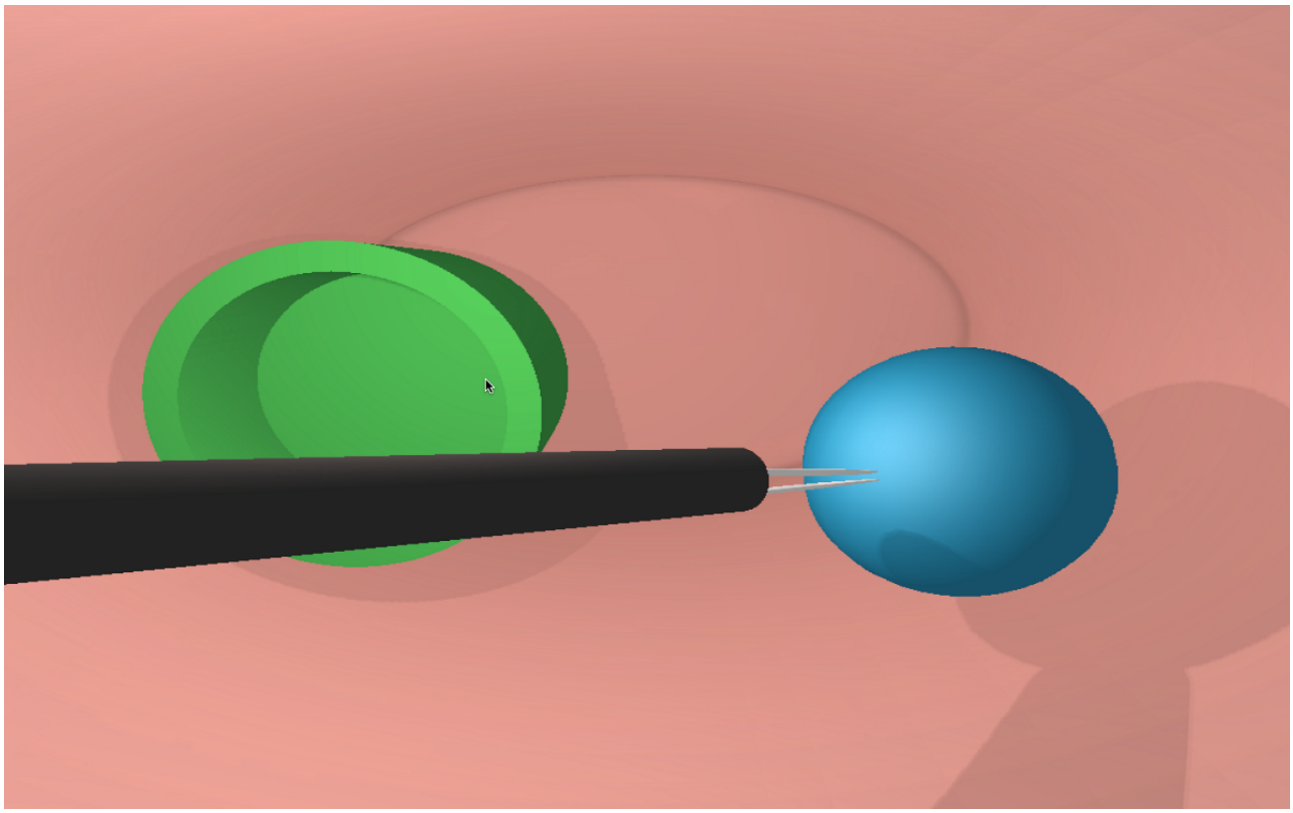


Figure 8. Good enough prototype of the web-based 3D virtual environment: Task 1.



XSL•FO
RenderX

Table 2. Process of obtaining the good enough prototype.

Element	Initial prototype	Problem	Functional prototype	Output
MIS[a] forceps	The shaft of the forceps is black	Difficulties in the detection of the forceps by the LMC[b]	The shaft of the forceps is white	Notable improvement in detection of the forceps by the LMC
Support devices	No support devices	Fulcrum effect not reproduced	Design of support devices	Reproduction of fulcrum effect
Mounting pad	No mounting pad	The hardware pieces (LMC and support devices) are independent, and there is no standard arrangement	Standardized integration of the pieces in the mounting pad	Physical stability of the model
Position of the LMC	Completely horizontal, 90 degree with regard to the screen	Difficulties in the detection of the forceps by the LMC	A forward 45-degree angle was applied with regard to the screen	Interference between the forceps when detected by the LMC was eliminated
First prototype of the 3D virtual environment (Figure 6)	Tests on the interaction between the forceps and the objects in the virtual environment	Difficulty for interaction between the forceps and the objects in the environment	Trial and error tests on interaction by modifying LMC and instrumental variables	Complete interaction achieved
Second prototype of the 3D virtual environment (Figure 7)	Functional environment in the 6 tasks	Quadrangular shapes in the environment	Circular shapes in the *good enough* environment (Figure 8)	An abstract 3D virtual environment with circular shapes
Good enough environment (Figure 8)	SQL[c] database engine not integrated into the simulation software	A software program should be installed in addition to the simulation program	Redesign of the model and data capture and storage	Feedback and metrics complete and integrated into the SQLite Integration of the SQL database engine into the simulation software

[a]MIS: minimally invasive surgery.

[b]LMC: Leap Motion Controller.

[c]SQL: Structured Query Language.

Phase 2: Evaluation of Fidelity to the Criterion and Subjective Validation of SIMISGEST-VR (Content Validity)

The next step in the process was the evaluation of fidelity to the criterion and the process of subjective content validity. The results are described below.

Demographics

A total of 30 people with an average age of 42 years (SD 2.2) participated in the study; 53% (n=16/30) were men. Those surveyed came from Colombia (n=14), Spain (n=8), Argentina (n=3), Brazil (n=2), Uruguay (n=2), and France (n=1).

Table 3 summarizes the participants' profiles according to the level of training and experience. The residents belonged to training programs in general and pediatric surgery; one of the participants was a biomedical engineer with extensive experience in the design of devices and simulators in MIS. The vast majority (n=28/30, 93%) of participants were right-handed, 1 was left-handed (n=9/30, 3%), and the other, ambidextrous (n=9/30, 3%).

In terms of the use of video games, most (n=22/30, 73%) of those surveyed had no experience with these app; 62% (n=5/8) of those who used video games were women. Of those with experience in video games (n=8), only 1 played them weekly, while the rest played them once a month (n=3) or occasionally (n=4). The mean age of those with no experience in video games was 44 years (SD 2.7), compared with 37 years (SD 3.5) for those with experience (*P*=.16).

Only 33% (n=10/30) of the participants had experience with VR devices, and only one-third used them occasionally.

Most of the surveyed participants had previous experience with simulators. In terms of the level of operating experience, 54% (n=14/26) of the respondents with experience with simulators had an intermediate or advanced operating level, followed by those with a basic operating level (n=10/26, 38%). Among participants who had experience with simulators [26], 62% (n=16/26) had used physical simulators, 23% (n=6/26) had used hybrid simulators, and only 15% (n=4/26) had used VR simulators. The average age of those who had no experience with simulators was 40 years (SD 6.7), compared with 42 years (SD 2.4) for those with previous experience (*P*=.83).

The demographic profile questionnaire can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Demographic profile according to the level of experience and training (N=30).

Demographic variable	Level of experience				Level of training		
	Basic manipulation (n=3)[a]	Basic operating level (n=11)[b]	Intermediate operating level (n=8)[c]	Advanced operating level (n=8)[d]	Practicing surgeon (n=21)	Resident (n=8)	Other (n=1)[e]
Age (years), mean (SD)	26 (0.6)	40 (4.3)	43 (3.3)	49 (2.9)	47 (2.2)	27 (0.6)	49 (—[f])
Sex, n							
Male	0	3	5	8	15	0	1
Female	3	8	3	0	6	8	0
Do you have regular experience with video games?, n							
Yes	1	1	4	2	6	1	1
No	2	10	4	6	16	6	0
Do you have previous experience with MIS[g] simulators?, n							
Yes	2	10	6	8	19	7	0
No	1	1	2	0	3	0	1
What type of simulator?, n							
Physical	2	6	3	5	12	3	1
Hybrid and augmented reality	0	2	2	2	4	2	0
Virtual reality	0	2	1	1	2	2	0
No experience	1	1	2	0	3	1	0

[a]Basic manipulation of the camera and/or retraction with forceps.

[b]Basic operating level (cholecystectomy and appendectomy).

[c]Intermediate operating level (fundoplication).

[d]Advanced operating level.

[e]Other: an engineer highly experienced in the design of instruments and devices for minimally invasive surgery simulation.

[f]Not available because there was only one observation.

[g]MIS: minimally invasive surgery.

Evaluation of Fidelity to Criterion

Tables 4 and 5 show that there were no significant differences in the different ratings when the level of training (Table 4) or experience (Table 5) was considered.

In terms of the fidelity to the criterion, none of the respondents strongly disagreed with any of the items asked. The rating of *disagree* was given by one participant to the question about relevance, by another to the assessment of how the movements of the physical instruments were represented in the virtual environment, and 3 assigned this score when rating the fulcrum effect.

In terms of ease of use, 73% (n=22/30) and 27% (n=8/30) assigned a rating of 5 and 4, respectively. The same results were obtained when the navigation menu was assessed. With regard to the relevance of the tool as a simulator, 73% (n=22/30) assigned a score of 5 and 20% (n=6/30) assigned a score of 4.

When assessing the capacity of the physical devices to simulate the fulcrum effect, 73% (22/30) assigned a score between 4 and 5, 17% (n=5/30) assigned a score of 3, and 10% (n=3/30) assigned a score of 2. For this last rating, in terms of the level of training, 2 were practicing surgeons and 1 was a resident, whereas in terms of the level of experience, one corresponded to basic manipulation, one to intermediate operating level, and another to advanced level.

In terms of how the movements of the forceps were represented in the virtual environment, 73% (n=22/30) rated this as 4 or 5, 23% (n=7/30) assigned a score of 3, and only one of the participants (n=9/30, 3%) assigned a score of 2 (level of training=practicing surgeon and level of experience=intermediate).

When assessing how appropriately the tool simulates the movements of MIS, 83% (n=25/30) rated the question as 4 or 5. All respondents (n=30/30, 100%) rated the design as attractive, with scores of 4 or 5. Almost all surveyed respondents (n=29/30, 97%) assigned ratings of 4 or 5 to the innovation factor, the capacity to provide feedback, and to the question of whether the latter was adequate.

The fidelity to the criterion study questions can be found in Multimedia Appendix 1. The fidelity to the criterion study result tables can be found in Multimedia Appendix 2.

XSL•FO
RenderX

Table 4. Fidelity to the criterion and content validity according to the level of training.

Variable	Resident (n=8)		Practicing surgeon (n=21)		Other[a] (n=1)		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Fidelity to the criterion[b]							
Ease of use	5	4.5-5	5	4-5	5	5-5	.88
Navigation menu	5	5-5	5	4-5	5	5-5	.62
Relevance as a learning tool	5	4-5	5	5-5	5	5-5	.73
Fulcrum effect	3.5	3-4	5	4-5	4	4-4	.13
Representation of the physical forceps in the virtual environment	4	3-5	4	4-5	5	5-5	.56
Simulation of the movements in MIS[c]	4	4-4	4	4 - 5	5	5-5	.18
Innovation	5	4.5-5	5	5-5	5	5-5	.90
Graphic design	4.5	4-5	5	4-5	5	5-5	.69
Feedback	5	4-5	5	5-5	5	5-5	.79
Relevance of the feedback	4	4-5	5	4-5	5	5-5	.43
Content validity[d]							
Hand-eye coordination	4.5	4-5	5	4-5	5	5-5	.66
Depth perception	4	3.5-5	5	4-5	5	5-5	.41
Basic psychomotor skills learning	4.5	4-5	5	5-5	5	5-5	.42
Basic steps of MIS	4	4-5	5	4-5	5	5-5	.64
Metrics	4	3-5	4	4-5	5	5-5	.43
Ubiquitous learning	4	4-5	5	4-5	5	5-5	.31
Tasks[e]							
Task 1	3.5	3-4	4	4-5	5	5-5	.19
Task 2	4	4-4.5	4	3-5	5	5-5	.41
Task 3	4	3.5-4	4	3-5	5	5-5	.40
Task 4	4	3-5	5	4-5	2	2-2	.21
Task 5	4.5	4-5	5	4-5	5	5-5	.65
Task 6	4	4-4	5	4-5	5	5-5	.02

[a]Other: An engineer highly experienced in the design of instruments and devices for minimally invasive surgery simulation.

[b]For fidelity to the criterion questions, see Multimedia Appendix 1.

[c]MIS: minimally invasive surgery.

[d]For content validity questions, see Multimedia Appendix 1.

[e]For task descriptions, see Table 1.

XSL•FO
RenderX

Table 5. Fidelity to the criterion and content validity according to the level of experience.

Variable	Basic manipulation (n=3)		Basic operating level (n=11)		Intermediate operating level (n=8)		Advanced operating level (n=8)		*P* value
	Median	IQR	Median	IQR	Median	IQR	Median	IQR	
Fidelity to the criterion[a]									
Ease of use	5	4-5	5	4-5	5	4.5-5	4.5	4-5	.84
Navigation menu	5	4-5	5	5-5	5	4.5-5	4.5	4-5	.51
Relevance as a learning tool	5	2-5	5	4-5	5	5-5	5	4-5	.83
Fulcrum effect	4	2-5	4	3-5	5	4-5	4	4-5	.66
Representation of the physical forceps in the virtual environment	5	3-5	4	4-5	4	3.5-5	4.5	3-5	.96
Simulation of the movements in MIS[b]	4	4-4	4	4-5	4.5	3.5-5	4	3.5-4.5	.70
Innovation	5	3-5	5	5-5	5	4.5-5	5	4.5-5	.95
Graphic design	4	4-5	5	4-5	5	4.5-5	4	4-5	.41
Feedback	5	4-5	5	4-5	5	5-5	4.5	4-5	.42
Relevance of the feedback	4	4-5	5	4-5	5	4.5-5	4.5	4-5	.66
Content validity[c]									
Hand-eye coordination	4	4-5	5	4-5	5	5-5	4.5	4-5	.77
Depth perception	5	4-5	5	4-5	5	4-5	4.5	4-5	.95
Basic psychomotor skills learning	4	3-5	5	4-5	5	5-5	5	4-5	.45
Basic steps of MIS	4	4-5	5	4-5	5	4.5-5	4	4-4.5	.33
Metrics	4	3-5	4	3-5	4.5	4-5	4.5	4-5	.75
Ubiquitous learning	4	4-5	5	4-5	5	5-5	4.5	4-5	.46
Tasks[d]									
Task 1	3	3-5	4	3-5	4	3.5-5	4	3.5-4.5	.88
Task 2	4	4-5	4	4-5	4	3.5-5	3.5	2-5	.76
Task 3	4	2-5	4	3-4	4.5	3.5-5	4	3.5 – 5	.76
Task 4	2	2-5	4	4-5	5	5-5	4	3.5-5	.18
Task 5	4	4-5	5	4-5	5	4.5-5	4.5	4-5	.70
Task 6	4	4-4	4	4-5	5	5-5	4.5	4-5	.12

[a]For fidelity to the criterion questions, see Multimedia Appendix 1.

[b]MIS: minimally invasive surgery.

[c]For content validity questions, see Multimedia Appendix 1.

[d]For task descriptions, see Table 1.

Content Validity

Table 6 summarizes the sources of validity evidence for the MIST-VR and the studies that have demonstrated such validity.

With regard to content validity, none of the items evaluated for the training capacity were rated as 1, although, in the case of hand-eye coordination by a practicing surgeon with an advanced operating level and the depth perception by a practicing surgeon with an intermediate operating level, the hand-eye coordination and depth perception were rated as 2. Almost all of those surveyed (n=28/30, 93%) rated the hand-eye coordination as 4 or 5, while 87% (n=26/30) gave this score for depth perception.

The highest-rated item was the one that considered that the prototype could be a solution for ubiquitous learning in MIS: 100% (n=30/30) of those surveyed rated it as 4 or 5. With regard to the evaluation of the metrics, 17% (n=5/30) of those surveyed rated them as 3, while the remaining participants (n=25/30) rated them as 4 or 5.

Almost all respondents (n=29/30, 97%) considered that the SIMISGEST-VR enables learning of basic psychomotor skills in MIS, with ratings of 4 and 5; whereas, 93% (n=28/30) agreed that the tasks reflect the basic steps of a minimally invasive procedure, with ratings of 4 and 5.

XSL•FO
RenderX

An analysis of the evaluation of the tasks, in general, showed that the following were rated between 4 and 5: Task 1 received this rating from 70% (n=21/30) of those interviewed; Task 2 from 77% (n=23/30); Task 3 from 73% (n=22/30); Task 4 from 77% (n=23/30); and Task 5 and Task 6 from 90% (n=27/30) of the participants.

For Task 6 (Table 4), a lower score was assigned by individuals with lower levels of training (P=.02).

The content validity study questions can be found in Multimedia Appendix 1. The tables of results of the content validity study can be found in Multimedia Appendix 3.

Table 6. Sources of validity evidence for the minimally invasive surgery training–virtual reality.

Source of validity evidence	Studies
Content evidence	[9,10,43,46-51]
Internal structure	[41,43,48,52-62]
Relationship to other variables	[4,10,12,13,37,43,47,50-57,60,63-116]
Consequences	[4,11,13,43,47,49,50,54,56,57,59,69,71,73,78,79,92,104,107,108,114,117-123]

Discussion

Principal Findings

Simulation as a tool for learning psychomotor skills in MIS has become a new model for education in surgery. The use of human or animal cadavers is becoming increasingly controversial for learning surgical maneuvers [124,125], resulting in an immense growth of simulation using virtual environments as a tool for learning psychomotor skills in MIS and for the simulation of full surgical procedures [5,126].

Simulators for psychomotor skills learning in MIS are classified into mechanical, hybrid/augmented reality, or VR [6,7]. Devices for gesture-based human-computer interaction are a new way of interacting with virtual environments. This study's simulator presents a new form of gesture-based simulation that is portable, low-cost, and enables ubiquitous learning and preoperative warm-up [14,127,128].

Development of Gesture-Mediated Simulator for Minimally Invasive Surgery—Virtual Reality

The development of SIMISGEST-VR was based on DBR principles. It was a *pragmatic* process because the researchers tried to resolve the problems of portability and the high cost of simulators for learning psychomotor skills in MIS. It was *grounded* in both theory and the real-world context, as we designed a functional simulator based on theories on simulation-based surgical skills training. It was *interactive*, in that during the simulator design stage, a *good enough* prototype was obtained through the participation of an interdisciplinary team (pediatric surgery, systems engineering, graphic design, and experts in education and psychology), as well as the comments and feedback provided by experts in MIS during the subjective validation study. Finally, the process was *iterative*, in that a process of analysis, design, evaluation, and redesign was applied (Table 1) until a *good enough* protocol was obtained that could be subjected to validation studies [129].

To develop this study's 3D virtual environment, the researchers adopted the principle of low fidelity, given that the model is envisaged for basic psychomotor skills learning. The term fidelity refers to the extent to which a simulation imitates reality (in the case of surgical simulation, the anatomy) and is considered a critical variable in the design of simulators. However, this statement is not necessarily completely true, as for novice learners, low-fidelity models that reproduce the essential constructs of a procedure allow a faster and more cost-effective learning curve to be achieved [35,130]. Thus, in the field of simulation in aviation, simple images reduce the learner's confusion when learning basic skills [131], while experts benefit from higher fidelity simulations [33,132].

The tasks were adapted from the MIST-VR, which is the only laparoscopic VR trainer that can act as a standard because it is the sole surgical VR system that has been reasonably validated [4,11,65,133]. MIST-VR has been shown to allow the learning of basic skills that can be transferred to the surgical environment at a more reasonable cost [4,11,52,73,134].

Metrics

Performance evaluation is a fundamental part of the learning process and is essential for certification. To obtain an objective evaluation of performance, the simulator should define metrics that must be valid, accurate, and relevant in terms of the procedure that is being taught. Evaluation using metrics and effective feedback are the most important elements of effective learning in a simulation environment [7]. Metrics allow an objective measure of motor performance to be obtained and enable the learning progress to be compared and tracked [10,43,44]. Accordingly, if the metrics lack sensitivity and validity, training on simulators will not be optimal and the learning will be affected [135]. In SIMISGEST-VR, the metrics were determined by time, the efficiency of movement, economy of diathermy, and error. This was an iterative process involving several pilot studies and modifications to the tasks and their metrics based on feedback provided by surgeons and education experts.

Feedback

Feedback is essential [136,137]. Training on a simulator should have 3 purposes: (1) to improve performance; (2) to make the performance consistent; and (3) to reduce the number of errors [57]. The metrics and feedback are essential for achieving these objectives. On the SIMISGEST-VR, the study adopted 3 types of feedback: (1) concurrent, which is provided while the task is being performed; (2) immediate, when the exercise is finished; and (3) terminal, which shows the final score when all the tasks have been completed [136,138-140].

XSL•FO
RenderX

Hardware

The design of the hardware components aimed to simulate the movements made by the surgeon during MIS. These movements are defined by the physical characteristics of the devices and, therefore, require the design of mechanical support devices that simulate the fulcrum effect (entry portals), add friction to the movements of the forceps, and limit arm movement during the performance of the tasks without interfering with the reading of the instrument movements by the LMC [141,142]. The portability and low cost were also taken into account.

Cost of Gesture-Mediated Simulator for Minimally Invasive Surgery-Virtual Reality

The VR or augmented VR simulators currently available in the market are not portable, and their cost ranges from US $2000 to US $100,000 (with annual maintenance costs of US $25,000) for a haptic VR simulator. The LMC costs approximately US $130, plus a further US $70 for the hardware elements, adding up to a total cost of approximately US $200 for the SIMISGEST-VR, software costs excluded.

Subjective Validation of Gesture-Mediated Simulator for Minimally Invasive Surgery–Virtual Reality

The second aim of this study was to evaluate fidelity to the criterion and a content validity study. Validity refers to the quality of the inferences, claims, or decisions taken from the scores given by an instrument, not the instrument itself. Validation for its part is a process through which the evidence that supports the quality, significance, and utility of the decisions and inferences that can be made from the scores provided by the instrument is drawn together and evaluated [143]. Validity is not an all-or-nothing statement, as it reflects a gradual appraisal that depends on the purpose of the measurement and the proper interpretation of the results. Validity is also not in itself a characteristic of the system, but the appropriate interpretation and use of the measurement results of the system. A single instrument may be used for many different purposes, and the resulting scores may be more valid for one purpose than for another [133].

Study of Fidelity to Criterion

Although it has been deemed that *face validity* should no longer be considered a type of validity or used as a term in validation studies [144,145], its assessment is extremely important during the design phase of any evaluation device [146,147]. Therefore, the use of an alternative term to denominate this type of evaluation has been suggested: *fidelity to the criterion* [148]. Despite such warnings, it is very striking to find that the term *face validity* is still being used in published literature on simulation in surgery [149,150].

Fidelity to the criterion evaluates to what point the test reflects the real-life situation, whether the simulator represents what it is supposed to represent (the realism of the simulator) or the extent to which a questionnaire or other measurement reflects the variable to be measured [125,151,152]. In the case of DBR, it is used in the initial phase of the construction of the test. The surveys that assess fidelity to the criterion feedback into the iterative design process, which allows the *good enough* prototype to be obtained [153]. Fidelity to the criterion is evaluated by experts and novices called referents [154,155].

In this study, the evaluation of fidelity to the criterion provided feedback on the initial design, and this was how the 3D virtual environment was redesigned until a *good enough* prototype was obtained. The quality of this evaluation is improved systematically when structured questionnaires and Likert scales are applied [154].

In all the items evaluated for fidelity to the criterion, most of those surveyed assigned scores of 4 or 5. There were no significant differences between the expert and referent groups (level of training) when rating fidelity to the criterion. The lowest scores were obtained for the item about the relevance (n=9/30, 3% of participants), the representation of the movements of the physical forceps in the virtual environment (n=9/30, 3%), and for the fulcrum effect (n=3/30, 10%).

Evidence Based on Test Content

The latest standards on validity and validation refer to sources of validity evidence, rather than distinct types of validity. Validity therefore refers to the degree to which the evidence and theory support the interpretations of test scores for the proposed uses of tests [156,157].

Evidence based on test content is an issue of representation and may be obtained from an analysis of the relationship between test content and the construct that is intended to be measured. In this study, the test content refers to the simulator's 6 specific tasks. Evidence can be obtained from logical or empirical analyses of how test content represents domain content and of the relevance of domain content to the proposed interpretation of test scores. Evidence may also come from experts' opinions on the relationship between the different test items and the construct when assessing whether the test contains the meaningful steps, skills, and materials used in the actual procedure [158] and determines whether the simulator can realistically teach what it is supposed to represent [159].

The question is, does the simulator realistically teach what it should teach? In other words, does the instrument represent all the ways in which it can be used to measure the content of a given construct? [160]. In summary, evidence based on test content judges the appropriateness of the simulator as a teaching modality or as a training tool within the domain that it seeks to measure [31,151,152].

This type of validation is highly recommended in the practice of DBR during the design phase of the *good enough* prototype. Content validity can be obtained from a literature review, an expert review, using content validity rates, and *Q sorting* [161].

The tasks within the surgical simulation should fulfill 3 criteria: objectivity, clarity, and completeness. To be objective, the definition of the task should refer to observable characteristics of the behavior; for it to be clear, the task should be unambiguous so that it can be read, understood, and reproduced equally by different observers; and finally, to meet the criterion of completeness, the definition of the task should delineate its start and end and make it clear when it was completed [162].

XSL•FO
RenderX

In this study, the 6 skill tasks were chosen for two main reasons: (1) these tasks are well-validated in many clinical studies [4,10,82,117] using the MIST-VR (Table 6); and (2) they contain laparoscopic skills and techniques that are usually present in many laparoscopic procedures (Table 1).

The vast majority of study participants considered that the SIMISGEST-VR was a useful tool for the development of hand-eye coordination and depth perception, with ratings of 4 and 5 on the Likert scale. Similarly, there was consensus about the capacity of the simulator to teach basic psychomotor skills and to reflect the basic steps in MIS. All the respondents considered the metrics to be adequate and envisaged that the simulator could become a solution to achieve ubiquitous learning of basic psychomotor skills in MIS.

In terms of the specific rating for each of the 6 tasks, this varied between 3.97 and 4.53. The participants considered all the items of the SIMISGEST-VR training system as good to excellent.

Finally, the study of fidelity to the criterion and content validity must be proven in the design stage of the artifact, before the criterion (concurrent and predictive) and construct validity (convergent and discriminative) can be confirmed. The evaluation of fidelity to the criterion, although somewhat subjective, is a necessary assessment during the initial phase of any high-stakes test construction and in this study, within the context of DBR, in the design phase of prototypes that will give a *good enough* prototype as a result [154,158,163]. In conclusion, the results of the study of fidelity to the criterion and content-related validity evidence showed overall positive scores.

Threats to Validity

The *Hawthorne effect* occurs when the opinion may be influenced by the attention paid to the respondent during his or her performance with the simulator, which may contribute to the occurrence of favorable responses or scores. This effect can be ameliorated by paying equal attention to each respondent. In addition, the *Pygmalion effect* occurs when the enthusiasm shown by the developers or because the novelty of the artifact affects the opinion of the respondent; the referent group is more prone to this latter effect [154,164]. In this study, the SIMISGEST-VR developer conducted the interviews and applied the Likert scale questionnaires; this may have influenced the ratings assigned by the participants (*Hawthorne effect)*.

Regarding the representation of the construct, in this study there was an underrepresentation—when compared with the learning models based on training boxes—referring to the *cut* skill of the *basic psychomotor skills* construct, which was because of technical reasons associated with the LMC (construct context). There was no overrepresentation of the construct [165].

Limitations

There are, however, limitations to this study. The sample size of this study was one of availability and, for the simulator to be portable and allow ubiquitous learning, the researchers disregarded some ergonomic principles applied to MIS [166,167]. Further research will be conducted using new motion metrics, new skill tasks, and the development of the web-based virtual environment for download as an app. In addition, the researchers of this study are working on the development of different difficulty levels for each exercise.

Future Work

The researchers of this study are currently conducting another study to show validity evidence for the *good enough* prototype described in this paper, using the new framework for validation in education [168,169]. This new study is expected to verify the sources of validity evidence for the internal structure, relationships between variables, and test consequences.

Once the metrics and the results of the performance scores have been validated as a useful tool for learning basic psychomotor skills in MIS, a model will be obtained to enable ubiquitous learning in MIS and preoperative warm-up by using the 3D reconstruction of patient images [14]. Studies conducted in this area have demonstrated that, generally speaking, preoperative warm-up exercises performed for at least 15 min before the procedure improve the surgeon's handling of soft tissue during cholecystectomy [170], bimanual skill, efficiency and smoothness of movement, and depth perception, at the same time as mistakes and operating time are reduced [171-177].

The large size and elevated costs of VR simulators currently available in the market prohibit their use in the operating theater. A portable, low-cost simulation solution, such as the SIMISGEST-VR, would allow surgeons to perform preoperative warm-up exercises anytime, anywhere (ubiquitous learning). In addition, the researchers aim to enable a surgeon to perform warm-up exercises based on 3D reconstructions of preoperative images of a specific patient, thus, practicing the procedure before performing the actual surgery. This could take place the night before in the surgeon's home or the operating theater on the day of the surgery [178-183]

Conclusions

This study demonstrated the feasibility of a portable, low-cost, gesture-based, functional simulator (SIMISGEST-VR) for learning basic psychomotor skills in MIS.

The results of the evaluation of fidelity to the criterion and content validity showed overall positive scores, which indicates that the SIMISGEST-VR would be acceptable to both the expert group and referent group as a training and learning device (including at home) to achieve ubiquitous learning in MIS.

The participants in the study agreed that content validity was acceptable, accurate, and representative in the field of basic psychomotor skills learning in MIS.

XSL•FO
RenderX

Authors' Contributions

All the authors contributed substantially to the study conception and design, data analysis, and interpretation of the findings and manuscript drafting. Fernando Álvarez López participated in the collection and assembly of data. Francesc Saigí-Rubió is the guarantor of the paper. All the authors have read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Application forms of the demographic survey, fidelity to the criterion, and content validity surveys.
[DOCX File , 20 KB - jmir_v22i7e17491_app1.docx]

Multimedia Appendix 2
Results of the fidelity to the criterion survey.
[DOCX File , 26 KB - jmir_v22i7e17491_app2.docx]

Multimedia Appendix 3
Results of the content validity survey.
[DOCX File , 38 KB - jmir_v22i7e17491_app3.docx]

References

1. Litynski GS. Profiles in laparoscopy: Mouret, Dubois, and Perissat: the laparoscopic breakthrough in Europe (1987-1988). JSLS-J Soc Laparoend 1999;3(2):163-167 [FREE Full text] [Medline: 10444020]
2. Yamashita Y, Kurohiji T, Kakegawa T. Evaluation of two training programs for laparoscopic cholecystectomy: incidence of major complications. World J Surg 1994;18(2):279-85; discussion 285. [doi: 10.1007/BF00294415] [Medline: 8042335]
3. Wolfe BM, Gardiner B, Frey CF. Laparoscopic cholecystectomy: a remarkable development. J Am Med Assoc 2015 Oct 6;314(13):1406. [doi: 10.1001/jama.2014.12014] [Medline: 26441196]
4. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/00000658-200210000-00008] [Medline: 12368674]
5. Dawe SR, Pena GN, Windsor JA, Broeders JA, Cregan PC, Hewett PJ, et al. Systematic review of skills transfer after surgical simulation-based training. Br J Surg 2014 Aug;101(9):1063-1076. [doi: 10.1002/bjs.9482] [Medline: 24827930]
6. Botden SM, Jakimowicz JJ. What is going on in augmented reality simulation in laparoscopic surgery? Surg Endosc 2009 Aug;23(8):1693-1700 [FREE Full text] [doi: 10.1007/s00464-008-0144-1] [Medline: 18813987]
7. Papanikolaou I. Assessment of medical simulators as a training programme for current surgical education. Hellenic J Surg 2013 Aug 9;85(4):240-248 [FREE Full text] [doi: 10.1007/s13126-013-0047-z]
8. Escamirosa FP, Flores RM, García IG, Vidal CR, Martínez AM. Face, content, and construct validity of the EndoViS training system for objective assessment of psychomotor skills of laparoscopic surgeons. Surg Endosc 2015 Nov;29(11):3392-3403. [doi: 10.1007/s00464-014-4032-6] [Medline: 25515985]
9. Sutton C, McCloy R, Middlebrook A, Chater P, Wilson M, Stone R. MIST VR. A laparoscopic surgery procedures trainer and evaluator. Stud Health Technol Inform 1997;39:598-607. [Medline: 10173070]
10. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. Stud Health Technol Inform 1998;50:124-130. [Medline: 10180527]
11. Ahlberg G, Heikkinen T, Iselius L, Leijonmarck C, Rutqvist J, Arvidsson D. Does training in a virtual reality simulator improve surgical performance? Surg Endosc 2002 Jan;16(1):126-129. [doi: 10.1007/s00464-001-9025-6] [Medline: 11961622]
12. Debes AJ, Aggarwal R, Balasundaram I, Jacobsen MB. A tale of two trainers: virtual reality versus a video trainer for acquisition of basic laparoscopic skills. Am J Surg 2010 Jun;199(6):840-845. [doi: 10.1016/j.amjsurg.2009.05.016] [Medline: 20079480]
13. Torkington J, Smith S, Rees B, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. Surg Endosc 2001 Oct;15(10):1076-1079. [doi: 10.1007/s004640000233] [Medline: 11727073]
14. Alvarez-Lopez F, Maina MF, Saigí-Rubió F. Use of commercial off-the-shelf devices for the detection of manual gestures in surgery: systematic literature review. J Med Internet Res 2019 Apr 14;21(5):e11925 [FREE Full text] [doi: 10.2196/11925] [Medline: 31066679]
15. Simon H. The Sciences of the Artificial. Third Edition. Cambridge, MA: MIT Press; 1996.
16. Manson N. Is operations research really research? ORiON 2006 Dec 1;22(2):-. [doi: 10.5784/22-2-40]

XSL•FO
RenderX

17. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75. [doi: 10.2307/25148625]
18. March ST, Smith GF. Design and natural science research on information technology. Decis Support Syst 1995 Dec;15(4):251-266. [doi: 10.1016/0167-9236(94)00041-2]
19. Hevner A, Chatterjee S. Design Research in Information Systems: Theory and Practice. New York, USA: Springer; 2010.
20. Dresch A, Pacheco-Lacerda D, Valle-Antunes J. Design Science Research: A Method for Science and Technology Advancement. New York, USA: Springer; 2015.
21. Takeda H, Veerkamp P, Tomiyama T, Yoshikawa H. Modeling design processes. AI Mag 1980 Dec 15;11(4):48. [doi: 10.1609/aimag.v11i4.855]
22. Vaishnavi V, Kuechler B, Petter S. Design Science Research in Information Systems. DESRIST. 2017. URL: http://www.desrist.org/design-research-in-information-systems/ [accessed 2020-05-20]
23. Ogura T, Sato M, Ishida Y, Hayashi N, Doi K. Development of a novel method for manipulation of angiographic images by use of a motion sensor in operating rooms. Radiol Phys Technol 2014 Jul;7(2):228-234. [doi: 10.1007/s12194-014-0259-0] [Medline: 24609904]
24. Mauser S, Burgert O. Touch-free, gesture-based control of medical devices and software based on the leap motion controller. Stud Health Technol Inform 2014;196:265-270. [Medline: 24732520]
25. Bachmann D, Weichert F, Rinkenauer G. Evaluation of the leap motion controller as a new contact-free pointing device. Sensors (Basel) 2014 Dec 24;15(1):214-233 [FREE Full text] [doi: 10.3390/s150100214] [Medline: 25609043]
26. Weichert F, Bachmann D, Rudak B, Fisseler D. Analysis of the accuracy and robustness of the leap motion controller. Sensors (Basel) 2013 May 14;13(5):6380-6393 [FREE Full text] [doi: 10.3390/s130506380] [Medline: 23673678]
27. Guna J, Jakus G, Pogačnik M, Tomažič S, Sodnik J. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors (Basel) 2014 Feb 21;14(2):3702-3720 [FREE Full text] [doi: 10.3390/s140203702] [Medline: 24566635]
28. Alvarez-Lopez F, Maina MF, Saigí-Rubió F. Natural user interfaces: is it a solution to accomplish ubiquitous training in minimally invasive surgery? Surg Innov 2016 Aug;23(4):429-430. [doi: 10.1177/1553350616639145] [Medline: 27009688]
29. Beck P. Accurate Three-dimensional Instrument Positioning. Google Patents. 2016. URL: https://patents.google.com/patent/US9918798B2/en [accessed 2020-05-29]
30. Takiyama K, Shinya M. Development of a portable motor learning laboratory (PoMLab). PLoS One 2016;11(6):e0157588 [FREE Full text] [doi: 10.1371/journal.pone.0157588] [Medline: 27348223]
31. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. Br Med J 2003 Nov 1;327(7422):1032-1037 [FREE Full text] [doi: 10.1136/bmj.327.7422.1032] [Medline: 14593041]
32. Mccloy R, Wilson M, Sutton C, Middlebrook A, Chater P, Stone R. MIST VR: apart-task virtual reality trainer for laparoscopic surgery. J Telemed Telecare 2016 Dec 2;3(1_Suppl):97. [doi: 10.1258/1357633971930652]
33. Seymour NE, Røtnes JS. Challenges to the development of complex virtual reality surgical simulations. Surg Endosc 2006 Nov;20(11):1774-1777. [doi: 10.1007/s00464-006-0107-3] [Medline: 16960667]
34. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005 Dec;19(12):1523-1532. [doi: 10.1007/s00464-005-0384-2] [Medline: 16252077]
35. Alessi S. Fidelity in the design of instructional simulations. J Comput Based Instr 1988;15(2):7. [doi: 10.5555/46756.46757]
36. Stone RJ. The (human) science of medical virtual learning environments. Philos Trans R Soc Lond B Biol Sci 2011 Jan 27;366(1562):276-285 [FREE Full text] [doi: 10.1098/rstb.2010.0209] [Medline: 21149363]
37. Kothari SN, Kaplan BJ, DeMaria EJ, Broderick TJ, Merrell RC. Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. J Laparoendosc Adv Surg Tech A 2002 Jun;12(3):167-173. [doi: 10.1089/10926420260188056] [Medline: 12184901]
38. Schijven M, Jakimowicz J. Virtual reality surgical laparoscopic simulators. Surg Endosc 2003 Dec;17(12):1943-1950. [doi: 10.1007/s00464-003-9052-6] [Medline: 14574546]
39. Bowman DA, Chen J, Wingrave CA, Lucas J, Ray A, Polys NF, et al. New directions in 3D user interfaces. Int J Virtual Real 2006 Jan 1;5(2):3-14. [doi: 10.20870/ijvr.2006.5.2.2683]
40. Schout BM, Hendrikx AJ, Scheele F, Bemelmans BL, Scherpbier AJ. Validation and implementation of surgical simulators: a critical review of present, past, and future. Surg Endosc 2010 Mar;24(3):536-546 [FREE Full text] [doi: 10.1007/s00464-009-0634-9] [Medline: 19633886]
41. Ali M, Mowery Y, Kaplan B, DeMaria E. Training the novice in laparoscopy. More challenge is better. Surg Endosc 2002 Dec;16(12):1732-1736. [doi: 10.1007/s00464-002-8850-6] [Medline: 12140638]
42. Sullivan K, Kantak S, Burtner P. Motor learning in children: feedback effects on skill acquisition. Phys Ther 2008 Jun;88(6):720-732. [doi: 10.2522/ptj.20070196] [Medline: 18339797]
43. Chaudhry A, Sutton C, Wood J, Stone R, McCloy R. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. Ann R Coll Surg Engl 1999 Jul;81(4):281-286 [FREE Full text] [Medline: 10615201]

XSL•FO
RenderX

44. Stylopoulos N, Vosburgh KG. Assessing technical skill in surgery and endoscopy: a set of metrics and an algorithm (C-PASS) to assess skills in surgical and endoscopic procedures. Surg Innov 2007 Jun;14(2):113-121. [doi: 10.1177/1553350607302330] [Medline: 17558017]
45. Oropesa I, de Jong T, Sánchez-González P, Dankelman J, Gómez E. Feasibility of tracking laparoscopic instruments in a box trainer using a leap motion controller. Measurement 2016 Feb;80:115-124. [doi: 10.1016/j.measurement.2015.11.018]
46. Wilson MS, Middlebrook A, Sutton C, Stone R, McCloy RF. MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. Ann R Coll Surg Engl 1997 Nov;79(6):403-404 [FREE Full text] [Medline: 9422863]
47. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg 2003 Feb;185(2):146-149. [doi: 10.1016/s0002-9610(02)01213-8] [Medline: 12559445]
48. Maithel S, Sierra R, Korndorffer J, Neumann P, Dawson S, Callery M, et al. Construct and face validity of MIST-VR, endotower, and CELTS: are we ready for skills assessment using simulators? Surg Endosc 2006 Jan;20(1):104-112. [doi: 10.1007/s00464-005-0054-4] [Medline: 16333535]
49. Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. Am J Surg 2006 Jan;191(1):128-133. [doi: 10.1016/j.amjsurg.2005.10.014] [Medline: 16399123]
50. van Sickle KR, Ritter EM, McClusky DA, Lederman A, Baghai M, Gallagher AG, et al. Attempted establishment of proficiency levels for laparoscopic performance on a national scale using simulation: the results from the 2004 SAGES minimally invasive surgical trainer-virtual reality (MIST-VR) learning center study. Surg Endosc 2007 Jan;21(1):5-10. [doi: 10.1007/s00464-006-0011-x] [Medline: 17111280]
51. Munz Y, Almoudaris AM, Moorthy K, Dosis A, Liddle AD, Darzi AW. Curriculum-based solo virtual reality training for laparoscopic intracorporeal knot tying: objective assessment of the transfer of skill from virtual reality to reality. Am J Surg 2007 Jun;193(6):774-783. [doi: 10.1016/j.amjsurg.2007.01.022] [Medline: 17512295]
52. Grantcharov TP, Rosenberg J, Pahle E, Funch-Jensen P. Virtual reality computer simulation. Surg Endosc 2001 Mar;15(3):242-244. [doi: 10.1007/s004640090008] [Medline: 11344422]
53. Paisley AM, Baldwin PJ, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. Br J Surg 2001 Nov;88(11):1525-1532. [doi: 10.1046/j.0007-1323.2001.01880.x] [Medline: 11683753]
54. Torkington J, Smith S, Rees B, Darzi A. The role of the basic surgical skills course in the acquisition and retention of laparoscopic skill. Surg Endosc 2001 Oct;15(10):1071-1075. [doi: 10.1007/s004640000183] [Medline: 11727072]
55. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001 Nov;25(11):1478-1483. [doi: 10.1007/s00268-001-0133-1] [Medline: 11760752]
56. Pearson A, Gallagher A, Rosser J, Satava R. Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. Surg Endosc 2002 Jan;16(1):130-137. [doi: 10.1007/s00464-001-8113-y] [Medline: 11961623]
57. Gallagher A, Satava R. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002 Dec;16(12):1746-1752. [doi: 10.1007/s00464-001-8215-6] [Medline: 12140641]
58. Kozlowski S, de Shon R. Scaled Worlds: development, validation, and applications. In: Inalas E, Elliott LR, Schflett SG, Coovert MD, editors. A Psychological Fidelity Approach to Simulation-based Training: Theory, Research, and Principles. Burlington, VT: Ashgate Publishing; 2004.
59. Gor M, McCloy R, Stone R, Smith A. Virtual reality laparoscopic simulator for assessment in gynaecology. BJOG 2003 Feb;110(2):181-187 [FREE Full text] [Medline: 12618163]
60. Gallagher AG, Smith CD, Bowers SP, Seymour NE, Pearson A, McNatt S, et al. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. J Am Coll Surg 2003 Sep;197(3):479-488. [doi: 10.1016/S1072-7515(03)00535-0] [Medline: 12946803]
61. Mackay S, Datta V, Chang A, Shah J, Kneebone R, Darzi A. Multiple objective measures of skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. Ann Surg 2003 Aug;238(2):291-300. [doi: 10.1097/01.sla.0000080829.29028.c4] [Medline: 12894024]
62. Stefanidis D, Korndorffer JR, Sierra R, Touchard C, Dunne JB, Scott DJ. Skill retention following proficiency-based laparoscopic simulator training. Surgery 2005 Aug;138(2):165-170. [doi: 10.1016/j.surg.2005.06.002] [Medline: 16153423]
63. Gallagher A, Satava R. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002 Dec;16(12):1746-1752. [doi: 10.1007/s00464-001-8215-6] [Medline: 12140641]
64. Taffinder N, McManus I, Gul Y, Russell R, Darzi A. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. Lancet 1998 Oct 10;352(9135):1191. [doi: 10.1016/s0140-6736(98)00034-8] [Medline: 9777838]
65. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. Stud Health Technol Inform 1998;50:124-130. [Medline: 10180527]
66. Jordan J, Gallagher AG, McGuigan J, McGlade K, McClure N. A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. Am J Surg 2000 Sep;180(3):208-211. [doi: 10.1016/s0002-9610(00)00469-4] [Medline: 11084131]

XSL•FO
RenderX

67. Gallagher AG, Hughes C, Reinhardt-Rutland AH, McGuigan J, McClure N. A case-control comparison of traditional and virtual-reality training in laparoscopic psychomotor performance. Minim Invasive Ther Allied Technol 2009 Jul 10;9(5):347-352. [doi: 10.3109/13645700009061457]
68. Jordan JA, Gallagher AG, McGuigan J, McClure N. Randomly alternating image presentation during laparoscopic training leads to faster automation to the 'fulcrum effect'. Endoscopy 2000 Apr;32(4):317-321. [doi: 10.1055/s-2000-7374] [Medline: 10774973]
69. Jordan J, Gallagher A, McGuigan J, McClure N. Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. Surg Endosc 2001 Oct;15(10):1080-1084. [doi: 10.1007/s004640000374] [Medline: 11727074]
70. McNatt S, Smith C. A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 2001 Oct;15(10):1085-1089. [doi: 10.1007/s004640080022] [Medline: 11727075]
71. Gallagher H, Allan J, Tolley D. Spatial awareness in urologists: are they different? BJU Int 2001 Nov;88(7):666-670 [FREE Full text] [doi: 10.1046/j.1464-4096.2001.02440.x] [Medline: 11890233]
72. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. Br Med J 2001 Nov 24;323(7323):1222-1223 [FREE Full text] [doi: 10.1136/bmj.323.7323.1222] [Medline: 11719413]
73. Hamilton E, Scott D, Fleming J, Rege R, Laycock R, Bergen P, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surg Endosc 2002 Mar;16(3):406-411. [doi: 10.1007/s00464-001-8149-z] [Medline: 11928017]
74. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc 2003 Jul;17(7):1082-1085. [doi: 10.1007/s00464-002-9176-0] [Medline: 12728373]
75. Eastridge BJ, Hamilton EC, O'Keefe GE, Rege RV, Valentine RJ, Jones DJ, et al. Effect of sleep deprivation on the performance of simulated laparoscopic surgical skill. Am J Surg 2003 Aug;186(2):169-174. [doi: 10.1016/s0002-9610(03)00183-1] [Medline: 12885613]
76. Madan AK, Frantzides CT, Shervin N, Tebbit CL. Assessment of individual hand performance in box trainers compared to virtual reality trainers. Am Surg 2003 Dec;69(12):1112-1114. [Medline: 14700302]
77. Bann S, Kwok K, Lo C, Darzi A, Wong J. Objective assessment of technical skills of surgical trainees in Hong Kong. Br J Surg 2003 Oct;90(10):1294-1299. [doi: 10.1002/bjs.4222] [Medline: 14515303]
78. Gonzalez R, Bowers SP, Smith CD, Ramshaw BJ. Does setting specific goals and providing feedback during training result in better acquisition of laparoscopic skills? Am Surg 2004 Jan;70(1):35-39. [Medline: 14964544]
79. Gallagher AG, Lederman AB, McGlade K, Satava RM, Smith CD. Discriminative validity of the minimally invasive surgical trainer in virtual reality (MIST-VR) using criteria levels based on expert performance. Surg Endosc 2004 Apr;18(4):660-665. [doi: 10.1007/s00464-003-8176-z] [Medline: 15026925]
80. Fried MP, Satava R, Weghorst S, Gallagher AG, Sasaki C, Ross D, et al. Identifying and reducing errors with surgical simulation. Qual Saf Health Care 2004 Oct;13(Suppl 1):i19-i26 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i19] [Medline: 15465950]
81. Ström P, Kjellin A, Hedman L, Wredmark T, Felländer-Tsai L. Training in tasks with different visual-spatial components does not improve virtual arthroscopy performance. Surg Endosc 2004 Jan;18(1):115-120. [doi: 10.1007/s00464-003-9023-y] [Medline: 14625735]
82. Uchal M, Raftopoulos Y, Tjugum J, Bergamaschi R. Validation of a six-task simulation model in minimally invasive surgery. Surg Endosc 2005 Jan;19(1):109-116. [doi: 10.1007/s00464-004-8145-1] [Medline: 15531971]
83. Adamsen S, Funch-Jensen PM, Drewes AM, Rosenberg J, Grantcharov TP. A comparative study of skills in virtual laparoscopy and endoscopy. Surg Endosc 2005 Feb;19(2):229-234. [doi: 10.1007/s00464-004-9090-8] [Medline: 15580316]
84. Avgerinos DV, Goodell KH, Waxberg S, Cao CG, Schwaitzberg SD. Comparison of the sensitivity of physical and virtual laparoscopic surgical training simulators to the user's level of experience. Surg Endosc 2005 Sep;19(9):1211-1215. [doi: 10.1007/s00464-004-8256-8] [Medline: 16132320]
85. Huang J, Payandeh S, Doris P, Hajshirmohammadi I. Fuzzy classification: towards evaluating performance on a surgical simulator. Stud Health Technol Inform 2005;111:194-200. [Medline: 15718726]
86. Madan AK, Frantzides CT, Tebbit C, Quiros RM. Participants' opinions of laparoscopic training devices after a basic laparoscopic training course. Am J Surg 2005 Jun;189(6):758-761. [doi: 10.1016/j.amjsurg.2005.03.022] [Medline: 15910733]
87. McClusky DA, Ritter EM, Lederman AB, Gallagher AG, Smith CD. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. Am Surg 2005 Jan;71(1):13-20; discussion 20. [Medline: 15757051]
88. Arora H, Uribe J, Ralph W, Zeltsan M, Cuellar H, Gallagher A, et al. Assessment of construct validity of the endoscopic sinus surgery simulator. Arch Otolaryngol Head Neck Surg 2005 Mar;131(3):217-221. [doi: 10.1001/archotol.131.3.217] [Medline: 15781761]

XSL•FO
RenderX

89. Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. Am J Surg 2005 Jun;189(6):753-757. [doi: 10.1016/j.amjsurg.2005.03.021] [Medline: 15910732]
90. Takiguchi S, Sekimoto M, Yasui M, Miyata H, Fujiwara Y, Yasuda T, et al. Cyber visual training as a new method for the mastery of endoscopic surgery. Surg Endosc 2005 Sep;19(9):1204-1210. [doi: 10.1007/s00464-004-8236-z] [Medline: 15942808]
91. DeMaria EJ, McBride CL, Broderick TJ, Kaplan BJ. Night call does not impair learning of laparoscopic skills. Surg Innov 2005 Jun;12(2):145-149. [doi: 10.1177/155335060501200213] [Medline: 16034504]
92. Hackethal A, Immenroth M, Bürger T. Evaluation of target scores and benchmarks for the traversal task scenario of the minimally invasive surgical trainer-virtual reality (MIST-VR) laparoscopy simulator. Surg Endosc 2006 Apr;20(4):645-650. [doi: 10.1007/s00464-004-2224-1] [Medline: 16424991]
93. Ström P, Hedman L, Särnå L, Kjellin A, Wredmark T, Felländer-Tsai L. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. Surg Endosc 2006 Sep;20(9):1383-1388. [doi: 10.1007/s00464-005-0545-3] [Medline: 16823652]
94. Goodell KH, Cao CG, Schwaitzberg SD. Effects of cognitive distraction on performance of laparoscopic surgical tasks. J Laparoendosc Adv Surg Tech A 2006 Apr;16(2):94-98. [doi: 10.1089/lap.2006.16.94] [Medline: 16646695]
95. Hedman L, Ström P, Andersson P, Kjellin A, Wredmark T, Felländer-Tsai L. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. Surg Endosc 2006 Aug;20(8):1275-1280. [doi: 10.1007/s00464-005-0036-6] [Medline: 16865624]
96. Hedman L, Klingberg T, Enochsson L, Kjellin A, Felländer-Tsai L. Visual working memory influences the performance in virtual image-guided surgical intervention. Surg Endosc 2007 Nov;21(11):2044-2050. [doi: 10.1007/s00464-007-9287-8] [Medline: 17522939]
97. Madan AK, Frantzides CT. Substituting virtual reality trainers for inanimate box trainers does not decrease laparoscopic skills acquisition. JSLS 2007;11(1):87-89 [FREE Full text] [Medline: 17651563]
98. Tanoue K, Ieiri S, Konishi K, Yasunaga T, Okazaki K, Yamaguchi S, et al. Effectiveness of endoscopic surgery training for medical students using a virtual reality simulator versus a box trainer: a randomized controlled trial. Surg Endosc 2008 Apr;22(4):985-990. [doi: 10.1007/s00464-007-9554-8] [Medline: 17710487]
99. Shane MD, Pettitt BJ, Morgenthal CB, Smith CD. Should surgical novices trade their retractors for joysticks? Videogame experience decreases the time needed to acquire surgical skills. Surg Endosc 2008 May;22(5):1294-1297. [doi: 10.1007/s00464-007-9614-0] [Medline: 17972136]
100. Schlickum MK, Hedman L, Enochsson L, Kjellin A, Felländer-Tsai L. Transfer of systematic computer game training in surgical novices on performance in virtual reality image guided surgical simulators. Stud Health Technol Inform 2008;132:210-215. [Medline: 18391288]
101. Windsor JA, Diener S, Zoha F. Learning style and laparoscopic experience in psychomotor skill performance using a virtual reality surgical simulator. Am J Surg 2008 Jun;195(6):837-842. [doi: 10.1016/j.amjsurg.2007.09.034] [Medline: 18417084]
102. Goldberg AE, Neifeld JP, Wolfe LG, Goldberg SR. Correlation of manual dexterity with USMLE scores and medical student class rank. J Surg Res 2008 Jun 15;147(2):212-215. [doi: 10.1016/j.jss.2008.02.050] [Medline: 18498872]
103. Botden SM, de Hingh IH, Jakimowicz JJ. Suturing training in augmented reality: gaining proficiency in suturing skills faster. Surg Endosc 2009 Sep;23(9):2131-2137 [FREE Full text] [doi: 10.1007/s00464-008-0240-2] [Medline: 19067051]
104. Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. Am J Surg 2009 Apr;197(4):447-449. [doi: 10.1016/j.amjsurg.2008.01.024] [Medline: 19217604]
105. Schlickum MK, Hedman L, Enochsson L, Kjellin A, Felländer-Tsai L. Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. World J Surg 2009 Nov;33(11):2360-2367. [doi: 10.1007/s00268-009-0151-y] [Medline: 19649553]
106. Moore A, Grow D, Bush R, Seymour N. Novices outperform experienced laparoscopists on virtual reality laparoscopy simulator. JSLS 2008;12(4):358-362 [FREE Full text] [Medline: 19275848]
107. van Sickle KR, Ritter EM, Baghai M, Goldenberg AE, Huang I, Gallagher AG, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. J Am Coll Surg 2008 Oct;207(4):560-568. [doi: 10.1016/j.jamcollsurg.2008.05.007] [Medline: 18926460]
108. Kanumuri P, Ganai S, Wohaibi EM, Bush RW, Grow DR, Seymour NE. Virtual reality and computer-enhanced training devices equally improve laparoscopic surgical skill in novices. JSLS 2008;12(3):219-226 [FREE Full text] [Medline: 18765042]
109. Madan AK, Harper JL, Frantzides CT, Tichansky DS. Nonsurgical skills do not predict baseline scores in inanimate box or virtual-reality trainers. Surg Endosc 2008 Jul;22(7):1686-1689. [doi: 10.1007/s00464-007-9691-0] [Medline: 18071808]
110. Arora S, Sevdalis N, Aggarwal R, Sirimanna P, Darzi A, Kneebone R. Stress impairs psychomotor performance in novice laparoscopic surgeons. Surg Endosc 2010 Oct;24(10):2588-2593. [doi: 10.1007/s00464-010-1013-2] [Medline: 20354878]

XSL•FO
RenderX

111. Gallagher AG, Boyle E, Toner P, Neary PC, Andersen DK, Satava RM, et al. Persistent next-day effects of excessive alcohol consumption on laparoscopic surgical performance. Arch Surg 2011 Apr;146(4):419-426. [doi: 10.1001/archsurg.2011.67] [Medline: 21502449]
112. Zhou M, Tse S, Derevianko A, Jones DB, Schwaitzberg SD, Cao CG. Effect of haptic feedback in laparoscopic surgery skill acquisition. Surg Endosc 2012 Apr;26(4):1128-1134 [FREE Full text] [doi: 10.1007/s00464-011-2011-8] [Medline: 22044975]
113. Hedman L, Schlickum M, Felländer-Tsai L. Surgical novices randomized to train in two video games become more motivated during training in MIST-VR and GI Mentor II than students with no video game training. Stud Health Technol Inform 2013;184:189-194. [Medline: 23400154]
114. Gallagher AG, Seymour NE, Jordan-Black J, Bunting BP, McGlade K, Satava RM. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. Ann Surg 2013 Jun;257(6):1025-1031. [doi: 10.1097/SLA.0b013e318284f658] [Medline: 23426342]
115. Schlickum M, Hedman L, Felländer-Tsai L. Visual-spatial ability is more important than motivation for novices in surgical simulator training: a preliminary study. Int J Med Educ 2016 Feb 21;7:56-61 [FREE Full text] [doi: 10.5116/ijme.56b1.1691] [Medline: 26897701]
116. Oussi N, Loukas C, Kjellin A, Lahanas V, Georgiou K, Henningsohn L, et al. Video analysis in basic skills training: a way to expand the value and use of BlackBox training? Surg Endosc 2018 Jan;32(1):87-95 [FREE Full text] [doi: 10.1007/s00464-017-5641-7] [Medline: 28664435]
117. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). Endoscopy 1999 May;31(4):310-313. [doi: 10.1055/s-1999-15] [Medline: 10376458]
118. Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training: a randomized controlled trial. Surg Endosc 2002 Jun;16(6):957-961. [doi: 10.1007/s00464-001-9132-4] [Medline: 12163963]
119. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004 Feb;91(2):146-150. [doi: 10.1002/bjs.4407] [Medline: 14760660]
120. Brunner WC, Korndorffer JR, Sierra R, Massarweh NN, Dunne J, Yau C, et al. Laparoscopic virtual reality training: are 30 repetitions enough? J Surg Res 2004 Dec;122(2):150-156. [doi: 10.1016/j.jss.2004.08.006] [Medline: 15555611]
121. Brunner W, Korndorffer J, Sierra R, Dunne B, Yau L, Corsetti R, et al. Determining standards for laparoscopic proficiency using virtual reality. Am Surg 2005 Jan;71(1):29-35. [Medline: 15757053]
122. Madan AK, Frantzides CT, Sasso LM. Laparoscopic baseline ability assessment by virtual reality. J Laparoendosc Adv Surg Tech A 2005 Feb;15(1):13-17. [doi: 10.1089/lap.2005.15.13] [Medline: 15772470]
123. Schijven MP, Jakimowicz JJ, Broeders IA, Tseng LN. The Eindhoven laparoscopic cholecystectomy training course--improving operating room performance using virtual reality training: results from the first EAES accredited virtual reality trainings curriculum. Surg Endosc 2005 Sep;19(9):1220-1226. [doi: 10.1007/s00464-004-2240-1] [Medline: 16132332]
124. Barnes RW. Surgical handicraft: teaching and learning surgical skills. Am J Surg 1987 May;153(5):422-427. [doi: 10.1016/0002-9610(87)90783-5] [Medline: 3578661]
125. Karaliotas C. When simulation in surgical training meets virtual reality. Hellenic J Surg 2012 Feb 19;83(6):303-316. [doi: 10.1007/s13126-011-0055-9]
126. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. Ann Surg 2008 Aug;248(2):166-179. [doi: 10.1097/SLA.0b013e318176bf24] [Medline: 18650625]
127. Sánchez-Margallo FM, Sánchez-Margallo JA, Moyano-Cuevas JL, Pérez EM, Maestre J. Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience. Minim Invasive Ther Allied Technol 2017 Oct;26(5):253-261. [doi: 10.1080/13645706.2017.1304964] [Medline: 28349758]
128. Xu X, Zheng Y, Yao S, Sun G, Xu B, Chen X. A low-cost multimodal head-mounted display system for neuroendoscopic surgery. Brain Behav 2018 Jan 14;8(1):e00891 [FREE Full text] [doi: 10.1002/brb3.891] [Medline: 29568688]
129. Den AJ, Bannan B, Kelly A, Plomp T. SLO Netherlands institute for curriculum development part B: illustrative cases. In: Educational Design Research. New York, USA: Springer; 2013.
130. Lahanas V, Loukas C, Georgiou K, Lababidi H, Al-Jaroudi D. Virtual reality-based assessment of basic laparoscopic skills using the leap motion controller. Surg Endosc 2017 Dec;31(12):5012-5023. [doi: 10.1007/s00464-017-5503-3] [Medline: 28466361]
131. Noble C. The relationship between fidelity and learning in aviation training and assessment. J Air Transp 2002;7(3):33-54 [FREE Full text]
132. Park A, Witzke D. Training and educational approaches to minimally invasive surgery: state of the art. Semin Laparosc Surg 2002 Dec;9(4):198-205. [doi: 10.1053/slas.2002.36468] [Medline: 12522776]
133. Schijven M, Jakimowicz J. Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 2003 May;17(5):803-810. [doi: 10.1007/s00464-002-9151-9] [Medline: 12582752]
134. Badash I, Burtt K, Solorzano CA, Carey JN. Innovations in surgery simulation: a review of past, current and future techniques. Ann Transl Med 2016 Dec;4(23):453 [FREE Full text] [doi: 10.21037/atm.2016.12.24] [Medline: 28090509]

XSL•FO
RenderX

135. Stefanidis D, Scott DJ, Korndorffer JR. Do metrics matter? Time versus motion tracking for performance assessment of proficiency-based laparoscopic skills training. Simul Healthc 2009;4(2):104-108. [doi: 10.1097/SIH.0b013e31819171ec] [Medline: 19444048]
136. Archer J. State of the science in health professional education: effective feedback. Med Educ 2010 Jan;44(1):101-108. [doi: 10.1111/j.1365-2923.2009.03546.x] [Medline: 20078761]
137. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. Acad Med 2013 Jun;88(6):872-883. [doi: 10.1097/ACM.0b013e31828ffdcf] [Medline: 23619073]
138. Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. Adv Health Sci Educ Theory Pract 2014 May;19(2):251-272. [doi: 10.1007/s10459-013-9462-8] [Medline: 23712700]
139. Walsh CM, Ling SC, Wang CS, Carnahan H. Concurrent versus terminal feedback: it may be better to wait. Acad Med 2009 Oct;84(10 Suppl):S54-S57. [doi: 10.1097/ACM.0b013e3181b38daf] [Medline: 19907387]
140. Stefanidis D, Korndorffer JR, Heniford BT, Scott DJ. Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. Surgery 2007 Aug;142(2):202-206. [doi: 10.1016/j.surg.2007.03.009] [Medline: 17689686]
141. Gallagher AG, McClure N, McGuigan J, Ritchie K, Sheehy NP. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. Endoscopy 1998 Sep;30(7):617-620. [doi: 10.1055/s-2007-1001366] [Medline: 9826140]
142. Crothers IR, Gallagher AG, McClure N, James DT, McGuigan J. Experienced laparoscopic surgeons are automated to the 'fulcrum effect': an ergonomic demonstration. Endoscopy 1999 Jun;31(5):365-369. [doi: 10.1055/s-1999-26] [Medline: 10433045]
143. Chan E. Standards and guidelines for validation practices: development and evaluation of measurement instruments. In: Zumbo BD, Chan EK, editors. Validity and Validation in Social, Behavioral, and Health Sciences. New York, USA: Springer International Publishing; 2014.
144. Mosier CI. A critical examination of the concepts of face validity. Educ Psychol Meas 1947;7(2):191-205. [doi: 10.1177/001316444700700201] [Medline: 20256558]
145. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006 Feb;119(2):166.e7-166.16. [doi: 10.1016/j.amjmed.2005.10.036] [Medline: 16443422]
146. Sweet RM, Hananel D, Lawrenz F. A unified approach to validation, reliability, and education study design for surgical technical skills training. Arch Surg 2010 Feb;145(2):197-201. [doi: 10.1001/archsurg.2009.266] [Medline: 20157089]
147. Broe D, Ridgway PF, Johnson S, Tierney S, Conlon KC. Construct validation of a novel hybrid surgical simulator. Surg Endosc 2006 Jun;20(6):900-904. [doi: 10.1007/s00464-005-0530-x] [Medline: 16738979]
148. Downing SM. Face validity of assessments: faith-based interpretations or evidence-based science? Med Educ 2006 Jan;40(1):7-8. [doi: 10.1111/j.1365-2929.2005.02361.x] [Medline: 16441314]
149. Rod J, Marret J, Kohaut J, Aigrain Y, Jais JP, de Vries P, et al. Low-cost training simulator for open dismembered pyeloplasty: development and face validation. J Surg Educ 2018;75(1):188-194. [doi: 10.1016/j.jsurg.2017.06.010] [Medline: 28778782]
150. Leijte E, Arts E, Witteman B, Jakimowicz J, de Blaauw I, Botden S. Construct, content and face validity of the eoSim laparoscopic simulator on advanced suturing tasks. Surg Endosc 2019 Nov;33(11):3635-3643 [FREE Full text] [doi: 10.1007/s00464-018-06652-3] [Medline: 30671668]
151. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. Curr Probl Surg 2002 Jun;39(6):573-659. [doi: 10.1067/mog.2002.123481] [Medline: 12037512]
152. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003 Oct;17(10):1525-1529. [doi: 10.1007/s00464-003-0035-4] [Medline: 14502403]
153. Bajka M, Tuchschmid S, Streich M, Fink D, Székely G, Harders M. Evaluation of a new virtual-reality training simulator for hysteroscopy. Surg Endosc 2009 Sep;23(9):2026-2033. [doi: 10.1007/s00464-008-9927-7] [Medline: 18437471]
154. Ayodeji ID, Schijven M, Jakimowicz J, Greve JW. Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum. Surg Endosc 2007 Sep;21(9):1641-1649. [doi: 10.1007/s00464-007-9219-7] [Medline: 17356944]
155. Schijven MP, Jakimowicz JJ. Validation of virtual reality simulators: key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005;14(4):244-246. [doi: 10.1080/13645700500221881] [Medline: 16754170]
156. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
157. Korndorffer JR, Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. Am J Surg 2010 Jan;199(1):99-104. [doi: 10.1016/j.amjsurg.2009.08.018] [Medline: 20103073]

XSL•FO
RenderX

158. Nicholson WJ, Cates CU, Patel AD, Niazi K, Palmer S, Helmy T, et al. Face and content validation of virtual reality simulation for carotid angiography: results from the first 100 physicians attending the emory neuroanatomy carotid training (ENACT) program. Simul Healthc 2006;1(3):147-150. [doi: 10.1097/01.sih.0000244457.30080.fc] [Medline: 19088583]
159. McDougall EM, Corica FA, Boker JR, Sala LG, Stoliar G, Borin JF, et al. Construct validity testing of a laparoscopic surgical simulator. J Am Coll Surg 2006 May;202(5):779-787. [doi: 10.1016/j.jamcollsurg.2006.01.004] [Medline: 16648018]
160. Goodwin LD. Changing conceptions of measurement validity: an update on the new standards. J Nurs Educ 2002 Mar;41(3):100-106. [Medline: 11939227]
161. Straub D, Gefen D. Validation guidelines for IS positivist research. Commun Assoc Inf Syst 2004;13:-. [doi: 10.17705/1CAIS.01324]
162. Gallagher A, O'Sullivan G. Metrics for the measurement of skill. In: Fundamentals of Surgical Simulation: Principles and Practice. London, UK: Springer London; 2012:123-153.
163. Fried GM. Simulators for laparoscopic surgery: a coming of age. Asian J Surg 2004 Jan;27(1):1-3 [FREE Full text] [doi: 10.1016/S1015-9584(09)60235-9] [Medline: 14709418]
164. Schijven M, Jakimowicz J. Face-, expert, and referent validity of the Xitact LS500 laparoscopy simulator. Surg Endosc 2002 Dec;16(12):1764-1770. [doi: 10.1007/s00464-001-9229-9] [Medline: 12098029]
165. Andreatta P, Gruppen L. Conceptualising and classifying validity evidence for simulation. Med Educ 2009 Nov;43(11):1028-1035. [doi: 10.1111/j.1365-2923.2009.03454.x] [Medline: 19874494]
166. Xiao D, Jakimowicz JJ, Albayrak A, Buzink SN, Botden SM, Goossens RH. Face, content, and construct validity of a novel portable ergonomic simulator for basic laparoscopic skills. J Surg Educ 2014;71(1):65-72. [doi: 10.1016/j.jsurg.2013.05.003] [Medline: 24411426]
167. Emam TA, Hanna GB, Kimber C, Dunkley P, Cuschieri A. Effect of intracorporeal-extracorporeal instrument length ratio on endoscopic task performance and surgeon movements. Arch Surg 2000 Jan;135(1):62-5; discussion 66. [doi: 10.1001/archsurg.135.1.62] [Medline: 10636349]
168. Messick S. Standards of validity and the validity of standards in performance assessment. Educ Meas Issues Pract 1995;14(4):5-8. [doi: 10.1111/j.1745-3992.1995.tb00881.x]
169. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ 2003 Sep;37(9):830-837. [doi: 10.1046/j.1365-2923.2003.01594.x] [Medline: 14506816]
170. Moldovanu R, Târcoveanu E, Dimofte G, Lupaşcu C, Bradea C. Preoperative warm-up using a virtual reality simulator. JSLS 2011;15(4):533-538 [FREE Full text] [doi: 10.4293/108680811X13176785204409] [Medline: 22643511]
171. Calatayud D, Arora S, Aggarwal R, Kruglikova I, Schulze S, Funch-Jensen P, et al. Warm-up in a virtual reality environment improves performance in the operating room. Ann Surg 2010 Jun;251(6):1181-1185. [doi: 10.1097/SLA.0b013e3181deb630] [Medline: 20485133]
172. Abdalla G, Moran-Atkin E, Chen G, Schweitzer MA, Magnuson TH, Steele KE. The effect of warm-up on surgical performance: a systematic review. Surg Endosc 2015 Jun;29(6):1259-1269. [doi: 10.1007/s00464-014-3811-4] [Medline: 25149639]
173. Kahol K, Satava RM, Ferrara J, Smith ML. Effect of short-term pretrial practice on surgical proficiency in simulated environments: a randomized trial of the 'preoperative warm-up' effect. J Am Coll Surg 2009 Feb;208(2):255-268. [doi: 10.1016/j.jamcollsurg.2008.09.029] [Medline: 19228538]
174. Moran-Atkin E, Abdalla G, Chen G, Magnuson TH, Lidor AO, Schweitzer MA, et al. Preoperative warm-up the key to improved resident technique: a randomized study. Surg Endosc 2015 May;29(5):1057-1063. [doi: 10.1007/s00464-014-3778-1] [Medline: 25249142]
175. Pike TW, Pathak S, Mushtaq F, Wilkie RM, Mon-Williams M, Lodge JP. A systematic examination of preoperative surgery warm-up routines. Surg Endosc 2017 May;31(5):2202-2214. [doi: 10.1007/s00464-016-5218-x] [Medline: 27633438]
176. Chen CC, Green IC, Colbert-Getz JM, Steele K, Chou B, Lawson SM, et al. Warm-up on a simulator improves residents' performance in laparoscopic surgery: a randomized trial. Int Urogynecol J 2013 Oct;24(10):1615-1622. [doi: 10.1007/s00192-013-2066-2] [Medline: 23575698]
177. Plerhoples TA, Zak Y, Hernandez-Boussard T, Lau J. Another use of the mobile device: warm-up for laparoscopic surgery. J Surg Res 2011 Oct;170(2):185-188. [doi: 10.1016/j.jss.2011.03.015] [Medline: 21529831]
178. Soler L, Marescaux J. Patient-specific surgical simulation. World J Surg 2008 Feb;32(2):208-212. [doi: 10.1007/s00268-007-9329-3] [Medline: 18066615]
179. Marescaux J, Clément JM, Tassetti V, Koehl C, Cotin S, Russier Y, et al. Virtual reality applied to hepatic surgery simulation: the next revolution. Ann Surg 1998 Nov;228(5):627-634. [doi: 10.1097/00000658-199811000-00001] [Medline: 9833800]
180. Shiozawa M, Sata N, Endo K, Koizumi M, Yasuda Y, Nagai H, et al. Preoperative virtual simulation of adrenal tumors. Abdom Imaging 2009;34(1):113-120. [doi: 10.1007/s00261-008-9364-z] [Medline: 18253779]
181. Willaert WI, Aggarwal R, Van Herzeele I, Cheshire NJ, Vermassen FE. Recent advancements in medical simulation: patient-specific virtual reality simulation. World J Surg 2012 Jul;36(7):1703-1712. [doi: 10.1007/s00268-012-1489-0] [Medline: 22532308]

XSL•FO
RenderX

182. Endo K, Sata N, Ishiguro Y, Miki A, Sasanuma H, Sakuma Y, et al. A patient-specific surgical simulator using preoperative imaging data: an interactive simulator using a three-dimensional tactile mouse. J Comput Surg 2014 Aug 6;1(1):-. [doi: 10.1186/s40244-014-0010-5]
183. Makiyama K, Yamanaka H, Ueno D, Ohsaka K, Sano F, Nakaigawa N, et al. Validation of a patient-specific simulator for laparoscopic renal surgery. Int J Urol 2015 Jun;22(6):572-576 [FREE Full text] [doi: 10.1111/iju.12737] [Medline: 25721949]

Abbreviations

MIS: minimally invasive surgery
MIST-VR: minimally invasive surgery training—virtual reality
LMC: Leap Motion Controller
DBR: design-based research
SIMISGEST-VR: gesture-mediated simulator for minimally invasive surgery—virtual reality

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by S Jung, Z He; comments to author 13.02.20; revised version received 26.02.20; accepted 29.02.20; published 14.07.20.

Please cite as:
Alvarez-Lopez F, Maina MF, Saigí-Rubió F
Use of a Low-Cost Portable 3D Virtual Reality Gesture-Mediated Simulator for Training and Learning Basic Psychomotor Skills in Minimally Invasive Surgery: Development and Content Validity Study
J Med Internet Res 2020;22(7):e17491
URL: http://www.jmir.org/2020/7/e17491/
doi:10.2196/17491
PMID:32673217

©Fernando Alvarez-Lopez, Marcelo Fabián Maina, Francesc Saigí-Rubió. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Building a Successful Massive Open Online Course About Multiple Sclerosis: A Process Description

Suzi B Claflin[1], PhD; Rachael Gates[1,2], MBA; Maree Maher[3], BSN; Bruce V Taylor[1], MD

[1]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

[2]Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Australia

[3]Multiple Sclerosis Limited, Blackburn, Australia

Corresponding Author:
Suzi B Claflin, PhD
Menzies Institute for Medical Research
University of Tasmania
17 Liverpool Street
Hobart, 7000
Australia
Phone: 61 0362267728
Email: suzi.claflin@utas.edu.au

Abstract

Background: Over the course of a year, we developed and tested a 6-week massive open online course (MOOC) on multiple sclerosis (MS) in consultation with the MS community. The course targeted the MS community and interested laypeople and was titled *Understanding MS*. The primary purpose of the course was to improve MS knowledge, health literacy, and resilience among participants. The final version of the MOOC made available for open enrollment was ranked first among all MOOCs released in 2019 (n>2400) based on participant reviews.

Objective: The aim of this study was to present a detailed description and assessment of the development process of the *Understanding MS* MOOC.

Methods: The development process included a course development focus group; the creation of more than 50 content videos and related text, quizzes, activities, and discussion prompts; the creation of original images and animations; a pilot study; and collaborations with people living with MS, MS nurses, allied health care practitioners, and neurologists and researchers from 4 universities.

Results: Overall, the process was efficient and effective. With a few small changes, we recommend this approach to those seeking to develop a similar course. This process led to the development of a highly reviewed MOOC with excellent user satisfaction.

Conclusions: We identified 5 key lessons from this process: (1) community support is essential, (2) stakeholder involvement improves content quality, (3) plan for research from the beginning, (4) coordination between the academic lead and project manager team ensures a consistent voice, and (5) a network of collaborators is a key resource.

(J Med Internet Res 2020;22(7):e16687) doi:10.2196/16687

KEYWORDS

multiple sclerosis; health education

Introduction

Background

Massive open online courses (MOOCs) have the potential to extend the classroom to people who cannot access traditional learning environments by capitalizing on the widespread availability of the internet [1]. The defining characteristics of MOOCs are that they are freely available and have virtually unlimited enrollment capacity. These features make MOOCs versatile; they can be used to present information on a range of topics to an array of different audience.

Over the past decade, there has been a rapid increase in the development of MOOCs. Class Central, the leading MOOC aggregator site, currently lists more than 13,000 MOOCs from more than 900 universities [2]. However, there are significant challenges in MOOC development and implementation,

XSL•FO
RenderX

including the cost and time involved [3,4], with universities spending US $200,000 or more per course [4]. These challenges suggest that it is worth evaluating the approaches used to create MOOCs to identify strategies that are cost-effective and result in successful courses.

Furthermore, little is known about the impact of MOOCs on learning outcomes. Addressing this knowledge gap requires meaningful comparisons of many courses [5]. These comparisons require an understanding of the similarities and differences in course development and content. However, little research has been done describing the process of developing an MOOC in detail [6,7]. Here, we present a detailed description and assessment of our work—developing an MOOC on multiple sclerosis (MS) for the MS community and interested laypeople—including all tools and materials that we used or produced, in the hope that it assists others in the creation of similar content and allows for more meaningful comparisons between courses.

MS is a disease of the central nervous system that affects approximately 2.3 million people worldwide [8]. People living with MS often experience mobility and fatigue-related symptoms that can interfere with their ability to access health care [9]. Consequently, several digital and remote communication disease management technologies have been developed for people living with MS [10]. MS symptoms may also impede the ability of people living with MS to access traditional educational offerings (eg, in-person courses), making it difficult for them to travel to a particular meeting place or maintain active participation over long sessions. Therefore, an MOOC may be a more appropriate vehicle for knowledge translation for people living with MS and other members of the MS community (eg, family and friends, caregivers, and health care providers), particularly for those who may feel isolated by living with a rare condition.

Objectives

Here, we describe and assess our experience in developing a successful 6-week English-language MOOC about MS. The first enrollment ran from April to June 2019 and the second ran from September to November 2019, with more than 8000 people enrolling and an average completion rate of 47% (SB Claflin, unpublished data, 2019), far exceeding the average for MOOCs, which fluctuates between 5% and 15% [11,12]. On the basis of participant reviews from these 2 enrollments, the course was ranked first among more than 2400 MOOCs released in 2019 and third among all health and medicine MOOCs globally by the leading online course aggregator site, Class Central [13,14]. In this paper, we described the process we used to develop the *Understanding MS* MOOC in detail, highlighting the strengths and weaknesses of our approach and making recommendations for the development of similar interventions.

Methods

Ethics

The work conducted for this project was approved by the Social Science Human Research Ethics Committee at the University of Tasmania (UTAS), including the focus group (H0017241) and a pilot study (H0017778).

Funding and Costs

The Understanding MS MOOC was collaboratively developed by the MS Flagship at the Menzies Institute for Medical Research (Menzies) at UTAS and Multiple Sclerosis Limited, a service organization providing resources to the MS communities of New South Wales, Victoria, Tasmania, and the Australian Capital Territory. These organizations cofunded the project, resulting in a total budget of approximately AUD $200,000 (US $139,000). The majority of these funds were used to pay the salary of a full-time postdoctoral research fellow who served as the academic lead of the project (SC) and the secondment of a MOOC project manager (RG) from the Wicking Center for Dementia Research and Education (WDREC). A small amount of funding was used to develop some of the course animations and to pay for travel costs and video transcription.

The remainder of the costs related to course development were in-kind, provided in either time or materials. The area experts who presented the course videos did it for free. The UTAS employs a videographer who filmed all but a few of the videos. These videos were primarily shot in studios that were generously offered at no cost by the local Australian Broadcasting Corporation and the University of Melbourne. All video editing was performed by the UTAS staff. All images were reproduced for free with permission from the original source or developed by the UTAS staff. A few animations were also produced in-house. Finally, the course was hosted on a custom UTAS-built MOOC platform, which was originally constructed to host MOOCs produced by the WDREC [5].

Management Structure

The development of the MOOC was primarily overseen by 2 full-time employees: the academic lead and the project manager. The academic lead was responsible for the development of course content, including syllabus, video outlines, text, quizzes, discussion prompts, and activities. The academic lead was also responsible for video presentations, both presenting by herself and liaising with other area experts. In addition, she was responsible for developing the research projects surrounding the MOOC and leading the focus group, pilot study, and development of the MS Knowledge Assessment Scale. The project manager was responsible for coordinating and overseeing video production (including editing and transcription; scheduling filming; and liaising with external contractors, such as the video animator), establishing branding and style guidelines for the course, liaising on technical requirements for the web-based learning management system, coordinating digital marketing strategies for the course, coordinating and implementing communication processes for participant engagement and retention, and general project planning and administration tasks. The academic lead and project manager worked closely together, usually meeting for the project several times per week and sometimes as often as several times per day.

The project was also overseen by an advisory group that met fortnightly throughout the project. The group included a clinician-researcher who served as the academic lead's direct supervisor (BT); Menzies MS Flagship project manager; Menzies MS Flagship communications officer; Menzies business

XSL•FO
RenderX

manager; MS researchers, including an epidemiologist, a neuroscientist, and a health economist; a web-based learning and systems support manager; and a liaison from Multiple Sclerosis Limited (MM). The group contained a wide range of MS community expertise, from the lived experience of people living with MS to the clinical expertise of a neurologist and an MS nurse. This group provided feedback on course content and study design, contributed their expertise to course videos, and recruited other area experts. All the members of the advisory group were asked to provide their feedback in preparation for this assessment.

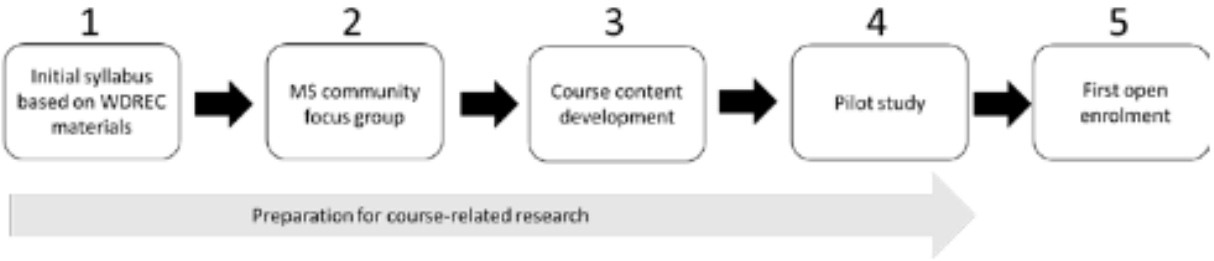
Core Principles and Development Outline

The advisory group developed 3 core principles that guided our work:

1. *Do no harm*. The course should not frighten participants and should be sensitive to participant anxiety.
2. *Keep it positive*. We used positive language in the course and course-related materials.
3. *Share the journey*. We wanted the course to be inclusive in design, content, and presentation and involve a diverse group of MS community members in all stages of development.

Course development included 4 distinct steps leading to the implementation of the course (initial syllabus development, focus group, content development, and pilot study) in the first open enrollment. Alongside this work, we carried out preparations for course-related research. The steps of the development process are illustrated in Figure 1.

Figure 1. Course development flowchart. MS: multiple sclerosis; WDREC: Wicking Center for Dementia Research and Education.



Initial Course Syllabus Development

The initial course syllabus was modeled on the work of the WDREC. WDREC previously developed 2 extremely successful MOOCs, Understanding Dementia and Preventing Dementia [15]. The syllabuses for these courses were used as an initial template for the Understanding MS course syllabus. This template was then adapted to fit the MS context. This included adjusting the content to ensure its relevance to people living with MS, as the WDREC courses are primarily taken by caregivers rather than people living with dementia. The initial course syllabus is available in the supplementary materials (Multimedia Appendix 1).

Focus Group

The main purpose of the focus group was to determine the content and modes of delivery that were most important and/or acceptable to the MS community, particularly people living with MS. The focus group participants were purposefully recruited through existing relationships with Multiple Sclerosis Limited to ensure that the focus group was broadly representative of the larger MS community (eg, representation of multiple MS phenotypes, male and female representation, and a variety of MS community roles). Before participation in the focus group, small group facilitators were given a 1.5-hour facilitator training by an experienced facilitator and a translational researcher. This training covered the protocol of the focus group and facilitation approaches and techniques.

The focus group followed a modified World Café [16] approach:

- Introduction: small group facilitators and MS MOOC team members greeted participants, and the academic lead gave a brief introduction presentation, including an agenda for the day, context and background for the discussion, rationale for developing an online course, and examples of the various modes of delivery available to the team (eg, video types, etc). This was followed by an introduction exercise and a large group discussion of ways of working to establish rules for the small group discussions. After this discussion, a catered morning tea was provided.
- Small group discussions: small group discussions followed the World Café format, except that, because of mobility issues, the participants remained at the same table throughout, and the facilitators moved between tables. Small groups discussed 4 questions, and each question was discussed for 15 to 20 min. Discussion questions are available in the supplementary materials (Multimedia Appendix 2). Discussions were followed by a catered lunch.
- Large group discussion and *dotacracy*: the small group facilitators summarized the main points of the discussions into bullet points, which were posted on a whiteboard and presented to the large group after lunch. Each participant was given a strip of 12 red dot stickers (3 per discussion question). They were told to place the stickers next to the summarized statements that they felt where most important for inclusion in the course. They were allowed to place more than one dot next to a given statement.
- Wrap-up: after voting, the participants were thanked and invited to participate in the course pilot study. The small group facilitators were asked to submit their feedback on the process within a week of the event.

The results of the focus group were analyzed quantitatively using summary statistics and qualitatively using textual analysis. Each group discussion was considered a data item (n=16). All discussion notes were transcribed and assessed using thematic

XSL•FO
RenderX

analysis. Any topic mentioned at least twice throughout the data set was considered to be a theme. Any theme mentioned by all 4 groups was considered a key theme. The data collected from the focus group were used to amend the initial syllabus and create the initial course outline, which served as the blueprint for the core course content. The results also informed the style and mode of delivery decisions for course materials.

Development of Course Materials

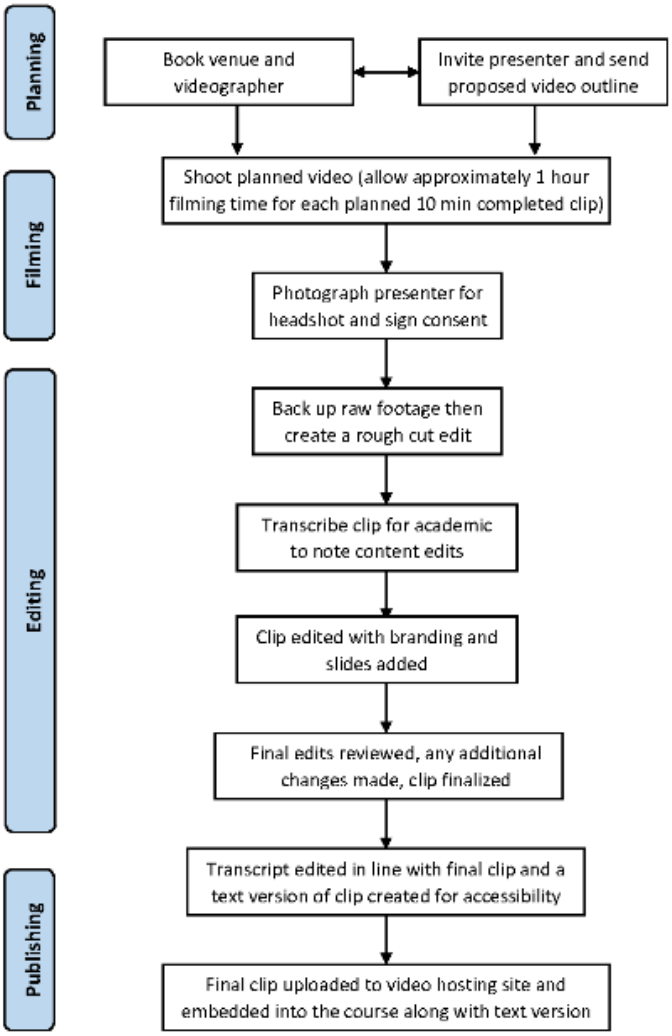
The course materials were designed in accordance with our core principles and integrated with the feedback of the focus group participants and the advisory group. The academic lead wrote the first draft of the core course content by elaborating on the course outline, breaking down larger topics into their component pieces, and giving greater detail on each one. The course was structured into 6 modules, with each module comprising several sections. The course outline was adapted into video scripts, with nearly every section containing at least one video, and was distributed to the area expert presenting the information. The area experts were given the opportunity to refine the content of the video with the academic lead before and during filming. These academic videos formed the core content of the course and served as the basis for the rest of the course material.

The academic videos were filmed in 3 formats:

1. Interview: the academic lead sat off-camera interviewing the area expert who was on camera. In the final version, tiles with the interview questions appear between the responses of the expert.
2. Conversational: The area expert and the academic lead were on the camera together, side-by-side. The academic lead interviewed the area expert.
3. Direct-to-camera: the academic lead presented a short (<5 min) scripted lecture on a topic. These videos were later animated.

After the videos were shot, they were edited and transcribed so that a text version of all videos could be made available to the course participants. Some videos were animated to illustrate the concepts that they covered. Figure 2 demonstrates the process used to develop video content.

Figure 2. Video production flowchart.

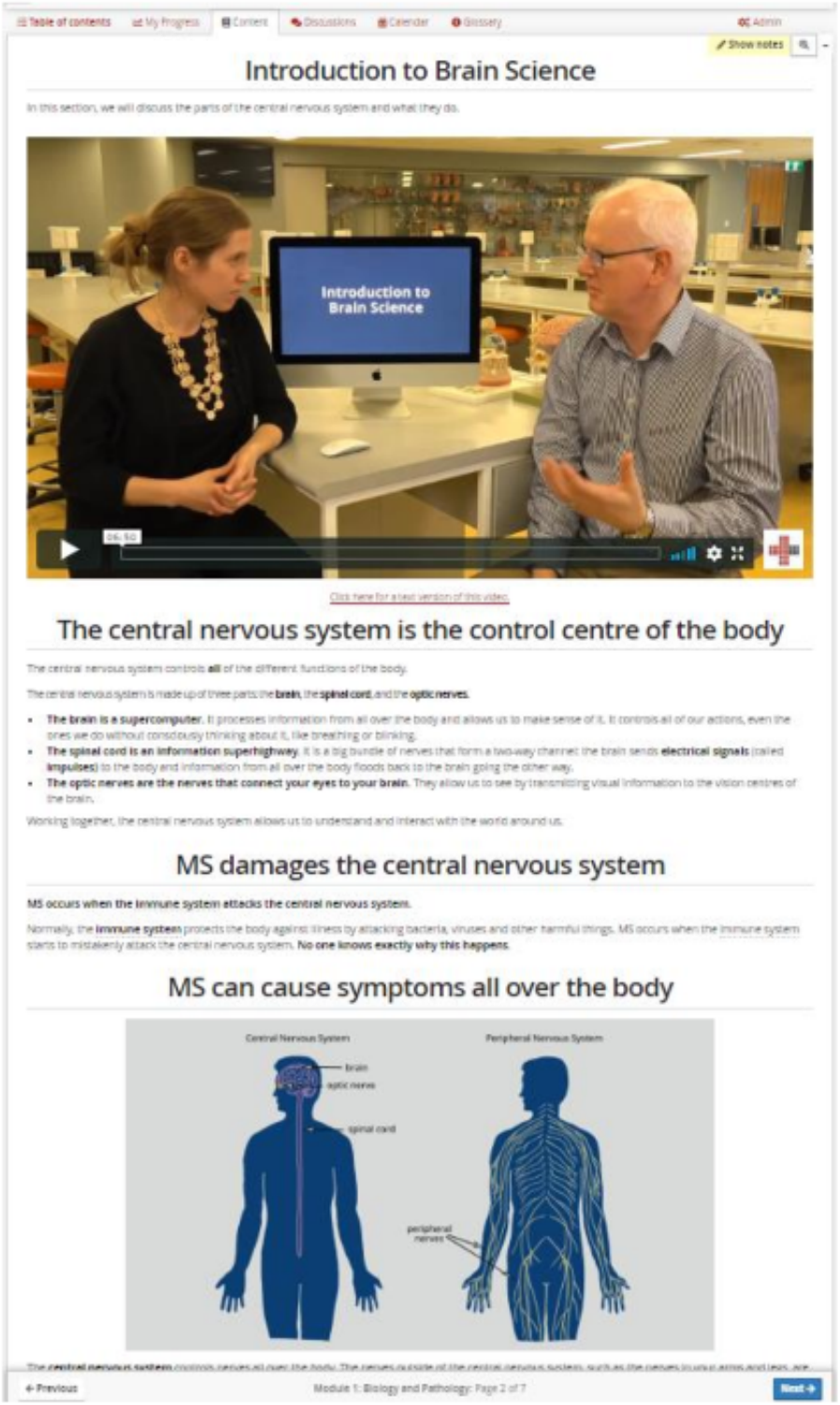


Text was written to reinforce the main messages of the videos and to add a greater context. The main messages were presented as headers, with more details given in the supporting text. The goal was to ensure that if a participant only read the headers, they would come away with the main points of the section. The text was pitched at a general audience and assumed participants

XSL•FO
RenderX

had completed a secondary school science education. Figure 3 shows a screen capture of one of the course pages.

Figure 3. Screen capture showing the layout of a single page of the course.



Activities were developed to encourage participant resilience (eg, goal setting) and develop their disease management skillset (eg, symptom tracking). Discussion prompts aimed to inspire reflection about the course content and gather feedback about the activities. End-of-module quizzes aimed to cover the main content of each module, aligning with the module learning objectives. Each quiz was made up of 10 questions, and all questions were multiple choice. Participants had to achieve a score of 70% or higher to move on to the next module, but they could take each quiz as many times as they liked. The correct answers to the quizzes were presented at the beginning of the following module.

In addition, 2 other video series were included in the course: the *community insights* series and *MS perspectives* series. The

XSL•FO
RenderX

MS perspective series included interviews with people living with MS recruited through pre-existing relationships with Multiple Sclerosis Limited and Menzies. Both were filmed in an interview format. They were each asked the same series of questions, which were related to the core content of the course. The interviews were then spliced together so that the final set of 5 videos showed the experience of multiple people living with MS on a particular subject. The *community insights* series included interviews with other members of the MS community, such as a caregiver and an MS nurse, about their experience working with and/or caring for people living with MS. These videos were placed in the course alongside the core content related to them.

Each module opened with a summary paragraph describing the content, a list of learning objectives, and a list of possible applications for the information presented in the module. The module sections usually began with a video under the section heading, followed by the supporting text. There were 1 to 2 discussion prompts and 1 activity in each module. The final section included a summary paragraph and downloadable summary PDF file, as well as a list of additional resources (mostly links to external websites) and the end-of-module quiz. An example quiz, a list of module activities and discussion prompts, and the finalized course outline are available in the supplementary materials (Multimedia Appendices 3-5). To minimize the burden on course participants, we aimed to keep the time required to complete each module under 2 hours. Due to this, the length of most of our videos was 7 min.

Pilot Study

The purpose of the pilot study was to gauge MS community members' satisfaction with the course and to test its technical components. The pilot study was advertised to MS community members via Menzies and Multiple Sclerosis Limited newsletters. People interested in taking part could register their interest on the course's landing webpage. People who indicated interest were contacted via email and invited to enroll in the pilot study, which began on January 21, 2019. Participants were given access to the course materials in the same fashion as course participants during regular enrollments, with 1 module opening each week accompanied by a reminder email.

Participants were given the option of providing feedback via a web survey at the end of each module and on the overall course. In these surveys, participants were asked to provide quantitative (eg, rating) and qualitative (eg, free text) feedback. Participants rated their satisfaction with each course component (videos, text, images and animations, discussions, quizzes, and technical aspects) on a 5-point Likert scale from very dissatisfied to very satisfied. They were also asked to provide their level of agreement with statements about each course component. The results were analyzed quantitatively using summary statistics (eg, percentage satisfied or in agreement with a given statement) and qualitatively using thematic analysis. An example evaluation survey is available in Multimedia Appendix 6. The results of this study were used to refine the course before the first public iteration, which opened on April 29, 2019.

Preparation for Course-Related Research

Along with providing community education, research is one of the primary purposes of the Understanding MS MOOC. To prepare for research projects around the first enrollment of the course, the academic lead prepared an implementation evaluation tool and an outcomes evaluation tool, led the development of the MS Knowledge Assessment Scale, and submitted ethics applications for course-related projects. These tools and related research will be described in more detail in future publications.

Results

In this section, we will assess the methodology we used to develop the Understanding MS MOOC by discussing the strengths and weaknesses of each step in the process.

Advisory Group and Massive Open Online Course Management Structure

The members of the advisory group reported that the group was effective. As the group contained a cross-section of expertise from all areas of the project (technical, academic, managerial, and clinical) and project stakeholders, as well as key decision makers, decisions were well informed and could be made swiftly. The group was inclusive and respectful of the members' time. There was an action item list for each meeting, and meetings were not held if there was nothing to discuss. The regular meeting time and buy-in from all members contributed to the group's success.

Group members reported one notable weakness: the function of the group could be improved by greater budget transparency, which would simplify decision making and purchase approvals.

Initial Course Syllabus Development

As discussed earlier, access to WDREC materials was a significant advantage when combined with our initial syllabus, as we were able to emulate the structure of successful MOOCs. The only substantial discussion around the initial syllabus was about the module order, particularly the placement of the biology and pathology module. This module is the most technical, and there was a concern that placing it first might discourage participants from completing the course. However, we concluded that placing it first made the most sense, as it provides information that underpins the following modules.

Focus Group

The small group facilitators agreed that the focus group was a success. They reported that the atmosphere was positive and that the participants enjoyed the process. The focus group had a 100% attendance rate; 21 people were invited and agreed to attend (the demographics of the attendees are provided in Table 1). This was the result of long-standing relationships with the MS flagship business manager, who acted as a community liaison. Furthermore, her initial contact with attendees set a casual and collegial tone for the focus group, which resulted in people arriving relaxed and open. This contributed significantly to the success of the event.

XSL•FO
RenderX

Table 1. Demographics of the focus group (n=21). Participants were able to select more than one multiple sclerosis (MS) community role and MS disease course.

Demographics	Values
Gender, n (%)	
Male	6 (29)
Female	15 (71)
Education level, n (%)	
Secondary school or less	5 (24)
Occupational certificate or diploma	8 (38)
Bachelor's degree	4 (19)
Postgraduate degree	4 (19)
MS[a] community roles, n (%)	
Person with MS	13 (62)
Family or friend	6 (29)
Caregiver	3 (14)
Allied health practitioner	2 (10)
Advocate	6 (29)
UTAS[b] staff member	2 (10)
Multiple Sclerosis Limited employee	1 (5)
Volunteer	1 (5)
Disease course, n (%)	
RRMS[c]	9 (82)
SPMS[d]	1 (9)
PPMS[e]	0 (0)
PRMS[f]	0 (0)
I am not sure	2 (18)
Age (years), mean (SD)	51.14 (11.33)
Disease duration, mean (SD)	13.18 (6.24)
Number of roles in the MS community, mean (SD)	1.62 (0.92)

[a]MS: multiple sclerosis.

[b]UTAS: University of Tasmania.

[c]RRMS: relapsing remitting multiple sclerosis.

[d]SPMS: secondary progressive multiple sclerosis.

[e]PPMS: primary progressive multiple sclerosis.

[f]PRMS: progressive relapsing multiple sclerosis.

Our method allowed us to identify the key themes from the small group discussions, such as accessibility, style, and exportable resources, and to compare these themes with the discussion summary statements that were most commonly voted as important by focus group participants (Table 2).

There were 2 main weaknesses in our focus group method. First, the small group facilitators reported some concern about fatigue. One facilitator suggested shortening the day by reducing the amount of time for questions at the end, whereas another suggested shortening the small group discussion periods from 15 to 10 min. Second, our selection of example videos was poor (ie, did not illustrate stylistic differences well), and we suggest that future projects choose videos more carefully to clearly show the differences in style. This might be accomplished by creating a few videos about the same subject in different styles, specifically for use in the focus group.

XSL•FO
RenderX

Table 2. The most important small group discussion summary statements (≥6 votes), as voted by focus group participants. The focus group participants were able to vote for the same statement multiple times.

Question	Votes, n
What topics/subjects would you like to see covered in an online course about MS[a]?	
Tips, management tools/activities to make living with MS easier, positive, and enhance quality of life	8
What to expect for certain scenarios (ie, lesion location)	7
Support and where to find it	7
Guides for partners, children, public personnel, establishments, and health care providers (eg, general practitioner and allied health)	6
Support services, events in local area	6
Guides for types, symptoms	6
What modes of delivery do you prefer?	
Video lecture (facing the audience)	12
Keep it simple	11
Link to other resources	6
What would make an online course about MS useful to you?	
Interview people with MS	13
Know more about MS	8
Cater for a variety of learning styles	6
What would make the course engaging for you?	
Plain English, simple language	14
Consider different learning styles, for example, visual and reading—either options or a mix of presentation styles	10
Speakers—engaging, warmth, humor, lived experience	9
Speakers must be engaging. Academics in chairs speaking to each other=boring	7

[a]MS: multiple sclerosis.

Development of Course Materials

In total, we created 53 videos, including 6 *MS perspectives videos* and 4 *community insights* videos, totaling 5 hours, 40 min, and 32 seconds of film. Only 4 used green screen animations. There were a total of 62 pages from the introduction through the completion of modules. On average, participants in the pilot study took approximately 2 hours to complete each module (Table 3), equating to 12 hours of total course content.

Our content development process had 4 main strengths. First, our core principles effectively shaped the tone of the course. We translated them into course-related materials in a variety of ways, including the following examples:

1. *Do no harm*. This influenced both the tone and the structure of the course. For example, while the content about MS symptoms, in general, was mandatory, videos detailing specific MS symptoms were made optional. This allowed participants to avoid exposure to information that might increase their symptom-related anxiety.
2. *Keep it positive*. In the feedback survey, instead of asking what was wrong with the course, we asked what could be improved.
3. *Share the journey*. Throughout the course, we choose to use the pronoun *we* instead of *you* to emphasize a sense of community. In addition, we individually greeted each participant who posted on the introduction discussion board in the first week of the open enrollments, thanking them for their participation and welcoming them to the course.

Second, having a single academic lead who oversaw all course content development was a very effective strategy, and we would recommend this approach to others. We found that it reduced duplication and contradiction and increased the continuity between sections. As all the written material (eg, text, quizzes, summaries, etc) had a single author, the course had a consistent style and pitch.

Third, we used a variety of different video styles. Not only is it more visually interesting but each video format also has particular advantages. The interview style is very flexible and allows for filming to be stopped between questions, which provides time for the academic lead and the presenter to workshop the upcoming question and response. Frequent pauses also mean that it is easy to reshoot the responses and only requires the presenter to remember their response to one question at a time. The conversational style allows for in-the-moment redirection of the presenter and follow-up questions by the academic lead. The conversational and interview styles require minimal upfront time investment by the presenters, which reduces the burden placed on them and may make them more likely to participate and enjoy participating. The direct-to-camera style allows for the concise delivery of information. However, it requires more upfront preparation,

XSL•FO
RenderX

including memorization of a script. For this reason, this style of video was only presented by the academic lead.

Fourth, strong selection of video presenters from across the MS community (eg, people living with MS, caregivers, researchers, neurologists, etc). We found that the area experts made the best presenters, as they were comfortable speaking about their area of expertise. Inviting the presenters to workshop the script ahead of shooting increased their involvement and preparation and made them more comfortable when filming. We found that filming was improved by keeping things casual. This made presenters more comfortable, which both improved their performance and made them more likely to enjoy themselves.

The largest weakness of our content development was our ambitious timetable, which allowed approximately 10 months leading up to the first open enrollment. Given the amount of content needed, this rushed the process. In addition, because of financial and logistical constraints, the course materials were only made available in English. This limited the accessibility of the course.

Table 3. Demographics of the pilot study participants who completed the feedback surveys.

Demographics	Module 1 (n=51)	Module 2 (n=44)	Module 3 (n=41)	Module 4 (n=41)	Module 5 (n=36)	Module 6 (n=39)	Overall (n=29)
Gender, n (%)							
Female	47 (92)	41 (93)	38 (93)	38 (93)	33 (92)	37 (95)	26 (90)
Male	4 (8)	3 (7)	3 (7)	3 (7)	3 (8)	2 (5)	3 (10)
Education, n (%)							
Secondary school	6 (12)	4 (9)	4 (10)	3 (7)	4 (11)	5 (13)	2 (7)
Occupational certificate or diploma	7 (14)	7 (16)	5 (12)	8 (20)	5 (14)	7 (18)	7 (24)
Undergraduate degree	16 (31)	14 (32)	14 (34)	13 (32)	10 (28)	12 (31)	8 (28)
Postgraduate degree	22 (43)	19 (43)	18 (44)	17 (42)	17 (47)	15 (39)	12 (41)
MS[a]community roles, n (%)							
Person with MS	32 (63)	29 (66)	25 (61)	27 (66)	24 (67)	26 (68)	17 (59)
Family member or friend	12 (24)	10 (23)	10 (24)	9 (22)	10 (28)	8 (21)	7 (24)
Caregiver	3 (6)	2 (5)	2 (5)	3 (7)	3 (8)	2 (5)	0 (0)
Allied health professional	4 (8)	5 (11)	3 (7)	4 (10)	2 (6)	4 (10)	4 (14)
MS nurse	2 (4)	2 (5)	2 (5)	2 (5)	2 (6)	2 (5)	2 (7)
General practitioner	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Neurologist	1 (2)	1 (2)	0 (0)	1 (2)	1 (3)	0 (0)	0 (0)
Advocate	1 (2)	1 (2)	0 (0)	1 (2)	1 (3)	1 (3)	0 (0)
Service provider	2 (4)	2 (5)	3 (7)	2 (5)	1 (3)	1 (3)	2 (7)
Researcher	4 (8)	2 (5)	2 (5)	1 (2)	0 (0)	0 (0)	0 (0)
Other	6 (12)	4 (9)	4 (10)	4 (10)	6 (17)	4 (10)	5 (17)
MS onset phenotype, n (%)							
Primary progressive	1 (3)	2 (5)	1 (2)	2 (5)	2 (6)	2 (5)	1 (3)
Progressive relapsing	1 (3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Relapsing remitting	18 (56)	16 (36)	15 (37)	15 (37)	12 (33)	14 (36)	11 (38)
Secondary progressive	9 (28)	8 (18)	8 (20)	8 (20)	8 (22)	8 (21)	3 (10)
I am not sure	3 (9)	3 (7)	1 (2)	2 (5)	2 (6)	2 (5)	2 (7)
Age (years), mean (SD)	52.1 (12.3)	51.9 (11.4)	52.8 (12.4)	53.2 (12.4)	55.3 (12.2)	52.8 (12.3)	51.9 (12.8)
MS disease duration, mean (SD)	10.7 (8.7)	11.7 (9.2)	11.4 (8.4)	10.2 (8.2)	11.5 (8.1)	11.0 (8.4)	7.8 (5.8)
Time to completion (hours), mean (SD)	1.9 (1.0)	2.1 (0.9)	1.9 (0.9)	1.6 (0.9)	2.1 (1.1)	1.7 (1.0)	2.2 (1.2)

[a]MS: multiple sclerosis.

XSL•FO
RenderX

Pilot Study

A total of 97 people enrolled in the course of the pilot study. Of those, 51 provided feedback on at least one section of the course. Across all surveys, a minimum of 90% (26/29) of the respondents were female, and a minimum of 59% (17/29) were people living with MS. The pilot study participant demographics are presented in Table 3.

The pilot study had 2 main strengths. First, the group was reasonably representative of the MS community, with a variety of MS community roles represented and a range of MS phenotypes. Second, we were able to pinpoint areas for improvement because we surveyed participants about each module and the overall course, which provided fine-grained data (shown in Multimedia Appendix 7). If we had simply asked about overall satisfaction, our ability to target underperforming areas would have been lost (compare part A of Multimedia Appendix 7 with part B of Multimedia Appendix 7).

The main weakness of the pilot study was the relative paucity of data. This was a fairly small study, and we did not make providing feedback a mandatory part of participating in the pilot. Therefore, we did not collect as much data as we could have. More data may have provided a more detailed or nuanced view of the course materials.

Discussion

Principal Findings

Overall, our process was effective and efficient. This resulted in the production of a successful health and medicine MOOC aimed at the MS community and interested laypeople with high participant satisfaction, retention, and recruitment and with no major technical errors. We believe that this project was a good value for money and that the funds were well spent. We estimate that without extensive in-kind support, the project costs would have exceeded AUD $500,000. The generosity of our project partners was essential to its success and to illustrate the community support for this work. We have identified 5 key lessons learned from our experience.

Community Support is Essential

The support that this project received from the community, UTAS, and Multiple Sclerosis Limited was critical. Without it, this project would not have been possible. UTAS' institutional support was particularly extensive. We were able to leverage the expertise, experience, and resources (particularly the custom online course platform) of the WDREC to achieve high cost-effectiveness. This was a sizable advantage; without it, the project would have been significantly more expensive, challenging, and time-consuming. Community buy-in and support also performed several other essential functions for the project. This allowed us to integrate the community from the outset. This led to collaboration, co-design, and engagement and allowed us to identify, understand, and tailor course content for our audience early on.

Stakeholder Involvement Improves Content Quality

There are many reasons to engage with stakeholders, including social justice and utilitarian motivations [17]. From a social justice perspective, there are compelling ethical reasons to involve stakeholders in the process early and often. From a utilitarian perspective, we found that early stakeholder involvement allowed us to identify the preferences of our target audience at the beginning of the process and helped to ensure that the content we created was appropriate. For example, we were surprised by the importance placed by focus group participants on the tone of the course. They indicated that they wanted a warm, good-natured tone. We incorporated this into the course content. This was a significant benefit to our project.

Plan for Research From the Beginning

We found that the best time to integrate research into an MOOC is at the beginning, while design and development are underway. This makes it possible to incorporate research tools directly into the course. It also allows sufficient time to produce tools that may be missing and require development. Early preparation also allowed the pilot study to function as a test of the course evaluation survey, which is now available for all course completers to provide feedback.

The Academic Lead and Project Management Team Ensure a Consistent Voice

A dedicated academic lead and project management team ensured that there was a consistent voice and style throughout the course content, accelerated course development, and minimized discrepancies in the course material. As this was their primary project throughout the year, they ensured that our ambitious timeline was met. They worked together to establish the textual, visual, and tonal styles of the course.

A Network of Collaborators Is a Key Resource

Our network of collaborators provided area experts from across the MS community, provided input on the course material, and, in the case of the WDREC, allowed us the immense benefit of their experience. Their generosity with their time made this project possible within the time and budget constraints presented above. Furthermore, collaborating at every level of this project has ensured that the community is involved and that stakeholders are represented, including in the authorship of this manuscript.

Strengths and Limitations

The main strength of this research was the diverse group of MS community members involved in all aspects of course development. The main limitations of this work were that it primarily involved Australian participants and was limited to those fluent in English. The web-based components further required the ability to access and use digital platforms.

Conclusions

The process behind the development of the Understanding MS MOOC was extensive, but each component contributed to a successful outcome. However, although this process was successful for us, it is important to recognize that this is a single project. We hope that by providing our material here, we will encourage others to use them in future work, allowing for meaningful comparisons in course development methodology and outcomes.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all of the people who contributed to the development of the Understanding MS MOOC. The authors would particularly like to thank the MS community members, who generously gave up their time to take part in the focus group, the course videos, and the pilot study; the course cosponsor; Multiple Sclerosis Limited; and the WDREC, whose technical expertise and substantial MOOC experience were invaluable.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial course syllabus. PwMS: person living with Multiple Sclerosis.
[DOCX File , 20 KB - jmir_v22i7e16687_app1.docx]

Multimedia Appendix 2
Focus group small group discussion questions.
[DOCX File , 14 KB - jmir_v22i7e16687_app2.docx]

Multimedia Appendix 3
Finalized course map for the first open enrolment.
[DOCX File , 314 KB - jmir_v22i7e16687_app3.docx]

Multimedia Appendix 4
Example quiz. Correct answers are bolded.
[DOCX File , 16 KB - jmir_v22i7e16687_app4.docx]

Multimedia Appendix 5
List of discussion prompts and activities from the first iteration of the Understanding Multiple Sclerosis online course.
[DOCX File , 15 KB - jmir_v22i7e16687_app5.docx]

Multimedia Appendix 6
An example pilot study survey querying participant feedback on the overall course.
[DOCX File , 6078 KB - jmir_v22i7e16687_app6.docx]

Multimedia Appendix 7
Pilot study results showing the percentage of participants who reported being satisfied or very satisfied with (A) the course sections (modules) and the course overall, and (B) each component (eg, video and text) of each course section and the course overall.
[PNG File , 42 KB - jmir_v22i7e16687_app7.png]

References

1. Liyanagunawardena TR, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014 Aug 14;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
2. Shah D. By The Numbers: MOOCs in 2019. Class Central. 2019. URL: https://www.classcentral.com/report/mooc-stats-2019/) [accessed 2020-05-29]
3. Cusumano MA. Are the costs of 'free' too high in online education? Commun ACM 2013 Apr;56(4):26-28. [doi: 10.1145/2436256.2436264]
4. Daniel J. Massive open online courses: what will be their legacy? FEMS Microbiol Lett 2016 Apr;363(8):55. [doi: 10.1093/femsle/fnw055] [Medline: 26940290]
5. Reich J. Education research. Rebooting MOOC research. Science 2015 Jan 2;347(6217):34-35. [doi: 10.1126/science.1261627] [Medline: 25554779]
6. Pickering JD, Henningsohn L, DeRuiter MC, de Jong PG, Reinders ME. Twelve tips for developing and delivering a massive open online course in medical education. Med Teach 2017 Jul;39(7):691-696. [doi: 10.1080/0142159X.2017.1322189] [Medline: 28485629]
7. Oakley BA, Sejnowski TJ. What we learned from creating one of the world's most popular MOOCs. NPJ Sci Learn 2019;4:7 [FREE Full text] [doi: 10.1038/s41539-019-0046-0] [Medline: 31240111]

XSL•FO
RenderX

8. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor B, et al. Atlas of multiple sclerosis 2013: a growing global problem with widespread inequity. Neurology 2014 Sep 9;83(11):1022-1024 [FREE Full text] [doi: 10.1212/WNL.0000000000000768] [Medline: 25200713]
9. Dobos K, Healy B, Houtchens M. Access to preventive health care in severely disabled women with multiple sclerosis. Int J MS Care 2015;17(4):200-205 [FREE Full text] [doi: 10.7224/1537-2073.2013-046] [Medline: 26300706]
10. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. JMIR Rehabil Assist Technol 2018 Apr 24;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
11. Reich J, Ruipérez-Valiente JA. The MOOC pivot. Science 2019 Jan 11;363(6423):130-131. [doi: 10.1126/science.aav7958] [Medline: 30630920]
12. Jordan K. MOOC Completion Rates: The Data. Katy Jordan. 2015. URL: http://www.katyjordan.com/MOOCproject.html [accessed 2020-05-29]
13. Shah D. Class Central's Best Online Courses of 2019. Class Central. 2020. URL: https://www.classcentral.com/report/best-free-online-courses-2019/?fbclid=IwAR0r7uq5dytkLQuiqH5135CBWf4UMwrVQDZBkSvm5DVA20L3l7muoOiXBsg [accessed 2020-05-29]
14. Health & Medicine Courses. Class Central. 2019. URL: https://www.classcentral.com/subject/health?sort=rating-up) [accessed 2020-05-29]
15. Goldberg LR, Bell E, King C, O'Mara C, McInerney F, Robinson A, et al. Relationship between participants' level of education and engagement in their completion of the understanding dementia massive open online course. BMC Med Educ 2015 Mar 26;15:60 [FREE Full text] [doi: 10.1186/s12909-015-0344-z] [Medline: 25889490]
16. Brown J, Isaacs D, World Cafe Community. The World Café: Shaping Our Futures Through Conversations that Matter. San Francisco: Berrett-Koehler Publishers; 2005.
17. Brunton G, Thomas J, O'Mara-Eves A, Jamal F, Oliver S, Kavanagh J. Narratives of community engagement: a systematic review-derived conceptual framework for public health interventions. BMC Public Health 2017 Dec 11;17(1):944 [FREE Full text] [doi: 10.1186/s12889-017-4958-4] [Medline: 29228932]

Abbreviations

Menzies: Menzies Institute for Medical Research
MOOC: massive open online course
MS: multiple sclerosis
UTAS: University of Tasmania
WDREC: Wicking Center for Dementia Research and Education

Edited by G Eysenbach; submitted 16.10.19; peer-reviewed by R Haase, G Giunti; comments to author 30.12.19; revised version received 19.02.20; accepted 10.04.20; published 29.07.20.

Please cite as:
Claflin SB, Gates R, Maher M, Taylor BV
Building a Successful Massive Open Online Course About Multiple Sclerosis: A Process Description
J Med Internet Res 2020;22(7):e16687
URL: https://www.jmir.org/2020/7/e16687
doi:10.2196/16687
PMID:32723726

©Suzi B Claflin, Rachael Gates, Maree Maher, Bruce V Taylor. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of Adaptive Feedback in a Smartphone-Based Game on Health Care Providers' Learning Gain: Randomized Controlled Trial

Timothy Tuti[1,2,3], BSc, MSc, DPhil; Niall Winters[1,3], BSc, MSc, PhD; Hilary Edgcombe[4], BSc, BMBCh, MSc; Naomi Muinga[2], BSc, MSc; Conrad Wanyama[2], BSc; Mike English[2,5], MD, MBBCHIR, FMedSci; Chris Paton[5], BMBS, MBA

[1]Kellogg College, University of Oxford, Oxford, United Kingdom

[2]KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya

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

[4]Nuffield Division of Anaesthetics, University of Oxford, Oxford, United Kingdom

[5]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Timothy Tuti, BSc, MSc, DPhil
Kellogg College
University of Oxford
60 Banbury Road
Oxford, OX2 6PN
United Kingdom
Phone: 44 7501966958
Email: timothy.tuti@kellogg.ox.ac.uk

Abstract

Background: Although smartphone-based emergency care training is more affordable than traditional avenues of training, it is still in its infancy, remains poorly implemented, and its current implementation modes tend to be invariant to the evolving learning needs of the intended users. In resource-limited settings, the use of such platforms coupled with gamified approaches remains largely unexplored, despite the lack of traditional training opportunities, and high mortality rates in these settings.

Objective: The primary aim of this randomized experiment is to determine the effectiveness of offering adaptive versus standard feedback, on the learning gains of clinicians, through the use of a smartphone-based game that assessed their management of a simulated medical emergency. A secondary aim is to examine the effects of learner characteristics and learning spacing with repeated use of the game on the secondary outcome of individualized normalized learning gain.

Methods: The experiment is aimed at clinicians who provide bedside neonatal care in low-income settings. Data were captured through an Android app installed on the study participants' personal phones. The intervention, which was based on successful attempts at a learning task, included adaptive feedback provided within the app to the experimental arm, whereas the control arm received standardized feedback. The primary end point was completion of the second learning session. Of the 572 participants enrolled between February 2019 and July 2019, 247 (43.2%) reached the primary end point. The primary outcome was standardized relative change in learning gains between the study arms as measured by the Morris G effect size. The secondary outcomes were the participants individualized normalized learning gains.

Results: The effect of adaptive feedback on care providers' learning gain was found to be g=0.09 (95% CI −0.31 to 0.46; *P*=.47). In exploratory analysis, using normalized learning gains, when subject-treatment interaction and differential time effect was controlled for, this effect increased significantly to 0.644 (95% CI 0.35 to 0.94; *P*<.001) with immediate repetition, which is a moderate learning effect, but reduced significantly by 0.28 after a week. The overall learning change from the app use in both arms was large and may have obscured a direct effect of feedback.

Conclusions: There is a considerable learning gain between the first two rounds of learning with both forms of feedback and a small added benefit of adaptive feedback after controlling for learner differences. We suggest that linking the adaptive feedback provided to care providers to how they space their repeat learning session(s) may yield higher learning gains. Future work might

XSL•FO
RenderX

explore in more depth the feedback content, in particular whether or not explanatory feedback (why answers were wrong) enhances learning more than reflective feedback (information about what the right answers are).

Trial Registration: Pan African Clinical Trial Registry (PACTR) 201901783811130; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=5836

International Registered Report Identifier (IRRID): RR2-10.2196/13034

(J Med Internet Res 2020;22(7):e17100) doi:10.2196/17100

KEYWORDS

neonatal mortality; education; emergency medical services; global health; smartphone; feedback; health workforce; developing countries; mobile phone

Introduction

Background

In low-income regions such as sub-Saharan Africa (SSA), the need for health workers to provide care by themselves is more frequent than in middle- and high-income settings, and it can be associated with negative neonatal outcomes [1]. Of the estimated 2.9 million neonatal deaths each year globally, SSA has the highest overall risk of death within the first 24 hours of life, accounting for 37% of global neonatal deaths [1]. Severe workforce shortages, coupled with the skill imbalance and maldistribution of the health workforce, and a lack of training opportunities [1,2] are likely to be key contributors to this high mortality rate. Additional training is needed to better prepare health care providers in these regions to provide effective emergency pediatric and neonatal care [3,4]. However, face-to-face training costs between US $80 and US $300 per person per day and is difficult to deliver at scale [5]. Only a small fraction of trained health providers with the basic requisite skills training for new-born resuscitation are usually found in these regions [6,7]. Therefore, new strategies are required to improve training access for over 1 million health providers across SSA. Any such approach needs to be updated efficiently in real-time as guidelines change in light of new evidence and (ideally) capture data on the number of health workers that are able to train within a certain time period [8,9].

There is little evidence about the implementation of digital learning interventions that are relevant to the context of low-income settings, that take into account health workers' initial and continuing clinical training needs, and that adapt learning content in the light of skill mastery and performance as learners continue to develop knowledge [9-15]. In nonclinical, high-resource contexts, adaptive instructional support has been shown to significantly outperform trainer-led large-group instruction, nonadaptive computer-based instruction, and paper-based instruction in producing learning gains [4,10,16,17]. Within low-resource contexts, investigation of learner models (cognitive models that try to model observed student learning behaviors) needed to support such tailored instructional approaches in clinical training settings is justified [10,11].

Health care training apps that have been developed to date and the approaches used can broadly be divided into two categories: those that simply replicate existing teaching strategies "on a screen," for example, by providing questions and answers for exam practice or displaying textbook graphics, that is, the "drill-and-practice" pedagogical approach, and others, that take advantage of features specific to digital devices, examples of which include the ability to respond with different pathways to user choices or the use of animations with which the user can interact, that is, a more learner-centered pedagogical approach [9]. Serious games, which are games with a specific, applied purpose (other than entertainment) that can be played on mobile phones, are one such way of providing training with the potential to affect health outcomes [18]. The rationale for using serious games is that, in a similar way to "first-person" computer games, emergency care training should enable health workers to follow highly structured pathways (such as clinical care algorithms) with pieces of information (cues) sought at each step that determine the correct actions to perform. With both clinical training and performance in computer games, executing cue-response sequences perfectly, rapidly, and automatically, with minimal help (eg, from corrective feedback and hints), demonstrates mastery. This type of mastery has been shown to support effective clinical care delivery, but the required frequency of rehearsal in this approach is difficult and expensive to maintain for face-to-face training [19]. By using a serious gaming approach, users may be more motivated to repeatedly play the serious game, using incentives such as rewards, increasing difficulty, and scores—techniques that have been successfully used to encourage repeated gameplay in nonserious computer games. There is a scarcity of evidence from the evaluation and assessment of serious gaming approaches in smartphones for health care training in low-income settings. Addressing this need was highlighted in the most recent systematic review in this area [15,20].

The Life-Saving Instruction for Emergencies (LIFE) project uses a smartphone-based serious game approach initially to provide training in the care of very sick newborns and children. The app extends the scenario-based Emergency Triage, Assessment, and Treatment Plus admission care training (ETAT+) training model [21,22] by incorporating more learner-centric intervention approaches. The aim of ETAT+ is to familiarize health care providers with clinical guidelines and the necessary knowledge and skills for triaging all sick children when they arrive at a health facility into those with emergency signs, with priority signs, or nonurgent cases and provide emergency treatment for those with life-threatening conditions [23]. The ETAT+ teaching model uses a face-to-face model to train health care workers in Africa and Southeast Asia and is explained in detail elsewhere [21-25].

XSL•FO
RenderX

Given the contextual challenges of enhancing health care providers' learning settings, such as SSA, LIFE is designed to develop health care providers' self-regulating learning (SRL) of ETAT+ content independent of any classroom or face-to-face tutoring facilitation. The potential for utilizing the digital and reusable nature of interventions such as LIFE to adapt the way health care professionals learn and receive feedback on their performance remains underexplored in SSA. Such personalization of learning could be used to maximize learning outcomes and to develop learners' skills [14,15]. The ubiquitous nature of smartphones as experimental tools offers access to a wider pool of study participants [26] and can minimize the cost of implementing, evaluating, and scaling educational platforms such as LIFE in a resource-constrained context [9]. Smartphones have also been shown to raise learners' interest in learning interventions [27].

SRL is a weakly recursively process that is facilitated by feedback at each stage [28-30] supporting learners' metacognitive regulation of learning strategies to help them regulate resources and emotions while learning [29]. In addition to regulation, metacognition consists of declarative knowledge and deals with the interplay between knowledge of one's abilities to perform tasks, the nature of the learning task, and the strategies one can employ to successfully perform the task [29]. It is theorized to have a limited capacity that renders the learners prone to making errors in complex or time-limited learning tasks (eg, delivery of emergency care in newborns) [31]. In the presence of such performance errors (eg, slipping or guessing) due to cognition's limited capacity [32], to guide SRL, a key objective is to infer the knowledge of the learning task being tutored. Feedback is posited to enhance learning when "...it provides further information to correct or modify action through the construction and activation of more appropriate [action sequences] ..." [33]. The information provided ought to move the learner to a deeper understanding of the learning task [34]. However, more feedback does not always equate to better learning: "...The amount of information given to the student must be what the student can use, rather than the amount the [tutor] may wish to give...," especially in light of the limited capacity of cognition [32,34-36]. Feedback that is too elaborate is more likely to produce cognitive overload. On the other hand, if it has low specificity, it is more likely to be perceived by learners as useless [32,37]. Effective feedback is posited to be specific but not too elaborate and presented in manageable units [37,38]. Its timing, specificity, frequency, and type have varying effects in enhancing learning [36]. In the absence of flexibility in determining the instruction challenge level or stratifying learning pathways within the LIFE project (due to the efforts to standardize delivery of clinical guidelines training content), feedback remains the most promising intervenable theory-aligned concept for enhancing SRL using smartphones in settings such as SSA.

Objectives

The primary objective of this randomized experiment was to investigate whether adaptive individualized feedback is superior to standardized feedback in mobile smartphone-based emergency neonatal care training. We hypothesized that health care providers randomized to receive adaptive feedback would have significantly higher learning gains than those randomized to receive standardized feedback. The secondary objective was to investigate learning gain in general and how learning gains when using LIFE are modified by health care providers' characteristics and how they space their learning.

Methods

Ethical Approval

The breakdown of the methods and analysis plan for this experiment are described in the published protocol for this study [39] and was approved by the Kenya Medical Research Institute's (KEMRI) Scientific and Ethical Review Committee (#3444) and the Central University Research and Ethics Committee of Oxford University (#ED-CIA-18-106). It follows the Consolidated Standards of Reporting Trials guidelines for reporting randomized experiments [40].

Study Design

The study was a parallel-group, double-blinded, randomized experimental design with an allocation ratio of 1:1. The participants were randomized to receive the intervention or to be in the control group when they launched the training app for the first time on their individual smartphone devices.

Eligibility Criteria

The participants were health care providers from any professional cadre involved in providing bedside patient care, who were either undergoing training (eg, students), or actively providing nursing, clinical or medical care. Health care providers who had retired from clinical practice, who practiced in high-income settings, or participants who were not health care providers were excluded from the study.

Study Setting and Recruitment

This study was confined to participants from low-income countries who stand to benefit from training in the management of pediatric emergencies (Multimedia Appendix 1). The distribution of the LIFE smartphone app was through the Google Play Store, with initial efforts directed toward face-to-face recruitment of participants in Kenya (more details are provided in the study protocol [39]). The recruitment of study participants endeavored to promote voluntary self-enrolment and used snowballing and purposive sampling strategy [41], which have been explained in detail in the study protocol [39].

The Intervention

The intervention in this study was the adaptive differentiated immediate feedback provided while learning through a smartphone-based serious gaming app. The content to be learnt was based on a neonatal resuscitation guideline course that is already offered in nine low-income countries [21,22,24]. The smartphone-based app was publicly available on the Google Play Store, where it was downloadable and installable to compatible Android-based smartphones. All study participants received a link to the mobile app hosted on the Google Play Store. The LIFE app was designed to target Android's SDK19 as the minimum version of Android supported as at February 2019 (which targets 100% of Android devices) [42]. The

XSL•FO
RenderX

smartphone app had already undergone alpha and beta testing on a pilot cohort of health care providers' smartphones from Kenya since October 2017 to ensure stability and reliability of its functioning on different mobile phones. More details of the intervention are provided in the Multimedia Appendix 2 [43-45].

The number of standardized feedback levels was determined by global health academics and expert medical trainers involved in ETAT+ training in SSA in consultations with the relevant medical professional bodies accrediting continuous professional development. The adaptive immediate feedback provided to the experimental group participants was designed to arouse meaningful reflective learning from continuous interaction between the learners and the smartphone-based training [46,47]. As didactic-procedural form of feedback [48] aligned to our theoretical framing, it was designed to force the health care provider to contemplate over the incorrect care provision choices they provided in their failed attempt (eg, "*Some* of the selected actions are not *appropriate* at this *stage*" focuses on the number of wrong choices and their placement within the clinical care-giving pathway and is meant to force reflection as to which stage they are most appropriate). This feedback was provided to the experiment arm after each incorrect attempt at a learning task with three cascading detail levels based on the predicted probability that the learner's next attempt was going to be correct. The modeling approach to support the data collected from 187 health care providers during the beta testing phase and is briefly described in the Multimedia Appendix 2 and explained in detail elsewhere [49]. The wording of the feedback provided was dependent on the number of incorrect choices the learner had selected and the actual incorrect choices themselves. This is illustrated in Multimedia Appendices 3-6. The control group study participants received standardized nonadapted immediate feedback after each incorrect attempt at a learning task, with the feedback on the first incorrect attempt asking the learner to retry (level 0, Multimedia Appendix 6) and the feedback on the second attempt giving a detailed explanation of the correct choices to select (level 2, Multimedia Appendix 6). Using multidimensional model of personalization by Holmes et al [50], in this experiment, the adaptive mechanism targets the personalization of how feedback is to be presented and when it is to be introduced.

The LIFE app was both the learning and the measurement tool. The learning task is synonymous with the quiz as learning is designed to be embedded in the formative evaluation. At the end of a successful completion of a learning session, the platform provided a performance score based on whether each learner's response to the learning tasks was correct on the first attempt. For this study, a learning session was conceptualized as every unique initiation (ie, iteration) of the neonatal resuscitation learning scenario training round on the LIFE smartphone app (illustrated in Scenario A in Multimedia Appendix 7).

Outcomes

The primary end point for both arms of the experiment was the completion of two learning sessions using LIFE, the first session being treated as pretest and the second session being treated as posttest. Both scores were converted into percentages. From the pre-post scores, the study's main comparative outcome was the learning effect size (g) [51], with the formula for its calculation provided in this study's protocol [39] and provided in the Multimedia Appendix 2. This effect size, also referred to as *Morris G* [51], represents the mean difference between the relative change within the study arms. From education literature [16,52,53], effect sizes of approximately 0.2, 0.5, and 0.8 are considered as small, moderate, and large effect sizes, respectively. These thresholds represent the magnitude of the effect and reflect our assumption that a statistically significant result is not necessarily important or meaningful. For example, for an effect size of approximately 0.2, the difference between the study groups is trivial even if it is statistically significant [52,54].

In addition to randomization, which eliminates or at least dramatically reduces biases influencing this study's primary outcome, the calculation of this study's primary outcome (1) is robust in managing preexisting knowledge differences among learners, (2) allows for the estimation of the intervention effect even when experimental and control groups are nonequivalent, and (3) considers the variances of both pretest and posttest scores. This contrasts with other forms of effect size calculation such as Hedges G and Becker D, which only use pretest or pooled variances [51]. In this model, the pretest and posttest variances were assumed to be homogeneous. The secondary outcome considered was the individualized learning gain of the study participants, defined as the relative change in performance score of health care providers divided by the maximum score they could have improved upon [55]. This was calculated from the performance scores from learning sessions following each other, with the performance from the first session treated as pretest score and the one from the last session treated as a posttest score. The formula for this calculation is provided in this study's protocol [39].

Participant Timeline

Enrolment of study participants began on February 1, 2019, and continued up to July 31, 2019. This study's rollout of LIFE's intervention was based on implementation study principles and outcomes [56] and was informed by self-regulated, self-directed learning [28,31,57]. It sought to understand and work within real-world conditions, rather than trying to control for adoption, acceptability, coverage, and sustainability conditions or to remove their influence on the study outcome [56]. Subsequently, no training sessions were planned for the study participants. Although LIFE is designed for low-income contexts, there was no limit set by geographical coverage for health care providers who might be interested in undertaking this self-directed training; anyone could download the app. However, we only analyzed health care providers from low-income countries. Participants without any geographic location data (due to refusing to grant the LIFE app the required Android permissions) were assumed to be from developing countries, given that our recruitment efforts were directed toward professional groups in these countries.

Sample Size Calculation

Similar interventions in other subject domains have been found to have a mean effect size of between 0.22 (95% CI 0.16 to

XSL•FO
RenderX

0.27) to 0.70 (95% CI −0.08 to 1.49) [16,58]. Drawing from these studies, to detect an effect size of 0.22 with a two-sided 5% significance level and a power of 80%, a sample size of 83 participants per group who reach the primary end point of the study was necessary. A sample size calculation for a one-way analysis of variance, together with one-sample and paired-sample *t* test analysis using the same effect, power, and significance parameters produced the same required sample size of 166 participants. The sample size calculation formula is provided in this study's protocol [39].

On the basis of the alpha tests of the LIFE smartphone app, we assumed a 50% dropout rate of study participants, with dropout defined as the incomplete or single use of the LIFE smartphone app, and planned to recruit at least 332 participants to account for this dropout rate. To encourage repeated usage of LIFE, all participants received email reminders from the time they were enrolled in the study once every 2 weeks, and this was suspended after they had received three reminders. From interviews of the study participants in the alpha and beta phases of the LIFE app development, those from lower clinical cadres tended to characterize the cost of phone data charge necessary for downloading the 231 MB smartphone app as too high. To mitigate the burden of saddling participants with this extra cost to their personal finances due to participating in the study, they could request for reimbursement of costs within a few weeks of using the LIFE smartphone app.

Demographic data was collected within the app in its initial use by study participants at the end of the first learning session if they consented. This was because demographic data were deemed more sensitive than trace play data and therefore required additional consent as per the ethical approval. This meant that for learners who dropped out before completing the first session, or chose not to fill-in those data, no demographic data were collected. From these data, statistical analyses were conducted to evaluate whether there was any systematic bias in the attrition of study participants. Study variables used in this analysis were study groups, exposure to previous neonatal training, clinical cadre, age, and level of experience. This was because given the differences in training pathways for clinical cadres and length of practice, level of expertise might produce differences in the effectiveness of the learning intervention [59]. In addition, age was included to evaluate whether it was associated with the pattern and effectiveness of smartphone-based learning given its novel nature requiring digital acuity [60]. We judged sex not to be theoretically influenced by the socio-cognitive framing of this experiment's research questions.

Randomization

For allocation of the participants, an in-app algorithm randomly generated a value between 0 and 3 when the smartphone app was launched for the first time. If the value was either 0 or 1, it was recoded to 0, otherwise recoded as 1. This algorithm was implemented using a randomization routine provided by the game engine for development reasons, which we assumed to be reliable [61]. The algorithm determined whether a participant was allocated to the control (if the recoded value was 0) or the experimental group (if the recoded value was 1). It also blinded both the study participants and staff to the allocation of participants to groups during the experiment, but not at the analysis stage. Sequence generation for random allocation was a computerized procedure pegged on a single instance (ie, smartphone app installation instance) that mimicked a coin-tossing procedure. Therefore, using permuted blocks of random sizes to assign participants to either the control or experimental group was not possible and therefore not implemented.

Statistical Methods and Analyses

For the primary end point, we used the Morris G effect size to analyze the differences between study arms of relative change in scores within the arm, as described in detail in the study protocol [39]. This was assessed after the second round of completing the training scenario through the smartphone app, with each round's performance score recorded. Secondary analysis was conducted using regression analysis, with the dependent variable being the normalized learning gain and the independent variables being health care providers' demographic characteristics and the game play characteristics (eg, spacing of repeat learning session, amount of time spent on learning task, and previous exposure to neonatal training), to evaluate their effect on learning gains.

The primary learning outcome used in this study could not be computed for study participants whose dropout was characterized by a lack of at least two complete learning sessions. Without a postbaseline assessment, "intention-to-treat" analysis could not be performed unless we imputed outcomes, which tends to produce biased estimates [62]. Therefore, we did not conduct an intention-to-treat analysis. However, dropout numbers are reported in relation to those who reached the study's primary end point, with their implications discussed considering self-regulated, self-directed learning [28,31,57].

Qualitative interviews were conducted in parallel to the experiment from a small sample (N=19) of the health care providers who participated in the study, regardless of whether they reached the primary end point or not. These interviews were used to explore health care providers perceptions of self-regulated learning that affected the contextual use of the smartphone-based learning platform. They provided a context for interpreting the observed learning outcomes from this study and will be reported separately.

Data Management

The primary data collected from the study participants' Android smartphone app are held on their devices with a back-up copy synchronized to *Google Firebase*, a secure distributed cloud-based database server, after being transmitted in an encrypted format. Data collected after the experiment were stored on encrypted password-protected USB devices and transferred to secure password-protected servers in Kenya and Oxford. Deidentified data were shared with study staff based at the research institutions listed on the ethical approval forms.

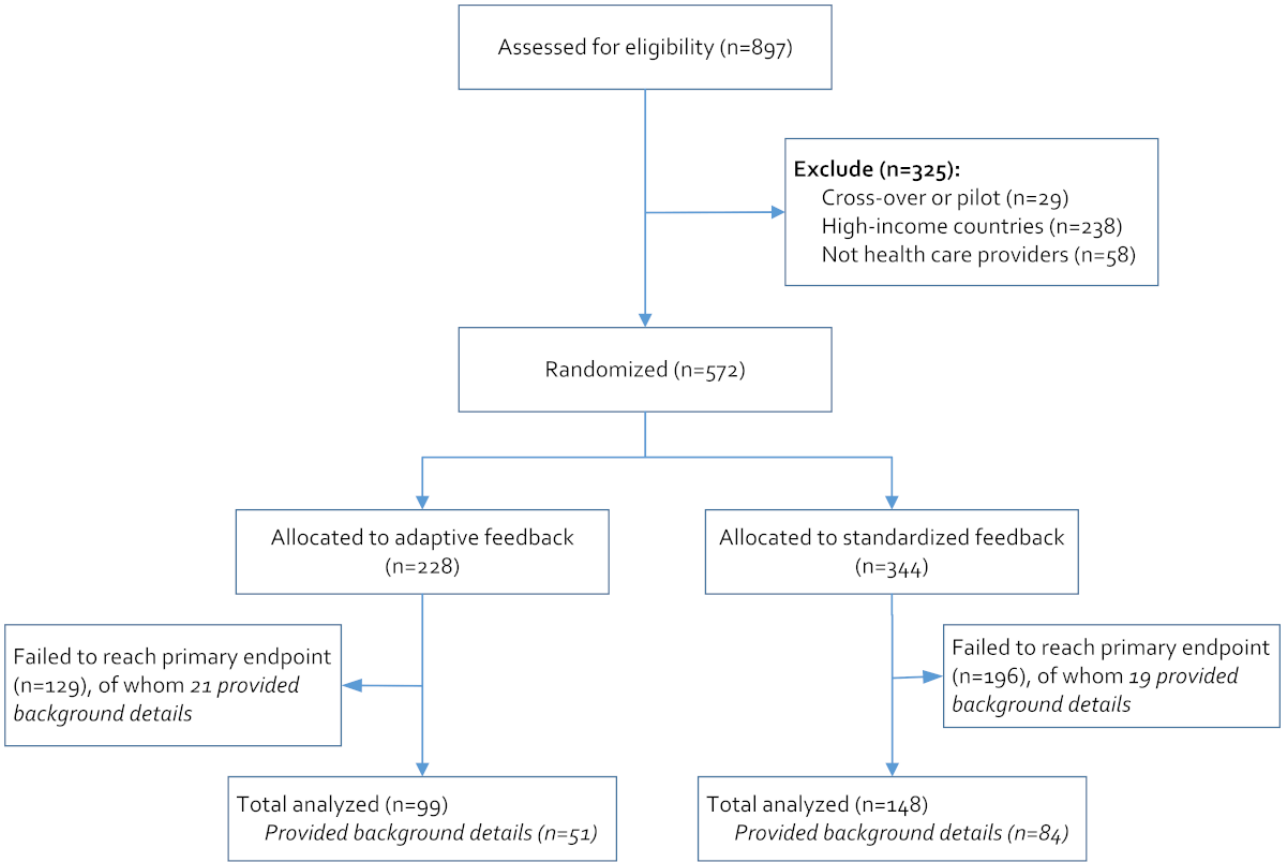
Informed Consent

Individual participant consent was elicited from within the mobile app before collection of any demographic data, in

XSL•FO
RenderX

addition to using explicit Android permission requests for collecting trace anonymized learning data from the app. This approach of informed consent in an app is not uncommon in medical research; it has been described in detail in a systematic review [63] as well as specifically for mobile app–based research [64]. The informed consent process also included information on the credentials and affiliations of the researchers and developers of the LIFE platform. Participants had no way of knowing whether they were receiving the "intervention of interest."

Figure 1. Experiment's enrolment flowchart.



Results

Study Sample

At the end of the data collection period, 572 of the enrolled 897 participants were eligible for the study and 247 (43.2%) ended up reaching the study's primary end point (Figure 1), as expected in our published protocol [39].

It was challenging to enforce the allocation ratio because the algorithm used for randomization was not centralized and therefore unable to implement a blocking mechanism in participant to group allocation. Randomization was happening on the individual health care provider's personal smartphone independent of how other participants were allocated and is illustrated in detail in Multimedia Appendix 8. In addition to the lack of an enforceable blocking mechanism, the software library used for the random generator function (C++ rand function) is most likely to have caused an unequal allocation ratio because it has varied efficiency across multiple smartphone devices that remain largely unknown (Multimedia Appendix 9) [61,65]. It is highly likely that this was what contributed to the observed disparity in allocation ratio. However, this did not render the allocation of study participants to the study arms nonrandom, as detailed below.

A total of 30.6% (175/572) of the enrolled participants provided their background information, and of these, 135 reached the experiment's primary end point (Figure 1). From the difference in means test and difference in proportions test, there was no significant difference between the experimental and control groups of both those who dropped out and those who reached the study's end point by LIFE use characteristics (Multimedia Appendix 10) or demographic characteristics (Table 1). However, the dropout rate was significant in health care providers in clinical officer cadres and those who had specialized training (Multimedia Appendix 11).

XSL•FO
RenderX

Table 1. Baseline characteristics of study participants where demographic data are available.

Indicator	All health care providers who provided background information (n=175)			Health care providers who reached the study's primary end point (n=135)		
	Control (n=103)	Experiment (n=72)	*P* value[a]	Control (n=84)	Experiment (n=51)	*P* value[b]
Age (years), mean (SD)	31.2 (8.26)	29.4 (6.99)	.12	30.81 (8.42)	29.18 (7.24)	.24
Experience (years), mean (SD)	7.21 (7.05)	6.44 (7.81)	.52	6.83 (6.66)	6.46 (8.33)	.79
Sessions started, mean (SD)	3.92 (3.02)	4.22 (3.37)	.55	4.32 (3.00)	5.35 (3.39)	.08
Clinical cadre, n (%)						
Doctor	37 (35.9)	30 (42)	.33	31 (37)	22 (43)	.24
Clinical officer	17 (16.5)	10 (14)	.26	14 (17)	5 (10)	.87
Nurse	40 (38.8)	28 (39)	.82	31 (37)	21 (41)	.31
Other	9 (8.7)	4 (5)	.53	8 (9)	3 (6)	.77
Clinical training level (whether completed general/specialty training and professionally registered), n (%)						
Specialized	23 (22.3)	11 (15)	.42	19 (23)	6 (12)	.94
General officer	38 (36.9)	29 (40)	.72	29 (35)	21 (41)	.22
Intern	9 (8.7)	7 (10)	.53	8 (9)	6 (12)	.34
Student	33 (32.0)	25 (35)	.32	28 (33)	18 (35)	.41
Done Emergency Triage, Assessment, and Treatment Plus admission care training?, n (%)	60 (58.3)	47 (65)	.28	49 (58)	33 (65)	.23

[a]*P* value from difference in means between study arms within progression category.

[b]*P* value from a difference in proportions test between study arms within progression category.

The effects of attrition bias appear to be minimal, as reported in Table 1 and Multimedia Appendix 10. There is good reason to infer that the randomization generally achieved its purposes of balancing participants from different backgrounds between the intervention arms while also mitigating against selection bias despite the unexpected allocation ratio.

Effect of Adaptive Feedback on Learning Gains Across Study Arms

The performance of the learners showed a substantive improvement of about 30% above pretest scores in both the control and experimental arms (Table 2). As the overall learning change from LIFE use was large, this may obscure a direct effect from the adaptive feedback. From Levene's test, the assumption that the variance of the performance scores was homogenous holds true [66].

Table 2. Performance of learners within study arms.

Study group	N	Pretest (%), mean (SD)	Posttest (%), mean (SD)	Life-Saving Instructions for Emergencies smartphone app effect[a]	
				Correlation[b]	Effect size (95% CI)
Control group	148	56.1 (23.2)	85.8 (15.6)	0.226	1.031 (0.789-1.274)
Experiment group	99	50.4 (21.4)	81.8 (17.7)	0.335	1.272 (0.966-1.577)

[a]Effect size based on intra-individual changes in test scores accounting for correlation between the scores in a single-group pretest posttest design [51].

[b]Pearson R correlation between the scores within the study arm.

The main outcome of interest, which was the effect of adaptive feedback on learning gain, was calculated after the second round of the completed learning scenario, and its calculation is explained in detail elsewhere [39]. Adaptive feedback had an intervention effect of 0.09 (95% CI −0.31 to 0.46; *P*=.47), which is both statistically and substantively insignificant. It is noteworthy that the calculation for the effect size computed here does not treat the scores as repeated measures but as independent data. It also assumes that pretest and posttest variances are homogeneous. To minimize this bias, a correction adjustment factor of 0.998 was applied to the effect size and is explained in the protocol [39,51]. This suggests that the degree of bias in the effect size calculation was minimal given that it is a very small adjustment. The number of feedback messages provided in the experiment over the learning sessions is illustrated in Table 3.

XSL•FO
RenderX

Table 3. Number of feedback messages provided.

Study arm	Minimal, n (%)	Reflective, n (%)	Detailed, n (%)	All, n (%)
Feedback messages for the 572 learners who were enrolled into the study				
Control	2067 (27.96)	0 (0.00)	2348 (31.76)	4415 (59.71)
Experiment	269 (3.64)	1097 (14.84)	1613 (21.82)	2979 (40.29)
All	2336 (31.59)	1097 (14.84)	3961 (53.57)	7394 (100.00)
Feedback messages[a] for the 247 learners who reached the study's end point				
Control	899 (29.4)	0 (0.00)	955 (31.23)	1854 (60.63)
Experiment	48 (1.57)	543 (17.76)	613 (20.05)	1204 (39.37)
All	947 (30.97)	543 (17.76)	1568 (51.28)	3058 (100.00)

[a]Counts are only for the first two complete learning sessions.

Effect of Adaptive Feedback on Individual Learning Gains

The primary outcome analysis results are likely to suffer from two forms of detection bias: subject-treatment interaction bias and differential time-effect bias [67], as illustrated in Multimedia Appendix 12. To mitigate against these biases and explore how LIFE use and learner background characteristics might affect learning gains, we conducted a secondary analysis using normalized learning gain at the individual level. This was consistent with the published protocol [39]. Details of our calculation of the normalized gain are explained elsewhere [39,55].

When considering differences in LIFE use at the individual health care provider level, learners who had a space of more than a week between subsequent use of LIFE had significantly lower normalized gains by −0.395 (95% CI −0.557 to −0.232; *P*<.001) than those who had spaced their learning to an hour or less. Any variation in spacing the self-directed use of LIFE, which was a week or less, did not produce significant changes to the normalized gains. The adaptive feedback mechanism had a significant effect on health care providers' normalized learning gains of 0.523 (95% CI 0.345-0.702; *P*<.001). Longer time spent on learning tasks and ratio of feedback hints provided per attempt at a learning task were significantly associated with lower normalized student learning gains. Health care providers' previous face-to-face training in ETAT+ content had no positive significant effect on their learning gain. This is illustrated in Table 4, model A.

When considering the demographic and background characteristics of the health care providers, their clinical cadre and level of practice/training had no significant effect on their learning gains, except for doctors whose learning gains were significantly higher by 0.14 (95% CI 0.016-0.265; *P*=.027). This is illustrated in model B of Table 4. Controlling for health care providers' background characteristics significantly increased the effect of adaptive feedback on individualized normalized learning gains to 0.644 (95% CI 0.347-0.941; *P*<.001). It also improved the proportion of variance for average student normalized gain, which was explained by the independent variables in the regression model A by 18.3% (Table 4). Overall, independent variables had low multicollinearity in both variants of the regression model in Table 4, as illustrated in Multimedia Appendix 13. Both models explained 34.4%-40.7% of the variance in the normalized learning gains of the health care providers using the LIFE smartphone app to train on neonatal emergency care. However, only model B in Table 4 satisfied all the statistical modeling assumptions.

A fraction of the learners had more than two rounds of play, as illustrated in Multimedia Appendices 14-16. However, with each round of play, the numbers of health care providers dropped by around 40%-60%. This meant that fitting a longitudinal model was not feasible because the variances of more than one linear combination of time effects were close to zero (ie, singular), indicating that the model would be overfitting [68]. Furthermore, health care providers' spacing of their learning was not standardized, and the learning iteration variable violated the sphericity assumption necessary for conducting a repeated measures analysis of variance test [69].

XSL•FO
RenderX

Table 4. Learning intervention effect on individual health care providers' normalized learning gains.

Indicator	Model A[a]: all learners (n=247)			Model B[b]: learners with demographic information (n=135)		
	β (SE)	*P* value	95% CI	β (SE)	*P* value	95% CI
Intercept	.79 (0.042)	<.001	0.707-0.872	0.851 (0.088)	<.001	0.677-1.026
Reference: Spacing ≤ 1 hour						
Spacing ≤1 day	.027 (0.054)	.61	−0.078 to 0.133	−0.045 (0.072)	.53	−0.188 to 0.098
Spacing ≤1 week	−0.142 (0.078)	.07	−0.294 to 0.011	−0.28 (0.097)	.005	−0.472 to −0.088
Spacing ≤1 month	−0.395 (0.082)	<.001	−0.557 to −0.232	−0.445 (0.129)	<.001	−0.7 to −0.19
Reference: Group=control						
Group=experiment	.523 (0.091)	<.001	0.345-0.702	.644 (0.15)	<.001	0.347-0.941
Time spent on learning task	−0.09 (0.023)	<.001	−0.135 to −0.046	−0.036 (0.038)	.35	−0.11 to 0.039
Help ratio[c]	−0.826 (0.133)	<.001	−1.087 to −0.565	−1.116 (0.219)	<.001	−1.549 to −0.683
Reference: Done Emergency Triage, Assessment, and Treatment Plus admission care training before=no						
Done Emergency Triage, Assessment, and Treatment Plus admission care training before=yes	−0.013 (0.04)	.75	−0.092 to 0.066	−0.092 (0.056)	.11	−0.204 to 0.02
Reference: Cadre=nurse						
Cadre=clinical officer	__[d]	—	—	0.01 (0.085)	.90	−0.15 to 0.179
Cadre=doctor	—	—	—	0.14 (0.063)	.03	0.016-0.265
Cadre=other	—	—	—	0.124 (0.107)	.25	−0.088 to 0.336
Reference: Level=student						
Level=intern	—	—	—	0.007 (0.096)	.95	−0.184 to 0.197
Level=general officer	—	—	—	−0.001 (0.07)	.99	−0.139 to 0.137
Level=specialized	—	—	—	−0.045 (0.085)	.60	−0.213 to 0.123
Age (years)	—	—	—	0.022 (0.027)	.42	−0.031 to 0.075
Experience (years)	—	—	—	−0.033 (0.031)	.30	−0.094 to 0.029

[a]Adjusted R^2 for model A was 0.344, and *P* value from the Breusch-Pagan test for homoscedasticity for model A was .02. Heteroskedasticity is indicated if *P* value is <.05

[b]Adjusted R^2 for model B was 0.407, and *P* value from the Breusch-Pagan test for homoscedasticity for model B was 0.61. Heteroskedasticity is indicated if *P* value is <.05

[c]The number of hints given as a ratio of number of tries a learner had in the second single learning session.

[d]They are estimates of indicators for the corresponding column heading.

Discussion

Summary of Findings

This study was used to explore the effectiveness of adaptive feedback for smartphone-based training of health care workers in low-income settings, which is a largely unexplored topic of medical education in this context. We found that although there was considerable learning gain with both forms of feedback (Table 2), adaptive feedback had a weak effect of 0.09 (95% CI −0.31 to 0.46; *P*=.47), which was not statistically significant. However, when considering the background characteristics of health care providers and the various self-directed spaced learning options, and using learning gains analyzed at the individual level as opposed to the group level, adaptive feedback had an effect size of 0.644 (95% CI 0.347-0.941; *P*<.001) on student normalized learning gains with immediate repetition. Spaced learning of a week or more was associated with a significant reduction in normalized gains by 0.28. Differences in clinical cadre, level of practice/training, and previous exposure to neonatal emergency care training had no significant effect on the individual health care providers' learning gains.

Relation to Other Studies

This experiment differs from previous similar studies in digital education for clinical practice in that it (1) used mobile devices for delivery of digital education interventions, (2) evaluated novel educational modalities enabling simulated learning such as adaptive feedback, (3) provided essential methodological information necessary for comparability purposes, (4) reported relative learning gains as done in this study except one (rather they tended to report differences of postintervention scores in

XSL•FO
RenderX

the study arms), and (5) were from low-income settings such as SSA or South-East Asia [11]. In the three studies that dealt with resuscitation identified by the systematic review by Car et al [11], only one study reported relative mean change but found no significant effect of different formulations of online content on learning gain [70], which is similar to our findings, although our primary outcome is calculated differently from theirs. In that study, a variant of individual students' normalized gain was used, such as the secondary outcome we used in this study where we found the effect on individual student learning gains was 0.644 (95% CI 0.347-0.941; $P<.001$) when considering individual learner characteristics.

This study has addressed some of the recommendations from a recent evidence review into gamified education in health, which proposed future studies to employ the use of a rigorous experimental design to evaluate learning interventions, and include more studies from low- and middle-income countries, two underrepresented aspects of the current evidence base [15]. Even when considering interventions looking into adaptive feedback in digital education, none that we know of are in the health domain [71], making this study unique. Given that the implementation of the learning intervention was available to all clinical cadres involved with bedside care provision, who had varying levels of experience (from students to consultants), from multiple low-income countries, representing a varied mix of geographical and resource settings, the diverse population of the clinical taskforce in this experiment ensures that those who would most benefit from using the presented learning tool for training in emergency care delivery are well represented. Furthermore, as the overall learning gain from LIFE use in both study arms was large, given that the learning platform minimized elements that would not be typically available in routine app settings [56], this experiment's findings are generalizable to emergency care training of the health care provider population from low-income countries using smartphone-based platforms.

Implications of Findings

At the group level, the effect of adaptive feedback on health care provider learning gains was not significant, but the opposite is true when evaluated at an individual level. This might imply that the intervention effect is strongly mediated by other factors (which were external to the smartphone app), chief among them being health care providers' individualized spacing of learning repetition. Together, with the inclusion of demographic characteristics, this increased the explainable variance of adaptive feedback on individual learning gains by 18.31%. Linking the level of adaptive feedback provided to health care providers' individualized learning repetitions conditioned on their level of experience and clinical role might explain the stronger intervention effect at the individual learner level.

This difference between the group and individual learning gain might also imply that the decision of when to collect postintervention or repeated measures is a significant factor in determining what is effective learning intervention. Considering spaced learning, intuitively, the intervention effect becomes more reflective of a mechanism that works on optimizing the recall rate as opposed to the level of internalizing/understanding content. Despite efforts to encourage the latter mechanism by using reflective cues (Multimedia Appendix 6), from early findings of an ongoing qualitative evaluation for this study, an alternative model using elaborative cues on why the choices of the health care providers were wrong would arguably enhance internalization of the learning content.

The high drop-off rate with each subsequent round of play despite a lack of maximization of the learning gains (Multimedia Appendices 14-16) was disconcerting and limited the assessment of learners' skill mastery over time. The high drop-off rate might imply that (1) effort regulation and motivation for self-directed learning using LIFE is very low, (2) health care providers self-assessed learning needs lead them to believe that they have attained content mastery and thus are no longer in need of this training, or (3) other factors that are external to this experiment affect the use of the LIFE app (eg, uninstalling the LIFE app to make more space for other apps on the learner's smartphone). Qualitative interviews will be used to explore whether any of these explanations are plausible and how they can be leveraged to encourage low drop-off rates.

From our findings, a gamified smartphone-based alternative to the low-dose-high-frequency clinical training approach commonly used in low-income settings that employs a self-regulated learning approach offers significant learning gains. This is useful where face-to-face training is not possible, costs of training are a concern, and learners prefer flexible learning schedules. In addition, the spaced repetition of such a learning approach can be encouraged after a week has passed, with the encouragement differentiated based on a knowledge tracing approach that is informed by the clinical role and experience in addition to learner progression. This can be explored in future research.

Study Limitations

In this randomized experiment, we used the standard Bayesian Knowledge Tracing (BKT) modeling approach, which is explained in detail in the Multimedia Appendix 2 and elsewhere [49], to determine the adaptive feedback threshold cutoffs. However, BKT has a set of problematic assumptions: it assumes that forgetting does not occur, the knowledge components (ie, quizzes in our case) are treated as being mutually independent, its typical implementation does not allow learners to have different learning rates, it assumes that all students have the same probability of knowing a particular skill at their first opportunity (Multimedia Appendix 4), and it suffers from the problem of multiple global maxima when trying to estimate model parameters [49,72]. In addition, we used learning data from 187 health care providers from the beta phase to train the model, which is a relatively small number. Together with the moderate predictive accuracy of the health care provider's knowledge (Multimedia Appendix 2), it is highly probable that the prescription of the level of feedback might have been biased, thereby underestimating the effect of the intervention. However, this modeling approach was used because it is more easily embedded within the smartphone app and was able to function offline compared with other common alternatives [49,73].

From interviews of the physically accessible study participants, those who were from lower clinical cadres found the one-off phone data cost necessary for downloading the smartphone app

XSL•FO
RenderX

as being too high. To mitigate against burdening participants with this extra personal cost due to participating in the study, they could request for reimbursement within a few weeks of using the smartphone app. This skewed the study participants numbers toward a specific country where the reimbursements were made available (Multimedia Appendix 1). Within these resource contexts, such costs play a role in generating study participants who might not be representative of those with the intention to participate, rather representing those with the "economic" ability to participate, which is difficult to mitigate against in a multicountry study. It is challenging to determine if the reimbursement favored any study arm because of the blinded participant allocation to the study arms coupled with a lack of linking participants (whose data were collected as anonymous at source) to reimbursement.

The risk of performance bias across arms was moderate given that allowing for the "real-world" use of the smartphone app in a self-directed manner would also make it difficult to comprehensively control for exposure to factors outside the intervention of interest. This is especially true where some of the participants were in the same peer groups, hospitals, or social circles and might have collaboratively used the smartphone app. Although randomization allocation did not result in an equal number of participants in the study arms, from our post-hoc analysis (Table 1; Multimedia Appendices 10 and 11), we do not believe that this biased this experiment's findings.

Conclusions

This study set out to evaluate the effect of adaptive feedback within a smartphone-based serious game on the learning gains of health care providers from low-income countries. From 247 health care providers, the effect on learning gain was found to be g=0.09 (95% CI −0.31 to 0.46; *P*=.47). When subject-treatment interaction and differential time effect were controlled for, the effect of the adaptive feedback on learning gains increased significantly to 0.644 (95% CI 0.347 to 0.941; *P*<.001). From our findings, we suggest that linking the level of adaptive feedback provided to health care providers to how they space their learning and their clinical level might yield a larger intervention effect at both the group and individual learner levels. For the feedback content itself, as an alternative to using reflective hints on what the right answers might be, elaborating why the health care providers' responses were wrong might enhance understanding of the learning content.

Acknowledgments

The authors would like to thank Jakob Rossner, the lead developer of LIFE, and the beta-testers based in Kenya, who continue to be instrumental in the development and scale-up of the LIFE platform. Funds from the *Economic and Social Research Council* (ESRC), Award Number: 1926841 awarded to TT as a DPhil fellowship supported this work. Additional funds from *Saving Lives at Birth Grand Challenges* awarded to ME and CP and Global Challenges Research Fund (GCRF) Incubator Award Number: 161/105 grant awarded to NW, ME, and CP supported aspects of this work. ME is supported by a Senior Research Fellowship from The Wellcome Trust (#207522). The funders had no role in drafting or submitting this study.

Conflicts of Interest

HE, CW, and CP are the project managers of the LIFE platform. ME and CP are the project owners of the LIFE platform. The rest of the authors have no conflicts of interest to declare.

Multimedia Appendix 1
Participants breakdown by region.
[DOCX File , 13 KB - jmir_v22i7e17100_app1.docx]

Multimedia Appendix 2
More details on the intervention.
[DOCX File , 23 KB - jmir_v22i7e17100_app2.docx]

Multimedia Appendix 3
Bayesian Knowledge Tracing model parameters used.
[DOCX File , 14 KB - jmir_v22i7e17100_app3.docx]

Multimedia Appendix 4
Example of learning curves from the Bayesian Knowledge Tracing model for five different health care providers from the same learning task which are used to determined feedback level at each opportunity.
[PNG File , 126 KB - jmir_v22i7e17100_app4.png]

Multimedia Appendix 5
Adaptive feedback mechanism allocation within Life-Saving Instructions for Emergencies app.
[PNG File , 291 KB - jmir_v22i7e17100_app5.png]

XSL•FO
RenderX

Multimedia Appendix 6
Feedback content provided for the individual quizzes in Life-Saving Instructions for Emergencies smartphone app for Neonatal Resuscitation Scenario A Training.
[DOCX File , 17 KB - jmir_v22i7e17100_app6.docx]

Multimedia Appendix 7
Snapshots of the Life-Saving Instructions for Emergencies smartphone app.
[PNG File , 693 KB - jmir_v22i7e17100_app7.png]

Multimedia Appendix 8
Random allocation procedure.
[PNG File , 160 KB - jmir_v22i7e17100_app8.png]

Multimedia Appendix 9
Random allocation ratio based on time of day (the seed used for the generator).
[PNG File , 119 KB - jmir_v22i7e17100_app9.png]

Multimedia Appendix 10
Tests of whether differences in allocation rate between study arms resulted in non-equivalent Life-Saving Instructions for Emergencies app use for all study participants included.
[DOCX File , 15 KB - jmir_v22i7e17100_app10.docx]

Multimedia Appendix 11
Tests of whether dropout rate is different between study arms using baseline characteristics of study participants where demographic data are available.
[DOCX File , 18 KB - jmir_v22i7e17100_app11.docx]

Multimedia Appendix 12
Learning spacing between first two sessions, n (%).
[DOCX File , 14 KB - jmir_v22i7e17100_app12.docx]

Multimedia Appendix 13
Multicollinearity tests for the learning gain outcome.
[DOCX File , 13 KB - jmir_v22i7e17100_app13.docx]

Multimedia Appendix 14
Learners in each study arm at each learning iteration.
[DOCX File , 14 KB - jmir_v22i7e17100_app14.docx]

Multimedia Appendix 15
Change of student learning gains over learning sessions by study arm.
[PNG File , 56 KB - jmir_v22i7e17100_app15.png]

Multimedia Appendix 16
Change of student learning gains over successive learning sessions by spacing options.
[PNG File , 52 KB - jmir_v22i7e17100_app16.png]

Multimedia Appendix 17
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1002 KB - jmir_v22i7e17100_app17.pdf]

References

1. UNICEF. 2018. Levels and Trends in Child Mortality Report 2018 URL: https://www.unicef.org/publications/index_103264.html [accessed 2018-11-03]

XSL•FO
RenderX

2. Crisp N, Gawanas B, Sharp I, Task Force for Scaling Up Education and Training for Health Workers. Training the health workforce: scaling up, saving lives. Lancet 2008 Feb 23;371(9613):689-691. [doi: 10.1016/S0140-6736(08)60309-8] [Medline: 18295028]
3. Hategekimana C, Shoveller J, Tuyisenge L, Kenyon C, Cechetto DF, Lynd LD. Correlates of performance of healthcare workers in emergency, triage, assessment and treatment plus admission care (ETAT+) course in Rwanda: context matters. PLoS One 2016;11(3):e0152882 [FREE Full text] [doi: 10.1371/journal.pone.0152882] [Medline: 27030974]
4. World Health Organization. 2011. Transformative Scale Up of Health Professional Education: An Effort to Increase the Numbers of Health Professionals and to Strengthen Their Impact on Population Health URL: https://www.who.int/hrh/resources/transformative_education/en/ [accessed 2018-11-23]
5. Wall SN, Lee AC, Niermeyer S, English M, Keenan WJ, Carlo W, et al. Neonatal resuscitation in low-resource settings: what, who, and how to overcome challenges to scale up? Int J Gynaecol Obstet 2009 Oct;107(Suppl 1):S47-62, S63 [FREE Full text] [doi: 10.1016/j.ijgo.2009.07.013] [Medline: 19815203]
6. Chaudhury S, Arlington L, Brenan S, Kairuki AK, Meda AR, Isangula KG, et al. Cost analysis of large-scale implementation of the 'helping babies breathe' newborn resuscitation-training program in Tanzania. BMC Health Serv Res 2016 Dec 1;16(1):681 [FREE Full text] [doi: 10.1186/s12913-016-1924-2] [Medline: 27908286]
7. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, Lancet Every Newborn Study Group. Every newborn: progress, priorities, and potential beyond survival. Lancet 2014 Jul 12;384(9938):189-205. [doi: 10.1016/S0140-6736(14)60496-7] [Medline: 24853593]
8. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
9. Edgcombe H, Paton C, English M. Enhancing emergency care in low-income countries using mobile technology-based training tools. Arch Dis Child 2016 Dec;101(12):1149-1152 [FREE Full text] [doi: 10.1136/archdischild-2016-310875] [Medline: 27658948]
10. Ma W. Summit: SFU's Institutional Repository. 2017. Intelligent Tutoring Systems and Learning Outcomes: Two Systematic Reviews URL: http://summit.sfu.ca/item/17246 [accessed 2018-10-13]
11. Car L, Soong A, Kyaw BM, Chua KL, Low-Beer N, Majeed A. Health professions digital education on clinical practice guidelines: a systematic review by digital health education collaboration. BMC Med 2019 Jul 18;17(1):139 [FREE Full text] [doi: 10.1186/s12916-019-1370-1] [Medline: 31315642]
12. Bollinger R, Chang L, Jafari R, O'Callaghan T, Ngatia P, Settle D, et al. Leveraging information technology to bridge the health workforce gap. Bull World Health Organ 2013 Nov 1;91(11):890-892 [FREE Full text] [doi: 10.2471/BLT.13.118737] [Medline: 24347719]
13. Brusilovsky P, Peylo C. Adaptive and intelligent web-based educational systems. Int J Artif Intell Educ 2003;13(2-4):159-172 [FREE Full text]
14. Greenhalgh T. Computer assisted learning in undergraduate medical education. Br Med J 2001 Jan 6;322(7277):40-44 [FREE Full text] [doi: 10.1136/bmj.322.7277.40] [Medline: 11141156]
15. Gentry SV, Gauthier A, Ehrstrom CL, Wortley D, Lilienthal A, Car LT, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994. [doi: 10.2196/12994] [Medline: 30920375]
16. Ma W, Adesope OO, Nesbit JC, Liu Q. Intelligent tutoring systems and learning outcomes: a meta-analysis. J Educ Psychol 2014;106(4):901-918. [doi: 10.1037/a0037123]
17. Fontaine G, Cossette S, Maheu-Cadotte M, Mailhot T, Deschênes MF, Mathieu-Dupuis G. Effectiveness of adaptive e-learning environments on knowledge, competence, and behavior in health professionals and students: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2017 Jul 5;6(7):e128 [FREE Full text] [doi: 10.2196/resprot.8085] [Medline: 28679491]
18. Mohan D, Schell J, Angus DC. Not thinking clearly? Play a game, seriously!. J Am Med Assoc 2016 Nov 8;316(18):1867-1868 [FREE Full text] [doi: 10.1001/jama.2016.14174] [Medline: 27825014]
19. Fitzgerald K. iFEET. 1997. Nurse as Educator: Principles of Teaching and Learning URL: http://www.ifeet.org/files/-Susan_Bacorn_Bastable-_Nurse_as_Educator_Princip.pdf [accessed 2020-05-22]
20. Maheu-Cadotte M, Cossette S, Dubé V, Fontaine G, Mailhot T, Lavoie P, et al. Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: a systematic review and meta-analysis protocol. BMJ Open 2018 Mar 16;8(3):e019871 [FREE Full text] [doi: 10.1136/bmjopen-2017-019871] [Medline: 29549206]
21. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A multifaceted intervention to implement guidelines and improve admission paediatric care in Kenyan district hospitals: a cluster randomised trial. PLoS Med 2011 Apr;8(4):e1001018 [FREE Full text] [doi: 10.1371/journal.pmed.1001018] [Medline: 21483712]
22. Irimu G, Wamae A, Wasunna A, Were F, Ntoburi S, Opiyo N, et al. Developing and introducing evidence based clinical practice guidelines for serious illness in Kenya. Arch Dis Child 2008 Sep;93(9):799-804 [FREE Full text] [doi: 10.1136/adc.2007.126508] [Medline: 18719161]

XSL•FO
RenderX

23. World Health Organization. 2005. Emergency Triage Assessment and Treatment (ETAT) Course URL: https://www.who.int/maternal_child_adolescent/documents/9241546875/en/ [accessed 2018-10-09]
24. Tuyisenge L, Kyamanya P, van Steirteghem S, Becker M, English M, Lissauer T. Knowledge and skills retention following emergency triage, assessment and treatment plus admission course for final year medical students in Rwanda: a longitudinal cohort study. Arch Dis Child 2014 Nov;99(11):993-997 [FREE Full text] [doi: 10.1136/archdischild-2014-306078] [Medline: 24925893]
25. Royal College of Paediatrics and Child Health. 2016. Emergency Triage, Assessment and Treatment plus (ETAT+) - Online Learning URL: https://www.rcpch.ac.uk/resources/emergency-triage-assessment-treatment-plus-etat-online-learning [accessed 2018-12-05]
26. Poushter J. Pew Research Center. 2016. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2020-05-22]
27. Hochberg K, Kuhn J, Müller A. Using smartphones as experimental tools—effects on interest, curiosity, and learning in physics education. J Sci Educ Technol 2018 Apr 6;27(5):385-403. [doi: 10.1007/s10956-018-9731-7]
28. Winne PH, Hadwin AF. Studying as self-regulated learning. In: Hacker DJ, Dunlosky J, Graesser AC, editors. Metacognition in Educational Theory and Practice. Abingdon, UK: Routledge; 1998:277-304.
29. Goldberg B, Sottilare R, Roll I, Lajoie S, Poitras E, Biswas G, et al. Enhancing Self-Regulated Learning through Metacognitively-Aware Intelligent Tutoring Systems. In: Proceedings of the International Conference on the Learning Sciences. 2014 Presented at: ISLS'14; June 23-27, 2014; Boulder, CO p. 1352-1361 URL: https://www.researchgate.net/publication/275580372_Enhancing_Self-Regulated_Learning_through_Metacognitively-Aware_Intelligent_Tutoring_Systems
30. Butler DL, Winne PH. Feedback and self-regulated learning: a theoretical synthesis. Rev Educ Res 2016 Jun 30;65(3):245-281. [doi: 10.3102/00346543065003245]
31. Winne HP. A cognitive and metacognitive analysis of self-regulated learning. In: Schunk DH, Zimmerman BJ, editors. Handbook of Self-Regulation of Learning and Performance. Abingdon, UK: Routledge/Taylor & Francis Group; 2011:29-46.
32. Brünken R, Seufert T, Paas F. Measuring cognitive load. In: Cognitive Load Theory. Cambridge, UK: Cambridge University Press; 2010:181-202.
33. Greenwood J, Winifreyda A. Two strategies for promoting clinical competence in pre-registration nursing students. Nurse Educ Today 1995 Jun;15(3):184-189. [doi: 10.1016/s0260-6917(95)80104-9] [Medline: 7616941]
34. Rn PA. 'Feedback. I listened, reflected and utilized': third year nursing students' perceptions and use of feedback in the clinical setting. Int J Nurs Pract 2000 Oct;6(5):247-252. [doi: 10.1046/j.1440-172x.2000.00218.x]
35. Fournier JF. Information Overload and Technology Education. In: Society for Information Technology & Teacher Education International Conference. 1996 Presented at: SITTE'96; March 29-April 2, 1996; Phoenix, AZ p. 412-417 URL: https://www.semanticscholar.org/paper/Information-Overload-in-Technology-based-Education%3A-Shrivastav-Hiltz/ff94d26c6b4711216b6e034072b0cbbc2fae10fd
36. Lam CF, DeRue DS, Karam EP, Hollenbeck JR. The impact of feedback frequency on learning and task performance: challenging the 'more is better' assumption. Organ Behav Hum Decis Process 2011 Nov;116(2):217-228. [doi: 10.1016/j.obhdp.2011.05.002]
37. Voerman L, Meijer PC, Korthagen FA, Simons RJ. Types and frequencies of feedback interventions in classroom interaction in secondary education. Teach Teach Educ 2012 Nov;28(8):1107-1115. [doi: 10.1016/j.tate.2012.06.006]
38. Shute VJ. Focus on formative feedback. Rev Educ Res 2008 Mar;78(1):153-189. [doi: 10.3102/0034654307313795]
39. Tuti T, Winters N, Muinga N, Wanyama C, English M, Paton C. Basedevaluation of adaptive feedback in a smartphone-based serious game on health care providers' knowledge gain in neonatal emergency care: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jul 26;8(7):e13034 [FREE Full text] [doi: 10.2196/13034] [Medline: 31350837]
40. Moher D, Schulz KF, Altman D, CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. J Am Med Assoc 2001 Apr 18;285(15):1987-1991. [doi: 10.1001/jama.285.15.1987] [Medline: 11308435]
41. Etikan I. Comparision of snowball sampling and sequential sampling technique. Biom Biostat Int J 2016 Jan 2;3(1):55. [doi: 10.15406/bbij.2016.03.00055]
42. Unreal Engine Documentation. 2018. Shipping With Android URL: https://docs.unrealengine.com/en-us/Platforms/Mobile/Android/ShippingWithAndroid#androidsdkversion [accessed 2020-04-30]
43. VanLehn K, Jordan P, Litman D. ISCA Speech. 2007. Developing Pedagogically Effective Tutorial Dialogue Tactics: Experiments and a Testbed URL: https://www.isca-speech.org/archive_open/archive_papers/slate_2007/sle7_017.pdf [accessed 2020-05-22]
44. Atukunda IT, Conecker GA. Effect of a low-dose, high-frequency training approach on stillbirths and early neonatal deaths: a before-and-after study in 12 districts of Uganda. Lancet Global Health 2017 Apr;5:S12. [doi: 10.1016/s2214-109x(17)30119-5]
45. Yudelson M. GitHub. Command-line Utility for Fitting Hidden Markov Models at Scale URL: https://github.com/myudelson/hmm-scalable [accessed 2020-10-21]

XSL•FO
RenderX

46. Kiili K, de Freitas S, Arnab S, Lainema T. The design principles for flow experience in educational games. Procedia Comput Sci 2012;15:78-91. [doi: 10.1016/j.procs.2012.10.060]
47. Dewey J. Experience and Education. Indianapolis: Kappa Delta Pi; 1998.
48. Holmes W, Mavrikis M, Hansen A, Grawemeyer B. Purpose and Level of Feedback in an Exploratory Learning Environment for Fractions. In: Proceedings of the International Conference on Artificial Intelligence in Education. 2015 Presented at: AIED'15; June 21-25, 2015; Madrid, Spain p. 620-623. [doi: 10.1007/978-3-319-19773-9_76]
49. Yudelson MV, Koedinger KR, Gordon GJ. Individualized Bayesian Knowledge Tracing Models. In: Proceedings of the International Conference on Artificial Intelligence in Education. 2013 Presented at: AIED'13; July 9-13, 2013; Memphis, TN, USA. [doi: 10.1007/978-3-642-39112-5_18]
50. Holmes W, Anastopoulo S, Schaumburg H, Mavrikis M. Open Research Online - The Open University. 2018. Technology-Enhanced Personalised Learning: Untangling the Evidence URL: http://oro.open.ac.uk/56692/1/TEPL_en.pdf [accessed 2019-11-01]
51. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. Organ Res Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
52. Ellis P. Hong Kong Polytechnic University. 2009. Thresholds for Interpreting Effect Sizes URL: http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html [accessed 2018-12-05]
53. Glass GV, Smith ML, McGaw B. Meta-Analysis in Social Research: Individual and Neighbourhood Reactions. Thousand Oaks, CA: Sage Publications; 1981.
54. Ian W. University of Bath. 2008. Statistics for Psychology URL: http://staff.bath.ac.uk/pssiw/stats2/page2/page14/page14.html [accessed 2020-05-01]
55. Marx JD, Cummings K. Normalized change. Am J Phys 2007 Jan;75(1):87-91. [doi: 10.1119/1.2372468]
56. Peters D, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. Br Med J 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]
57. Winne PH, Nesbit JC, Kumar V, Hadwin AF, Lajoie SP, Azevedo RA, et al. Supporting self-regulated learning with gStudy software: the learning kit project. Technol Instruct Cogn Learn 2006;3(1/2):105-113 [FREE Full text]
58. Fontaine G, Cossette S, Maheu-Cadotte M, Mailhot T, Deschênes MF, Mathieu-Dupuis G, et al. Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. BMJ Open 2019 Aug 28;9(8):e025252 [FREE Full text] [doi: 10.1136/bmjopen-2018-025252] [Medline: 31467045]
59. Wijk H, Ponzer S, Heikkilä K, Kihlström L, Nordquist J. Factors influencing effectiveness in postgraduate medical education - a qualitative study of experiences of the responsible clinical consultants. BMC Med Educ 2019 Jan 3;19(1):3 [FREE Full text] [doi: 10.1186/s12909-018-1433-6] [Medline: 30606174]
60. Pickering JD. Developing the evidence-base to support the integration of technology-enhanced learning in healthcare education. Med Sci Educ 2017 May 29;27(4):903-905. [doi: 10.1007/s40670-017-0424-2]
61. QuantStart. 2019. Random Number Generation via Linear Congruential Generators in C++ URL: https://www.quantstart.com/articles/Random-Number-Generation-via-Linear-Congruential-Generators-in-C [accessed 2020-10-21]
62. Abraha I, Cherubini A, Cozzolino F, de Florio R, Luchetta ML, Rimland JM, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. Br Med J 2015 May 27;350:h2445 [FREE Full text] [doi: 10.1136/bmj.h2445] [Medline: 26016488]
63. Moore S, Tassé AM, Thorogood A, Winship I, Zawati M, Doerr M. Consent processes for mobile app mediated research: systematic review. JMIR Mhealth Uhealth 2017 Aug 30;5(8):e126 [FREE Full text] [doi: 10.2196/mhealth.7014] [Medline: 28855147]
64. Grady C, Cummings SR, Rowbotham MC, McConnell MV, Ashley EA, Kang G. Informed consent. N Engl J Med 2017 Mar 2;376(9):856-867. [doi: 10.1056/NEJMra1603773] [Medline: 28249147]
65. Lavavej S. Channel 9. 2013. Rand() Considered Harmful URL: https://channel9.msdn.com/Events/GoingNative/2013/rand-Considered-Harmful [accessed 2020-10-21]
66. Schultz BB. Levene's test for relative variation. Syst Zool 1985 Dec;34(4):449. [doi: 10.2307/2413207]
67. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
68. Bolker B. R Documentation. 2019. Test Fitted Model For (Near) Singularity URL: https://www.rdocumentation.org/packages/lme4/versions/1.1-21/topics/isSingular [accessed 2020-10-21]
69. Misangyi VF, LePine JA, Algina J, Goeddeke F. The adequacy of repeated-measures regression for multilevel research. Organ Res Methods 2016 Jun 29;9(1):5-28. [doi: 10.1177/1094428105283190]
70. Leszczyński P, Charuta A, Łaziuk B, Gałązkowski R, Wejnarski A, Roszak M, et al. Multimedia and interactivity in distance learning of resuscitation guidelines: a randomised controlled trial. Interact Learn Environ 2017 Jun 14;26(2):151-162. [doi: 10.1080/10494820.2017.1337035]
71. Bimba AT, Idris N, Al-Hunaiyyan A, Mahmud RB, Shuib NL. Adaptive feedback in computer-based learning environments: a review. Adaptive Behavior 2017 Aug 24;25(5):217-234. [doi: 10.1177/1059712317727590]
72. Gong Y, Back JE, Heffernan NT. How to construct more accurate student models: comparing and optimizing knowledge tracing and performance factor analysis. Int J Artif Intell Educ 2011;21(1-2):27-46. [doi: 10.5555/2336135.2336138]

XSL•FO
RenderX

73. Chi M, Koedinger KR, Gordon GJ, Jordan P, VanLehn K. Instructional Factors Analysis: a Cognitive Model for Multiple Instructional Interventions. In: Proceedings of the 4th International Conference on Educational Data Mining. 2011 Presented at: EDM'11; July 6-8, 2011; Eindhoven, The Netherlands p. 61-70.

Abbreviations

BKT: Bayesian Knowledge Tracing
ETAT+: Emergency Triage, Assessment, and Treatment Plus admission care training
KEMRI: Kenya Medical Research Institute
LIFE: Life-Saving Instructions for Emergencies
SRL: self-regulating learning
SSA: sub-Saharan Africa

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by R Dekova, S Kevany, S Choi; comments to author 27.01.20; revised version received 13.03.20; accepted 08.04.20; published 06.07.20.

Please cite as:
Tuti T, Winters N, Edgcombe H, Muinga N, Wanyama C, English M, Paton C
Evaluation of Adaptive Feedback in a Smartphone-Based Game on Health Care Providers' Learning Gain: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17100
URL: https://www.jmir.org/2020/7/e17100
doi:10.2196/17100
PMID:32628115

©Timothy Tuti, Niall Winters, Hilary Edgcombe, Naomi Muinga, Conrad Wanyama, Mike English, Chris Paton. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Internet-Based Psychological Intervention With a Serious Game to Improve Vitality, Psychological and Physical Condition, and Immune Function in Healthy Male Adults: Randomized Controlled Trial

Lemmy Schakel[1,2], PhD; Dieuwke S Veldhuijzen[1,2], PhD; Henriët van Middendorp[1,2], PhD; Corine Prins[3], BSc; Anne M H F Drittij[3], MSc; Frank Vrieling[3], MSc; Leo G Visser[3], Prof Dr; Tom H M Ottenhoff[3], Prof Dr; Simone A Joosten[3], PhD; Andrea W M Evers[1,2,4], Prof Dr

[1]Health, Medical and Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden, Netherlands

[2]Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands

[3]Department of Infectious Diseases, Leiden University Medical Centre, Leiden, Netherlands

[4]Department of Psychiatry, Leiden University Medical Centre, Leiden, Netherlands

Corresponding Author:
Andrea W M Evers, Prof Dr
Health, Medical and Neuropsychology Unit
Institute of Psychology
Leiden University
Wassenaarseweg 52
Leiden
Netherlands
Phone: 31 715273627
Email: a.evers@fsw.leidenuniv.nl

Abstract

Background: Recently, internet-based cognitive behavioral therapy (ICBT) and serious gaming interventions have been suggested to enhance accessibility to interventions and engagement in psychological interventions that aim to promote health outcomes. Few studies, however, have investigated their effectiveness in the context of simulated real-life challenges.

Objective: We aimed to examine the effectivity of a guided ICBT combined with a serious gaming intervention in improving self-reported psychophysiological and immunological health endpoints in response to psychophysiological and immune-related challenges.

Methods: Sixty-nine healthy men were randomly assigned to the intervention condition, receiving ICBT combined with serious gaming for 6 weeks, or the control condition, receiving no intervention. Self-reported vitality was the primary endpoint. Other self-reported psychophysiological and immunological endpoints were assessed following various challenges, including a bacillus Calmette-Guérin vaccination evoking pro-inflammatory responses, 1 and 4 weeks after the intervention period.

Results: Although the intervention did not affect vitality-associated parameters, self-reported sleep problems (P=.027) and bodily sensations (P=.042) were lower directly after the intervention compared with controls. Furthermore, wellbeing (P=.024) was higher in the intervention group after the psychophysiological challenges. Although no significant group differences were found for the psychophysiological and immunological endpoints, the data provided preliminary support for increased immunoglobulin antibody responses at the follow-up time points (P<.05). Differential chemokine endpoints between conditions were observed at the end of the test day.

Conclusions: The present study provides some support for improving health endpoints with an innovative ICBT intervention. Future research should replicate and further extend the present findings by consistently including challenges and a wide range of immune parameters into the study design.

Trial Registration: Nederlands Trial Register NTR5610; https://www.trialregister.nl/trial/5466

(J Med Internet Res 2020;22(7):e14861) doi:10.2196/14861

XSL•FO
RenderX

KEYWORDS

BCG vaccination; cognitive behavioral therapy; immune system; stress; ICBT

Introduction

Psychological interventions have been shown to be effective in improving self-reported health outcomes [1,2] and immune status [3-5]. Two meta-analytic reviews found modest support for improved immune function after psychological interventions [6,7]. However, it is difficult to draw conclusive findings from these studies due to the large heterogeneity in the incorporated interventions (ie, various types of relaxation, conditioning, disclosure, and stress management interventions) and immunological outcomes (ie, quantitative and qualitative). Moreover, novel developments in psychological treatments may potentially further enhance the effectiveness of psychological interventions in improving self-reported and immunological health outcomes [7]. A rather novel development focuses on providing psychological interventions based on cognitive behavioral therapy (CBT) via the internet (ICBT). CBT focuses on enhancing coping and problem-solving skills and therefore can be applied to improve health management strategies [8]. Furthermore, cognitive behavioral strategies can adjust standards of health and in turn improve quality of life [8]. A meta-analysis showed that the effectiveness of guided ICBT interventions is comparable to face-to-face interventions in patients with chronic somatic conditions [9]. Advantages of ICBT over face-to-face interventions are the increased convenience in use and enhanced flexibility of the specific location and time for users regarding where the intervention sessions are completed [10]. However, adherence rates in ICBT are lower compared to face-to-face interventions; therefore, engagement should be taken into account [11,12].

Engagement can possibly be enhanced by applying persuasive internet-based technologies, such as serious gaming. Serious gaming aims to provide education in an entertaining manner and can therefore be intrinsically motivating [13,14]. A meta-analysis showed evidence for the effectiveness of serious gaming in promoting a healthy lifestyle [15]. Serious games are able to tap into multiple learning processes such as explicit behavior change strategies (eg, goalsetting and transferring knowledge [16]), but also imply more implicit behavior change strategies (eg, priming and evaluative conditioning). Serious gaming could therefore be an interesting technique to add to ICBT interventions to further optimize their effectiveness.

Most research has focused on basal health outcomes; however, health outcomes assessed in situations that challenge actual health status might better represent reactions to real-time stressors [17]. Few studies have incorporated immunological or psychophysiological challenges in their study design. Immunological challenges may comprise in vitro exposure to a chemical substance (eg, lipopolysaccharide or pokeweed mitogen [18,19]) to obtain insights into the cellular responses after a psychological intervention. Immunological challenges can also be applied in vivo to observe subsequent responses (eg, antibody responses after vaccination [20] or healing process of experimentally created wounds [21]). Psychophysiological challenges can provide insights into participants' responses to stress after a psychological intervention (eg, exposure to a social evaluative stressor). The studies that incorporated challenges into their design often focused on incorporating one specific challenge and did not combine and compare effects on both in vitro and in vivo immunological as well as psychophysiological challenges [17]. For example, a recent systematic review of studies that specifically examined wound healing after a psychological intervention provided some support for the optimization of immunological markers after this challenge [22]. It is possible that incorporating multiple challenges will provide a more concise view on the effectiveness of a psychological intervention and therefore provide further insights into the link between psychological and immunological mechanisms.

The aim of this randomized controlled trial was therefore to investigate whether an ICBT intervention combined with serious gaming to optimize its effectiveness and engagement can effectively improve self-reported psychophysiological and immunological health endpoints in response to in vitro and in vivo immunological and psychophysiological challenges [17,23]. To gather more insights into the potential effectiveness of an ICBT intervention combined with serious gaming intervention in preventing adverse health endpoints and improving immune function, we included healthy participants in this study. Including healthy participants also provided us with the opportunity to incorporate multiple immunological and psychosocial challenges to gather more insights into the mechanisms underlying the effects of psychological interventions on health endpoints. It was hypothesized that participants in the intervention condition would show higher self-reported vitality and related health endpoints after the intervention compared with the control condition. In addition, improved self-reported psychophysiological and immune-related health endpoints after the in vitro and in vivo immunological and psychosocial challenges were expected in the intervention condition compared with the control condition.

Methods

Ethical Considerations

The study protocol was approved by the Medical Ethical Committee of Leiden University Medical Centre (registration number P15.099/NL52434.058.15) and preregistered at the Netherlands National Trial Register (NTR5610). The study was conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonisation Guidelines on Good Clinical Practice. Details on the study protocol and design have been published previously [23].

Study Population

To gather more insights into the potential effectiveness of ICBT combined with serious gaming in preventing adverse health endpoints and improving immune function, we included healthy participants in this study. The inclusion and exclusion criteria are described in detail in a previously published article on the

XSL•FO
RenderX

study protocol and design [20]. Briefly, healthy male participants between 18 and 35 years of age without any somatic or psychological conditions interfering with the study protocol were eligible to participate in the study. We only included male participants as the menstrual cycle is known to affect immune function [23,24]. Participants were recruited through digital and printed flyers at various faculties of Dutch universities from February 2016 until April 2018.

Procedure

Participants were informed that the study was about the effectiveness of a psychological training program directed at optimizing immune function. After signing an informed consent form, participants completed the self-reported and psychophysiological endpoints, and venous blood was collected. Participants who met the inclusion criteria were randomly assigned to a 6-week intervention or control condition. In the week following the 6-week intervention or control period (ranging from 1 to 7 days after completion of the intervention period), all participants again completed the self-reported and psychophysiological endpoints, and blood was collected. Directly afterwards, participants were vaccinated with bacillus Calmette-Guérin (BCG). One day later, a test day followed, on which psychophysiological stress challenges (ie, Paced Auditory Serial Addition Task, Cold Pressor Test, and Trier Social Stress Test) were administered. At the start and end of the test day, self-reported and psychophysiological endpoints were assessed (see Table 1), and blood was again collected. After 4 weeks, a follow-up measurement was conducted, including self-reported endpoints as well as psychophysiological endpoints and collection of a blood sample. Total time investment for participating in the study was around 15-20 hours, depending on the group allocation. This also included 4 visits to the study center. Participants received €200 for their participation.

Table 1. Details of the endpoints at each measurement point.

Endpoint	Baseline	After intervention /pre-vaccination	Start test day	End test day	Follow-up
Self-reported endpoints	SVS[a], CIS-20[b], RAND-36, PILL[c], MOS[d] Sleep, PANAS[e], and NRS[f]	SVS, CIS-20, RAND-36, PILL, MOS Sleep, PANAS, and NRS	SVS, CIS-20, PANAS, and NRS	PANAS and NRS	SVS, CIS-20, RAND-36, PILL, MOS Sleep, PANAS, and NRS
Psychophysiological endpoints	Heart rate variables, skin conductance, cortisol, and alpha amylase	N/A[g]	Heart rate variables, skin conductance, cortisol, and alpha amylase	Heart rate variables, skin conductance, cortisol, and alpha amylase	Heart rate variables, skin conductance, cortisol, and alpha amylase
Immune endpoints	Unstimulated as well as LPS[h]-stimulated serum samples	Unstimulated as well as LPS-stimulated serum samples	Unstimulated as well as LPS-stimulated serum samples	Unstimulated serum sample	Unstimulated serum sample

[a]SVS: Subjective Vitality Scale.

[b]CIS-20: Checklist Individual Strength.

[c]PILL: Pennebaker Inventory of Limbic Languidness.

[d]MOS: Medical Outcomes Study.

[e]PANAS: Positive and Negative Affect Schedule.

[f]NRS: numeric rating scale.

[g]N/A: not applicable.

[h]LPS: lipopolysaccharide.

Randomization and Blinding

Participants were randomized to the intervention or control condition based on a 1:1 allocation ratio. A block randomization was performed by the first author with random.org (block size of 4) to control for seasonal influences [23]. The test leader on the test day was blinded for group allocation.

Intervention

See Table 2 for an overview of the intervention. Participants in the intervention group received a guided ICBT intervention for 6 weeks [23], which was based on an ICBT intervention for chronic somatic diseases developed in our research group [25,26]. The intervention was delivered on the internet platform and software hosted by Karify [27]. The intervention could be accessed individually by the participant. It was provided for free, and the website was password protected. Access was granted by providing participants with an email link in which participants were invited to set up a personal account. The intervention contained 6 modules (goal setting, healthy food and exercise, relaxation, sleep, cognition and worldview, and long-term goals). These modules were guided by a therapist (psychologist supervised by a CBT psychologist) from whom participants received homework assignments and asynchronously provided feedback messages. In addition, participants in the intervention condition played a serious game (ViaNova©) that incorporated comparable modules as the guided intervention (ie, healthy food and exercise, sleep, relaxation, and long-term goals) as part of the ICBT. A subset of these games that focused specifically on food-related health behavior was tested in a previous study that demonstrated preliminary support for the effectiveness of a single serious gaming session in optimizing virtual food choice and implicit food preference [28]. Two weeks after the intervention, participants received a booster session by telephone which lasted

XSL•FO
RenderX

15-30 minutes and focused on relapse prevention by asking participants how they worked on their goals after the last online session and what strategies they used to keep up with their goals. The control condition did not receive any training.

Table 2. Overview of the intervention.

Week	Module	Actions
1	Module goal setting	Face-to-face intake with the therapist and setting goals for the online intervention
2	Module healthy food and exercise	Reading information, keeping a diary on goal progress, reading the information provided online by the therapist, completing assignments provided by the therapist
3	Module relaxation	Reading information, keeping a diary on goal progress, reading the information provided online by the therapist, completing assignments provided by the therapist
4	Module sleep	Reading information, keeping a diary on goal progress, reading the information provided online by the therapist, completing assignments provided by the therapist
5	Module cognition and world view	Reading information, keeping a diary on goal progress, reading the information provided online by the therapist, completing assignments provided by the therapist
6	Module long-term goals	Reading information, keeping a diary on goal progress, reading the information provided online by the therapist, completing assignments provided by the therapist

Challenges

In Vitro and In Vivo Immunological Challenges

As an in vitro immunological challenge, heparinized whole blood samples were stimulated in vitro with lipopolysaccharide (LPS) to stimulate cytokine production at baseline (before the intervention), at the start of the vaccination day, and one day later at the start of the test day [23]. The process consisted of stimulating 1 mL of sodium-heparinized blood in BD Vacutainer blood collection tubes (BD, Franklin Lakes, NJ) with LPS (*Escherichia coli*, ultra-pure, Invivogen, Toulouse, France) at a final concentration of 100 ng/mL or as a control without LPS, and samples were incubated at 37 °C for 6 hours. Tubes were spun at 3400 rpm for 10 minutes, and plasma was collected and stored until testing at –80 °C.

In addition, in the week following the intervention (or similar time frame for the control arm), all participants were vaccinated with *Mycobacterium bovis* BCG, a live-attenuated vaccine used against tuberculosis. This vaccine was incorporated as an in vivo challenge to the immune system. BCG (Intervax, via RIVM, Bilthoven, The Netherlands) was administered by intradermal injection (0.1 mL) in the upper arm.

Psychophysiological Challenges

On the day post-vaccination, participants were exposed to 3 psychophysiological challenges in the following order: a modified version of the Paced Auditory Serial Addition Task [29], a Cold Pressor Test [30], and the Trier Social Stress Test [31]. All challenges are known to reliably induce psychophysiological stress responses [30-33]. More information regarding these challenges has been published previously [23].

Primary Endpoints

Self-Reported Vitality

The Subjective Vitality Scale (SVS) [34] and Checklist Individual Strength (CIS-20) [35,36] were used to measure self-reported vitality. The composite score of the SVS and CIS-20 was used as a primary endpoint in this study, to gather a rather complete view on vitality. This composite score was determined by subtracting the standardized sum score of the CIS-20 from the standardized sum score of the SVS. Scores on the composite scale can be interpreted as higher scores representing higher self-reported vitality. The SVS and CIS-20 have been shown to be reliable and valid in previous research [37,38] and had good internal reliability in the present study (Cronbach α=.84 and .87, respectively).

Secondary Endpoints

Self-Reported Quality of Life

In addition, the RAND-36 was used to assess physical and mental health-related quality of life by determining sum scores of the subscales physical functioning and emotional wellbeing [39], which has been shown to be reliable and valid in previous literature [40]. Standardized T-scores were computed for both scales, with higher scores representing higher self-reported quality of life.

Self-Reported Bodily Sensations

Bodily sensations were measured with the Pennebaker Inventory of Limbic Languidness [41]. The Pennebaker Inventory of Limbic Languidness showed good internal reliability in the present study (Cronbach α=.89).

Self-Reported Sleep Problems

Sleep problems were assessed with 9 items of the Medical Outcomes Study Sleep Scale [42], which showed good internal reliability previously [42]. Higher scores on this scale represent lower levels of self-reported sleep problems. Although this questionnaire yielded sufficient internal reliability at follow-up (Cronbach α=.73), the internal reliability in the present study was low at baseline and after the intervention (Cronbach α=.45 and .36, respectively); therefore, the results of this scale in the present study should be interpreted with caution.

Self-Reported Wellbeing

Wellbeing was assessed using the 20-item Positive and Negative Affect Schedule [43] and a 7-item numeric rating scale (NRS) on wellbeing. The Positive and Negative Affect Schedule was subdivided into the positive affect scale and the negative affect

XSL•FO
RenderX

scale, which both showed good reliability and validity in previous literature [44] as well as good reliability in the present study (Cronbach α =.88 and .70, respectively). On the NRS that was used to measure wellbeing, scores ranged from 0 (not at all) to 10 (very much). Higher scores on this questionnaire represent higher levels of self-reported wellbeing. The NRS showed good internal reliability in the present study (Cronbach α=.80).

Psychophysiological Endpoints

Heart rate (HR), heart rate variability (HRV), and skin conductance were assessed continuously with a BIOPAC MP150 system (BIOPAC Systems Inc, Goleta, CA) using AcqKnowledge software version 4.1.1. Furthermore, HR, HRV, and skin conductance were measured at a resting state for 5 minutes at specific time points (see Table 1). Recording of the electrocardiogram signal was performed with an ECG100C module set at 1000Hz. The high pass filter was set at 0.05 Hz and the low pass filter at 35Hz. For HR, electrodes were attached at the sternum and somewhat below the left lower rib. To measure skin conductance, Ag/Agcl electrodes were attached at the medial phalanges of two fingers of the non-dominant hand (ie, the middle and index fingers). A GSR100C module was used to measure skin conductance, set at 1000 Hz. Gain was set at 5 μΩ/V and the low pass filter at 10 Hz. The Physio Data Toolbox Version 0.4 was used for visual inspection of the data as well as for calculating the mean HR, HRV, and skin conductance levels at each time point [45].

Saliva samples were collected to measure cortisol and alpha amylase. Samples were stored at –80 °C until analyzed. Cortisol was assessed in saliva with a competitive electrochemiluminescence immunoassay using a Modular Analytics E602 immunoanalyzer (Roche Diagnostics, Mannheim, Germany). Cortisol activities were measured and expressed in nmol/L. Determination of salivary alpha amylase was performed using a kinetic colorimetric assay for total amylase activity (Cat Nr. 03183742, Roche Diagnostics, Mannheim, Germany) on a routine clinical chemistry analyzer. Amylase activity was measured and expressed in U/L.

Immune Endpoints

Blood samples were collected in cloth-activating tubes (BD Vacutainer) at baseline, after the intervention/pre-vaccination, post-vaccination, and at the 4-week follow-up. Samples were clotted for an hour at room temperature before centrifugation at 2500 rcf for 10 minutes, and serum was collected and aliquoted for storage at –80 °C.

The list of cytokines and chemokines that were analyzed is specified in Multimedia Appendix 1. Cytokine and chemokine levels were measured in serum as well as in stimulated or control plasma samples using the multiplex bead array (Bio-Plex Pro Human Chemokine Panel, 40-Plex #171AK99MR2, Bio-Rad laboratories, Veenendaal, The Netherlands [46]). C-reactive protein concentrations were determined in serum by ELISA according to the instructions of the manufacturer (Abnova, Heidelberg, Germany) at baseline, at the start of the vaccination day, at the start of the test day, and at follow-up.

In addition, immunoglobulin G (IgG) antibody levels were evaluated at baseline and 4 weeks after vaccination. Purified protein derivative (5 μg/mL; Statens Serum Institute, Copenhagen, Denmark) was coated to 96 well Microlon plates (Greiner, Alphen aan den Rijn, The Netherlands). Sera were diluted 1 to 25 and incubated overnight. IgG antibody binding was detected using HRP-labelled polyclonal rabbit anti-human IgG (Dako, Glostrup, Denmark), staining with TMB substrate buffer (Sigma Aldrich, Zwijndrecht, The Netherlands), stopped with H2SO4 and OD450 reading [47].

Statistical Analyses

All analyses were performed using SPSS Statistics (version 25; IBM Corp, Armonk, NY). As described in our design paper [23], a total sample size of 60 participants was deemed sufficient to detect scientifically and clinically relevant differences in the incorporated primary endpoint. An analysis of covariance (ANCOVA) with condition (intervention vs control) as the between-subjects factor, vitality after the intervention as the dependent variable, and baseline vitality as the covariate was conducted to assess the primary hypothesis that participants in the intervention condition would show higher self-reported vitality after the intervention (pre-vaccination) compared with the control condition. In addition, when a significant effect was found in the ANCOVA, it was investigated whether the effects were also present at the other time points. This was done with repeated measures analysis of variance (ANOVA) with condition (intervention vs control) as the between-subjects factor and time (ie, baseline, after intervention/pre-vaccination, after vaccination, follow-up) as the within-subjects factor. For the repeated measures ANOVAs, we were specifically interested in the interaction effects between time and condition, as well as in the main effects of time, which are therefore specified in the Results section. To examine at which time point(s) groups differed on vitality, represented by a significant interaction effect between time and condition in the repeated measures ANOVA, Holm's corrected ANOVAs were performed to compare the intervention condition with the control condition at specific time intervals by calculating difference scores between baseline and each of the other time points. Since we did not observe substantial missing data or deviations from the actual timeline within participants, we decided to test the secondary endpoints in a similar way (repeated measures ANOVA) as for the primary endpoint instead of the preplanned multilevel analyses for the secondary endpoints [23]. The results for bodily sensations, quality of life, and sleep problems were analyzed as described at 3 time points (ie, baseline, after intervention/pre-vaccination, follow-up). As the items on these questionnaires were based on the experiences of the last 4 weeks, these questionnaires were not completed post-vaccination.

To test any group differences for wellbeing and positive and negative affect in response to the test day, repeated measures ANOVAs were performed for wellbeing and positive and negative affect with condition (intervention vs control) as the between-subjects factor and the 4 time points (ie, baseline, start of the test day, end of the test day, follow-up) as the within-subjects factor. Data on cortisol, alpha amylase, HR, HRV, and skin conductance were analyzed in a similar way.

XSL•FO
RenderX

For both serum and LPS whole blood stimulation assay, principal component analysis was performed to identify and subsequently exclude extreme outliers. Interleukin (IL)-6 and IL-8 were excluded from the LPS whole blood stimulation analysis. For each time point comparison, two types of linear models were fitted. The first was a linear multiple regression model using Δ-cytokine concentrations at different time points (ie, pg/mL at start of the test day – pre-vaccination, pg/mL at the end of the test day – pre-vaccination, and pg/mL at follow-up – baseline) as dependent variables to estimate the effect of the intervention as an independent variable on changes in cytokine concentrations while correcting for age. The second was a linear mixed model with a random intercept per subject to estimate the effect of time on cytokine levels in either the control or intervention group while correcting for age. Resulting *P* values were false discovery rate (FDR)–corrected to obtain *q* values. Data were mean-centered and scaled to standard deviation units for the generation of volcano plots. Finally, principal component analysis, fitting of multiple linear regression models and linear mixed models, and plotting of analysis results were performed using R version 3.5.0 with the following packages: 'mixOmics' [48], 'lme4' [49], 'lmerTest' [50], and 'ggplot2' [51].

Results

Sample Characteristics

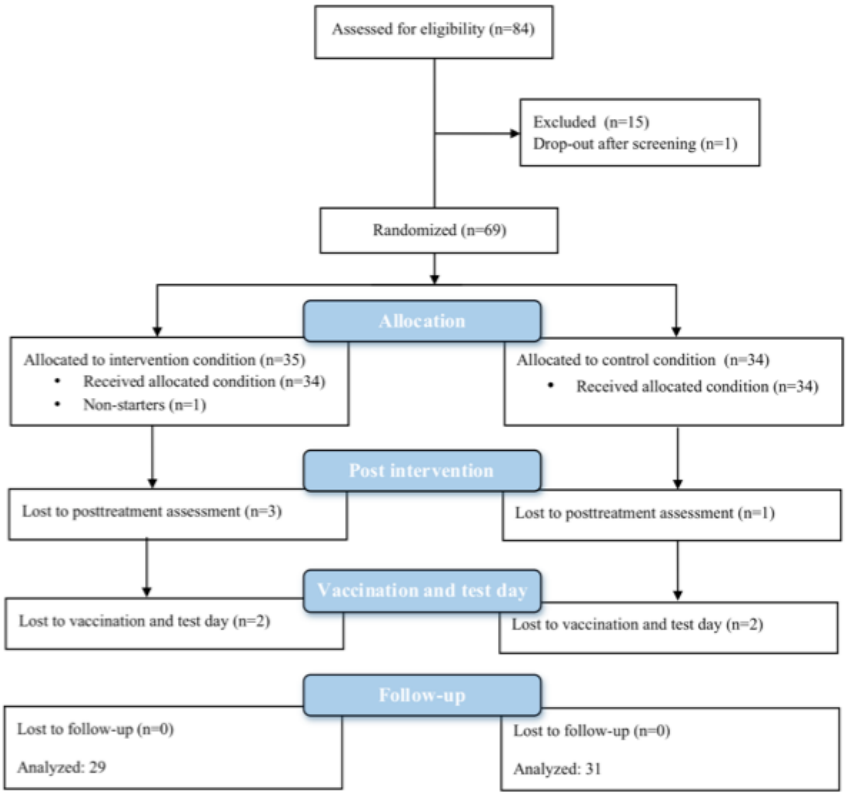
Of the 84 participants assessed for eligibility, 14 participants did not meet the inclusion criteria (7 due to somatic morbidity, 5 due to psychological morbidity, and 2 due to practical planning issues). One participant dropped out of the study directly after the screening. Therefore, 69 participants were randomized to one of the two conditions in the present study (see Figure 1). Then, 3 participants dropped out of the study, 1 in the control condition and 2 in the intervention condition. Additionally, 1 participant did not start in the intervention condition after group allocation, due to time constraints. Due to global production problems of the BCG vaccine, 2 participants in the intervention condition and 2 participants in the control condition dropped out of the study after completion of the primary endpoint. Furthermore, 1 participant in the intervention condition dropped out of the study after completion of the intervention due to time constraints. This resulted in 31 participants in the control condition and 29 participants in the intervention condition that completed all visits. Analyses were performed for available data. No significant differences were found in age or BMI between the participants in the control and intervention conditions (*P*>.05). See Table 3 for an overview of the age, BMI, and baseline level of vitality of the participants.

Primary Endpoint

Vitality

No significant differences were found between the groups for self-reported vitality within 1 week after the intervention (pre-vaccination; *P*=.43). The descriptive results for vitality at all time points are displayed in Figure 2.

Figure 1. Flow diagram.

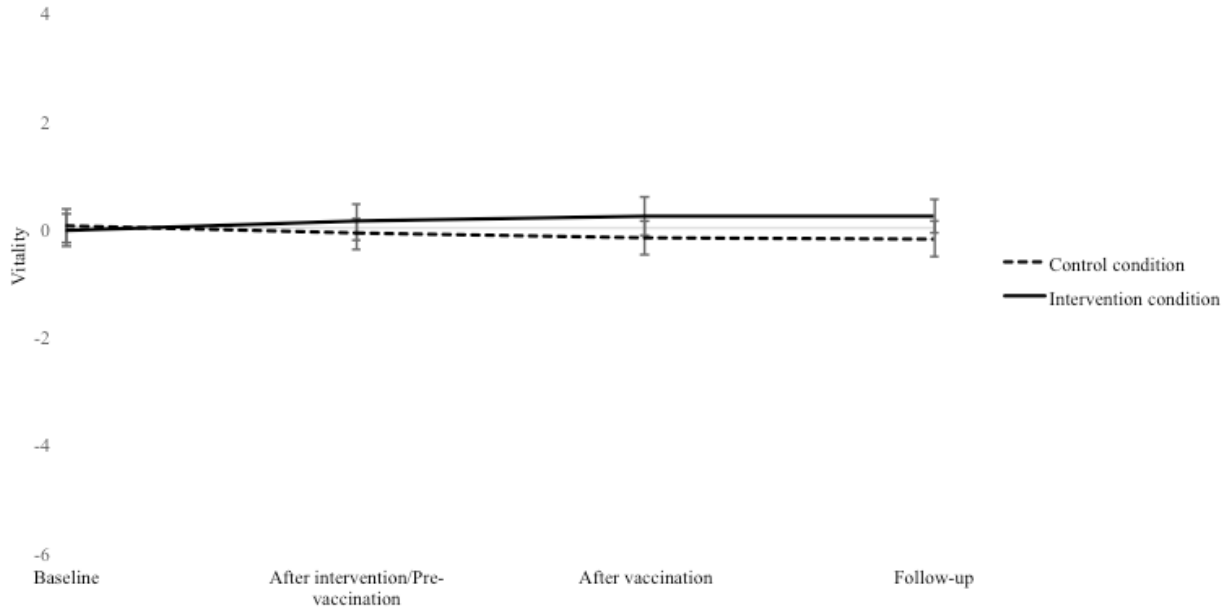


XSL•FO
RenderX

Table 3. Demographic characteristics for the control and intervention conditions.

Characteristics	Control condition	Intervention condition
Age (years), mean (SD)	22.9 (4.1)	22.5 (2.3)
BMI (kg/m^2), mean (SD)	23.0 (2.8)	22.5 (2.4)
Standardized vitality score, mean (SD)	0.05 (1.84)	–0.05 (1.66)

Figure 2. Mean and standard error of self-reported vitality at baseline, after intervention (pre-vaccination), after vaccination, and at follow-up, separately for the control condition and the intervention condition.



Secondary Endpoints

Self-Reported Quality of Life, Bodily Sensations, Sleep, Positive and Negative Affect, and Wellbeing

In Multimedia Appendix 2, the results for quality of life for the physical (A) and mental (B) quality of life subscales are shown. Both ANCOVAs did not yield any significant group differences (*P*=.92, *P*=.24, respectively).

Figure 3 depicts the results for bodily sensations. The ANCOVA yielded a significant main effect for condition ($F_{1,62}$=4.30, *P*=.042, n^2 =.56), indicating fewer bodily sensations for the intervention condition compared with the control condition directly after the intervention (pre-vaccination). The repeated measures ANOVA yielded a significant main effect of time ($F_{1.65,79.03}$=7.30, *P*=.002). Irrespective of condition, Holms corrected pairwise comparisons showed a significant decrease from baseline to after the intervention (pre-vaccination; t_{64}=3.16, adjusted *P*=.004), as well as a significant decrease from baseline to follow-up (t_{49}=2.43, adjusted *P*=.019). No significant interaction effect between time and condition was found (*P*=.36).

The results for sleep problems are presented in Figure 4. The ANCOVA showed a trend for an effect of the intervention ($F_{1,62}$= 3.30, *P*=.074, n^2=.44). The repeated measures ANOVA did not yield a significant effect of time (*P*=.18) but showed a significant interaction between time and intervention ($F_{1.66,104.74}$=4.02, *P*=.027, n^2=.06). Holms corrected pairwise comparisons showed a significant difference between the intervention condition and control condition from baseline to after the intervention (pre-vaccination; $F_{1,63}$=4.60, adjusted *P*=.036, n^2=.07), as well as from baseline to follow-up ($F_{1,63}$=6.23, adjusted *P*=.030, n^2=.09), indicating fewer sleep problems directly after the intervention (pre-vaccination) and also at follow-up for the intervention condition compared with the control condition.

The results for positive and negative affect are shown in Multimedia Appendix 3A and Multimedia Appendix 3B, respectively. For positive affect, no significant interaction effect between time and condition was found, *P*=.69. Negative affect also yielded no significant interaction between time and condition, *P*=.15.

For wellbeing, the results are shown in Figure 5. The repeated measures ANOVA yielded a significant main effect of time ($F_{2.38,138.04}$=18.97, *P*<.001) and a significant interaction effect between time and intervention ($F_{2.38,138.04}$= 3.27, *P*=.033, n^2=.14). Holms corrected pairwise comparisons showed a significant difference between the intervention condition and control condition from baseline to the end of the test day ($F_{1,58}$=7.45, adjusted *P*=.024, n^2=.11), indicating less of a decrease in self-reported wellbeing from baseline to the end of the test day for the intervention compared with the control condition.

XSL•FO
RenderX

Figure 3. Mean and standard error of the mean of self-reported bodily sensations at baseline, after intervention (pre-vaccination), and at follow-up, separately for the control condition and intervention condition.

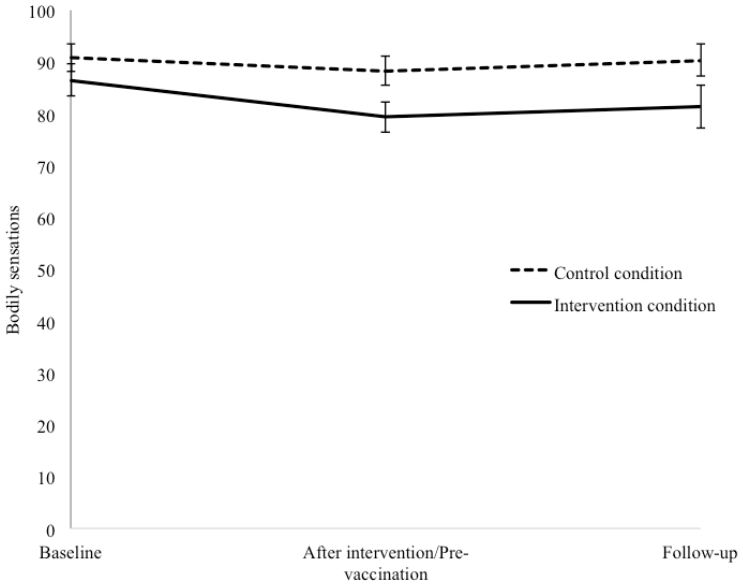
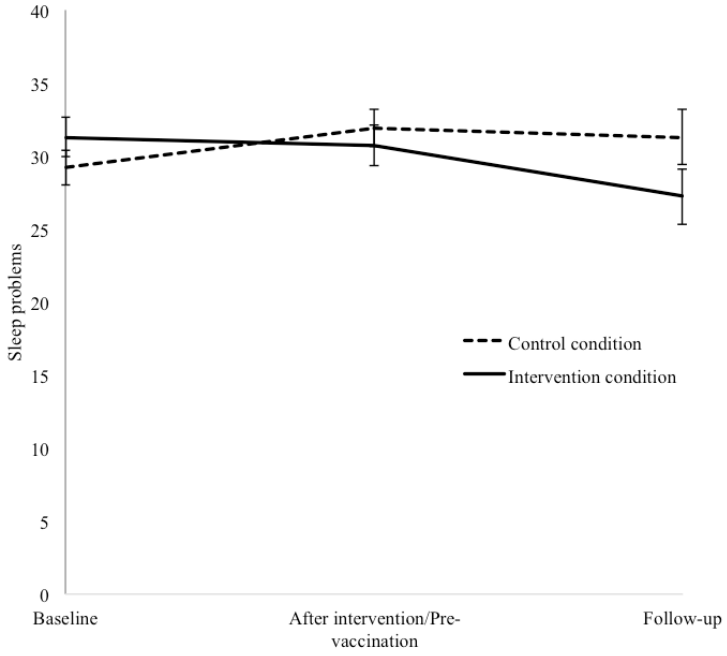
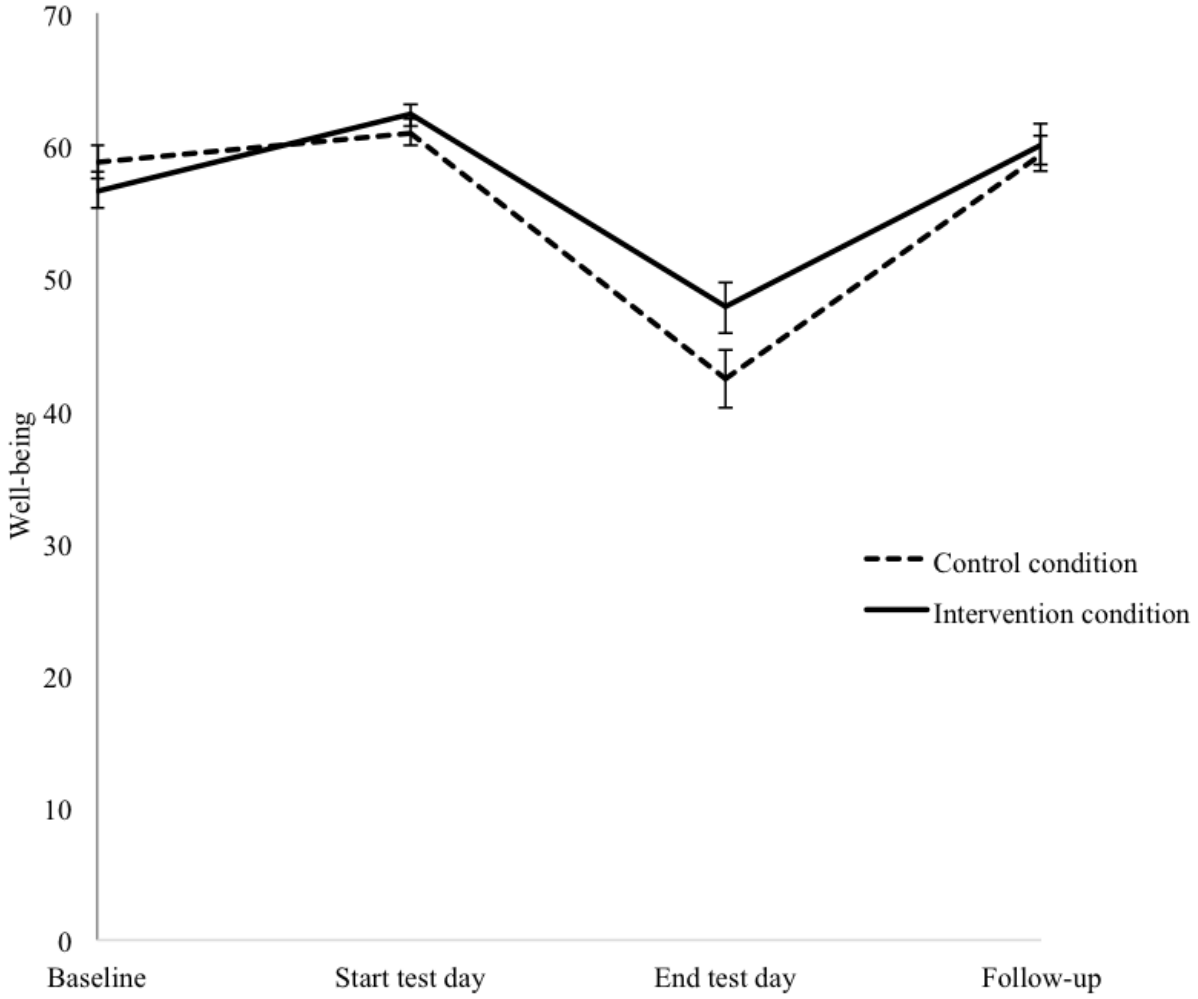


Figure 4. Mean and standard error of the mean of sleep problems at baseline, after intervention (pre-vaccination), and at follow-up, separately for the control condition and the intervention condition.



XSL•FO
RenderX

Figure 5. Mean and standard error of the mean of self-reported wellbeing at baseline, the start of the test day, the end of the test day, and at follow-up, separately for the control condition and intervention condition.



Psychophysiological Endpoints

Multimedia Appendix 4 shows the descriptive statistics for HR, skin conductance, HRV, cortisol, and alpha amylase for the control and intervention groups. For cortisol, the repeated measures ANOVA yielded no significant interaction between time and condition (P=.11). Similar results were found for alpha amylase, as no significant interaction effect between time and condition was found (P=.90).

For HR, a significant main effect of time was found ($F_{2.40,132.24}$=11.37, P<.001). Irrespective of the conditions, Holms corrected pairwise comparisons showed a significant decrease from baseline to the end of the test day (t_{56}=–3.78, adjusted P<.001). A trend was found for an interaction effect between time and condition ($F_{2.40,132.24}$=2.44, P=.081), indicating a lower HR at follow-up in the intervention condition compared with the control condition. For HRV, a significant main effect of time was found ($F_{1.49,80.29}$=4.74, P=.019), which varied over time. Holms corrected pairwise comparisons indicated no significant differences over time. No significant interaction effect was found between time and condition (P=.15). For skin conductance, no significant main effect of time (P=.46) nor an interaction effect between time and condition was found (P=.26).

Immune Endpoints

Figure 6 shows the volcano plots of significantly upregulated and downregulated serum analytes between pre-vaccination to the end of the test day. Negative values indicate analytes that are downregulated at the end of the test day compared with pre-vaccination, and positive values indicate upregulated analytes at the end of the test day compared to pre-vaccination. Analytes with an estimated effect <0.1 were not considered, since those estimates frequently represent very small changes in cytokine levels below the detection limits of variation in technical duplicates.

The multivariate linear regressions yielded no significant differences between the intervention and control groups at any time point. We therefore exploratively investigated the kinetic changes of the control and intervention conditions individually. Within the control or intervention group, significant changes over time were identified for unique sets of analytes. For the control condition, significant increases for various cytokines and chemokines (ie, IL-2, IL-10, chemokine [C-C motif] ligand [CCL]1, CCL17, CCL19, CCL23, CCL25, CCL26, chemokine [C-X-C motif] ligand [CXCL]2, CXCL6, CXCL13, CX3CL1,

XSL•FO
RenderX

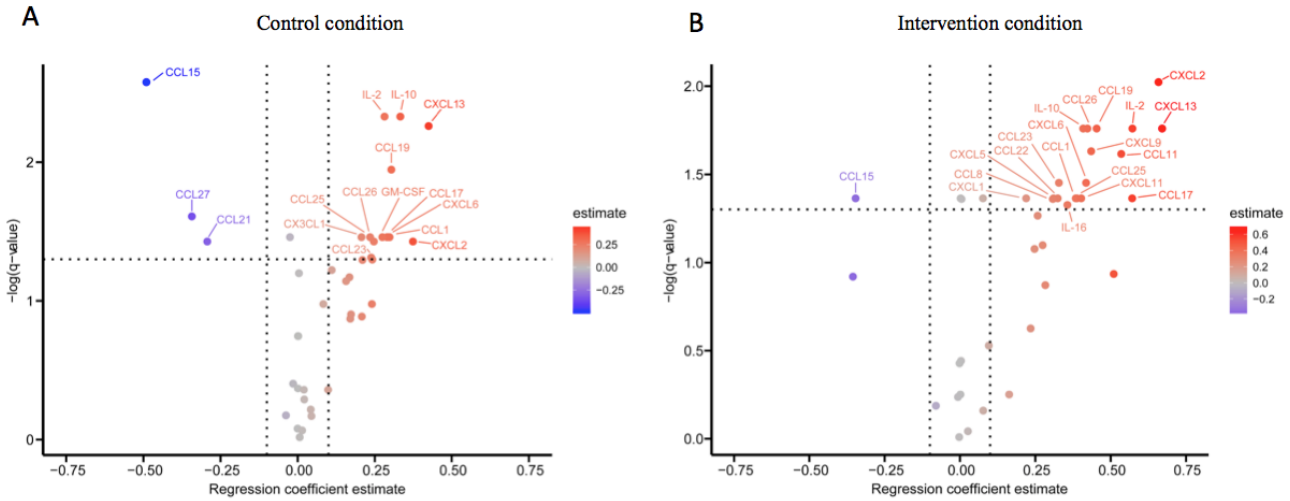
granulocyte-macrophage colony-stimulating factor) as well as significant decreases in other chemokines (ie, CCL2, CCL15, CCL21, CCL27; all FDR-corrected $P<.05$) between pre-vaccination and end of the test day were found. For the intervention condition, significant increases were also found for various cytokines and chemokines from pre-vaccination to end of the test day (ie, IL-1β, IL-2, IL-8, IL-10, IL-16, CCL1, CCL8, CCL11, CCL17, CCL19, CCL22, CCL23, CCL25, CCL26, CXCL1, CXCL2, CXCL5, CXCL6, CXCL9, CXCL11, CXCL13, macrophage migration inhibitory factor, tumor necrosis factor [TNF]-α) and a significant decrease in CCL15 (all FDR-corrected $P<.05$). The results for upregulated IL-8, CXCL5, and TNF-α as well as for downregulated CCL15 are shown in Multimedia Appendix 5, as these analytes showed the most prominent group differences. Similar results were found from the start of the test day to the end of the test day. No significant differences were found from baseline to follow-up in the control condition, although the intervention condition showed significant increases in serum IL-10, CCL19, and CXCL9 concentrations as well as a significant decrease in CCL15 (all FDR-corrected $P<.05$).

The results for the IgG antibody levels are displayed in Multimedia Appendix 6. The multivariate linear regressions yielded no significant differences between the intervention and control groups at any time point for IgG antibody levels. When looking at the within-group changes over time separately for the intervention and control conditions, no significant differences were found from baseline to follow-up in the control condition, whereas the intervention condition showed significant increases in purified protein derivative–specific IgG levels (FDR-corrected $P<.05$).

Serum C-reactive protein levels were not significantly different between groups (data not shown).

LPS stimulation of whole blood samples did not induce significant differences between the intervention and control groups (all FDR-corrected $P>.05$). In an explorative analysis, we investigated the intervention condition and control condition separately for the different time ranges. For the control condition from baseline to the test day, we found significant increases in IL-1β and TNF-α (both FDR-corrected $P<.05$), whereas no significant differences were found for the intervention condition.

Figure 6. Volcano plots for the comparison between pre-vaccination to the end of the test day for the (A) control and (B) intervention conditions. Significance is displayed on the y axis, and estimate of variance is displayed on the x axis.



Discussion

Overview

The aim of the present study was to investigate the effects of an ICBT combined with a serious gaming intervention on optimizing self-reported psychophysiological and immunological endpoints in response to in vitro and in vivo immunological and psychophysiological challenges. No significant differences between the intervention and control conditions were found for self-reported vitality. The intervention group did show fewer bodily sensations and fewer sleep problems after the intervention. Furthermore, the intervention group had higher self-reported wellbeing after different psychophysiological stressors compared with the control group. No significant group differences were found for the psychophysiological and immunological endpoints, although some preliminary support was found for improved outcomes on HR variables as well as increased IgG antibody responses at follow-up and differential chemokine endpoints at the end of the test day in the intervention compared with the control condition. The present study thus provides a first step towards unraveling the effectiveness of an electronic health psychological intervention combined with serious gaming elements on optimizing various self-reported psychophysiological and immunological health endpoints.

Primary Findings

Although the intervention condition showed improved self-reported vitality and the control condition did not, no significant group differences were found. Also, no significant group differences were found for quality of life; however, these scores were already rather high at baseline for both groups. We included a healthy population, which presumably already possessed a good quality of life that could not be maximized further by our psychological intervention. In contrast, bodily

XSL•FO
RenderX

sensations, including headache, itch, and other negative sensations, and sleep problems were significantly decreased after the intervention, compared with the control condition. This is in line with previous studies that showed that ICBT can decrease sleep problems and headache symptoms in patients with insomnia or chronic headache [52-55]. As bodily sensations and sleep problems affect general health outcomes [56,57], the intervention was effective in optimizing precursors of health. Due to the variations in the findings for bodily sensations, sleep problems, quality of life, and vitality, however, no conclusive view on the effectiveness of the intervention in improving self-reported health endpoints can be formulated. Furthermore, as the internal reliability of the Medical Outcomes Study Sleep Scale was insufficient in the present study, the outcomes on sleep should be interpreted with caution. However, possible health benefits may become especially clear when the system is challenged. The present study therefore also investigated the results of self-reported endpoints in response to in vitro and in vivo immunological as well as psychophysiological challenges. Although no significant differences were found between conditions in positive and negative affect, a higher self-reported wellbeing was found at the end of the test day for the intervention condition compared with the control condition. It is possible that the healthy population included here already possessed sufficient resilience and skills to handle the immunological and psychophysiological challenges. Future studies should therefore also include participants at risk for health problems, including participants with chronic somatic conditions or with (sub)clinical levels of anxiety or depression to see whether they may benefit from such a psychological intervention [58].

Secondary Findings

When specifically assessing the psychophysiological health endpoints (ie, HR, HRV, skin conductance, cortisol, and alpha amylase), preliminary evidence for improved endpoints after the intervention was found. Particularly, the intervention condition had a lower HR at follow-up as compared with the control condition. Although not significant, the results for HRV showed a similar pattern. As lower HR and higher HRV can be seen as biomarkers for better stress-related health outcomes [59-61], these data cautiously support the effectiveness of the psychological intervention on optimizing health. However, no significant effects were found for skin conductance, cortisol, and alpha amylase. The results therefore provide limited support for optimizing the response of the sympathetic-adrenal-medullar axis, but no support for influencing the hypothalamic-pituitary-adrenal axis, whereas the sympathetic-adrenal-medullar axis and hypothalamic-pituitary-adrenal axis are known to interact with each other in order to keep allostasis [62]. In addition, as no group differences were found on the test day for HR, HRV, cortisol, and alpha amylase, more research is needed on the external validity and clinical relevance of the present findings on psychophysiological health endpoints.

For the immune endpoints, the between-group analyses yielded no significant findings. The explorative analyses showed significant alterations in several cytokines and chemokines from baseline to follow-up in the intervention condition, whereas no significant alterations were found in the control condition between these time points, providing some cautious support for higher responses for most analytes at the follow-up in the intervention condition. Previous literature on the effectiveness of psychological interventions on optimizing immune function have not yet focused specifically on cytokines and chemokines [6]. Cytokines and chemokines are known to have a significant influence on inflammatory processes, as they provide directional cues for the movement and tissue homing of leukocytes [63,64]. To make more conclusive statements on the effectiveness of psychological interventions in optimizing chemokine functioning, future research should incorporate a wide range of analytes with varying immunological characteristics into the study design, in order to replicate the present findings and to gather more insights in the mechanisms underlying differential immune responses after a psychological intervention.

Concerning the in vivo challenge (ie, the BCG vaccination), we found increased IgG antibody levels from baseline to follow-up for the intervention condition, whereas no such significant differences were observed in the control condition. This finding provides some preliminary support for an altered host response to the BCG vaccine after the intervention. This preliminary finding is in line with a previous study from Petrie et al [20] who found higher antibody levels in response to a hepatitis B vaccine in the intervention condition receiving an emotional disclosure intervention compared to a control condition receiving no intervention. In contrast to a hepatitis B vaccine, the BCG vaccine, being a live vaccine, actually is a human challenge model and as such approximates immune responses that are observed after natural infections [23]. Since antibody titers in the present study were not different in the between-group analyses, the findings need to be interpreted with caution. The present study was the first to incorporate BCG vaccination, and future studies incorporating BCG into the study design should provide further insights into the effects of training towards this infectious challenge.

When looking at the in vitro immunological challenge, the between-group analyses on LPS-stimulated cytokines and chemokines yielded no significant differences. In exploratory analyses, we found that IL-1β, IL-8, CXCL5, and TNF-α were significantly increased from pre-vaccination to the start of the test day in the intervention but not in the control group. Furthermore, CCL2, CCL21, and CCL27 were significantly decreased from pre-vaccination to end of the test day only in the control group but not in the intervention group. Those findings suggest differential immune activation between the groups. However, the data do not support altered immune function following a psychological intervention in response to LPS as an in vitro immunological challenge. Moreover, LPS is a rather strong immune activator, possibly having masked subtle immunologic differences between the intervention and control groups.

Limitations

Despite the innovative features of the present study (ie, the combination of innovative intervention components directed at both automatic and conscious information processing and behavior change, multiple in vitro and in vivo immunological

XSL•FO
RenderX

and psychophysiological challenges, as well as the inclusion of a wide range of self-reported and psychophysiological endpoints), it has some limitations that should be mentioned. First, the present study population consisted of healthy men between 18 and 35 years of age. This represents a homogeneous healthy sample; however, future research should investigate whether the intervention might be (more) effective in other (at-risk) populations. Second, the present study design does not allow us to unravel the effectiveness of separate intervention components or separate challenges. Previously, preliminary support for the effectiveness of serious games on virtual food choice and implicit food preference was found [65]. Future research could further investigate the add-on effectiveness of serious games in optimizing engagement with the intervention and subsequent health endpoints by comparing ICBT with serious games versus ICBT alone. Moreover, future studies may investigate the effectiveness of serious games for adherence to the ICBT treatment. Third, although we tried to keep track of the time participants spent on the serious game by saving log files of the gaming activity, those log files were saved offline by participants themselves, and we did not receive log files from each participant, meaning that we could not verify whether they actually played the game 5 days a week. Although the therapist that guided the intervention tried to keep track of the gaming frequency by asking participants to report on their gaming activities in the online electronic health intervention, future studies should attempt to receive live tracking via online electronic records. Fourth, although we blinded the test leader to group allocation on the test day, the test leaders for the other measurement points were not blinded to group allocation. Although we do not have any indications that this has influenced our results, this cannot be excluded. Finally, although we asked participants not to use drugs and alcohol 48 hours before each measurement and we checked this by verbally asking them whether they used alcohol or drugs, we cannot be entirely sure that participants did not violate these rules. As consumption of alcohol and drugs can alter cytokine responses [66], future research should include quantification of alcohol and drug consumption with objective tests.

Conclusion

In conclusion, although the present study did not find support for the optimization of vitality, it did find some support for the effectiveness of an ICBT combined with a serious gaming intervention in decreasing bodily sensations and sleep problems. Also, the present study showed that the intervention participants had higher levels of self-reported wellbeing in response to the psychophysiological challenges than control participants. Additionally, specific IgG antibody levels were increased at 4 weeks after BCG vaccination in the intervention condition. As this is one of the first studies incorporating multiple challenges to evaluate the effects of a psychological intervention on health endpoints, the present study provides a first step towards improving health endpoints with a psychological intervention, although clearly more research is needed on this topic. Future research should further investigate whether tailoring the intervention to specific populations, including patients with chronic somatic conditions or participants with (sub)clinical levels of stress or anxiety issues, enhances efficacy and impacts relevant disease-related parameters and biomarkers. Given the innovative study design, combining multiple new elements, future studies should consistently incorporate challenges and a wide range of immune parameters into the study design in order to get a more complete view on the effects of innovative psychological interventions.

Acknowledgments

The authors acknowledge Chantal Eckhardt, Dion de Hoog, Sander van den Oever, Shirley de Wit, and Rafael Bidarra of Delft University of Technology for their help in designing and developing the serious game ViaNova©. Additionally, the authors acknowledge Anne Bruinings, Paige Crompvoets, Annemarie Danton, Lianne Keuning, Jelle van Leusden, Miranda Lutz, Meriem Manai, Rebecca Maurits, Stefanie Meeuwis, Kaya Peerdeman, Maaike de Ronde, Aleksandrina Skvortsova, Ikrame Tajioui, and Judith Tekampe for their help with collecting the data. This work was supported by a European Research Council Consolidator Grant (ERC-2013-CoG-617700_EXPECT HEAL-TH) and an NWO Vici Grant (016.Vici.170.152), both granted to AWME. The funders had no role in the data acquisition, data analyses, or the content of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of chemokines and other cytokines that were analyzed in the 40-plex assay.
[DOCX File , 13 KB - jmir_v22i7e14861_app1.docx]

Multimedia Appendix 2
Mean and standard error of the mean of self-reported physical quality of life (A) and mental quality of life (B) T-scores at baseline, after intervention (pre-vaccination), and at follow-up, separately for the control condition and the intervention condition. A higher score on the y-axis represents a higher quality of life.
[PNG File , 15 KB - jmir_v22i7e14861_app2.png]

XSL•FO
RenderX

Multimedia Appendix 3
Mean and standard error of the mean of the standardized scores for self-reported positive affect (A) and negative affect (B) at baseline, the start of the test day, the end of the test day, and at follow-up, separately for the control condition and the intervention condition. A higher score on the y-axis represents a higher level of self-reported positive affect and negative affect, respectively.
[PNG File , 23 KB - jmir_v22i7e14861_app3.png]

Multimedia Appendix 4
Means and standard deviations for heart rate, skin conductance, heart rate variability, as well as cortisol and alpha amylase, separately for the control condition and the intervention condition.
[DOCX File , 15 KB - jmir_v22i7e14861_app4.docx]

Multimedia Appendix 5
Boxplots for the upregulated IL-8 (A), CXCL5 (B) and TNF- α (C), as well as for the downregulated CCL15 (D) for the control condition (left graph) and intervention condition (right graph) separately at baseline, pre-vaccination, start test day, end test day and follow-up. A higher level in pg/ml on the y-axis represents a higher cytokine/chemokine level.
[PNG File , 136 KB - jmir_v22i7e14861_app5.png]

Multimedia Appendix 6
Boxplots with the OD450 readings for the control condition (left graph) and intervention condition (right graph) separately with the PPD specific IgG antibody levels at baseline and follow-up. A higher OD450 reading on the y-axis represents a higher IgG antibody level.
[PNG File , 42 KB - jmir_v22i7e14861_app6.png]

Multimedia Appendix 7
CONSORT‐eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2201 KB - jmir_v22i7e14861_app7.pdf]

References

1. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]
2. Knoerl R, Lavoie Smith EM, Weisberg J. Chronic Pain and Cognitive Behavioral Therapy: An Integrative Review. West J Nurs Res 2016 May;38(5):596-628. [doi: 10.1177/0193945915615869] [Medline: 26604219]
3. Rosenkranz MA, Davidson RJ, Maccoon DG, Sheridan JF, Kalin NH, Lutz A. A comparison of mindfulness-based stress reduction and an active control in modulation of neurogenic inflammation. Brain Behav Immun 2013 Jan;27(1):174-184 [FREE Full text] [doi: 10.1016/j.bbi.2012.10.013] [Medline: 23092711]
4. Witt K. Psychological Treatment Can Modulate the Skin Reaction to Histamine in Pollen Allergic Humans. Dermatol Psychosom 2003 Apr 4;4(1):33-37. [doi: 10.1159/000070533]
5. Robinson H, Jarrett P, Vedhara K, Broadbent E. The effects of expressive writing before or after punch biopsy on wound healing. Brain Behav Immun 2017 Mar;61:217-227. [doi: 10.1016/j.bbi.2016.11.025] [Medline: 27890660]
6. Miller GE, Cohen S. Psychological interventions and the immune system: A meta-analytic review and critique. Health Psychology 2001;20(1):47-63. [doi: 10.1037/0278-6133.20.1.47]
7. Schakel L, Veldhuijzen D, Crompvoets P, Bosch J, Cohen S, van Middendorp H, et al. Effectiveness of Stress-Reducing Interventions on the Response to Challenges to the Immune System: A Meta-Analytic Review. Psychother Psychosom 2019 Aug 6;88(5):274-286 [FREE Full text] [doi: 10.1159/000501645] [Medline: 31387109]
8. de Ridder D, Geenen R, Kuijer R, van Middendorp H. Psychological adjustment to chronic disease. The Lancet 2008 Jul;372(9634):246-255. [doi: 10.1016/s0140-6736(08)61078-8]
9. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014 Mar 27;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
10. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
11. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
12. Ferwerda M, van Beugen S, van Riel PC, van de Kerkhof PC, de Jong EM, Smit JV, et al. Measuring the Therapeutic Relationship in Internet-Based Interventions. Psychother Psychosom 2016;85(1):47-49. [doi: 10.1159/000435958]

XSL•FO
RenderX

13. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious Games and Gamification for Mental Health: Current Status and Promising Directions. Front Psychiatry 2016 Jan 10;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
14. Kato PM. Video Games in Health Care: Closing the Gap. Review of General Psychology 2010 Jun;14(2):113-121. [doi: 10.1037/a0019441]
15. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
16. Schakel L, Veldhuijzen DS, Manai M, Beugen SV, Vaart RVD, Middendorp HV, et al. Editor's choice: Optimizing healthy food preferences by serious gaming. Psychol Health 2020 Apr 12;35(4):405-424. [doi: 10.1080/08870446.2019.1675657] [Medline: 31607172]
17. Schakel L, Veldhuijzen D, Crompvoets P, Bosch J, Cohen S, van Middendorp H, et al. Effectiveness of Stress-Reducing Interventions on the Response to Challenges to the Immune System: A Meta-Analytic Review. Psychother Psychosom 2019 Aug 6;88(5):274-286 [FREE Full text] [doi: 10.1159/000501645] [Medline: 31387109]
18. Elsenbruch S, Langhorst J, Popkirowa K, Müller T, Luedtke R, Franken U, et al. Effects of mind-body therapy on quality of life and neuroendocrine and cellular immune functions in patients with ulcerative colitis. Psychother Psychosom 2005 Aug 10;74(5):277-287. [doi: 10.1159/000086318] [Medline: 16088265]
19. Kiecolt-Glaser JK, Glaser R, Williger D, Stout J, Messick G, Sheppard S, et al. Psychosocial enhancement of immunocompetence in a geriatric population. Health Psychology 1985;4(1):25-41. [doi: 10.1037/0278-6133.4.1.25]
20. Petrie KJ, Booth RJ, Pennebaker JW, Davison KP, Thomas MG. Disclosure of trauma and immune response to a hepatitis B vaccination program. Journal of Consulting and Clinical Psychology 1995;63(5):787-792. [doi: 10.1037/0022-006x.63.5.787]
21. Koschwanez H, Robinson H, Beban G, MacCormick A, Hill A, Windsor J, et al. Randomized clinical trial of expressive writing on wound healing following bariatric surgery. Health Psychol 2017 Jul;36(7):630-640. [doi: 10.1037/hea0000494] [Medline: 28383927]
22. Robinson H, Norton S, Jarrett P, Broadbent E. The effects of psychological interventions on wound healing: A systematic review of randomized trials. Br J Health Psychol 2017 Nov 03;22(4):805-835. [doi: 10.1111/bjhp.12257] [Medline: 28670818]
23. Schakel L, Veldhuijzen DS, van Middendorp H, Prins C, Joosten SA, Ottenhoff THM, et al. The effects of a psychological intervention directed at optimizing immune function: study protocol for a randomized controlled trial. Trials 2017 May 26;18(1):243. [doi: 10.1186/s13063-017-1983-7]
24. Oertelt-Prigione SIMC. Immunology and the menstrual cycle. Autoimmun Rev 2012 May;11(6-7):A486-A492. [doi: 10.1016/j.autrev.2011.11.023] [Medline: 22155200]
25. Ferwerda M, van Beugen S, van Middendorp H, Spillekom-van Koulil S, Donders ART, Visser H, et al. A tailored-guided internet-based cognitive-behavioral intervention for patients with rheumatoid arthritis as an adjunct to standard rheumatological care: results of a randomized controlled trial. Pain 2017 May;158(5):868-878. [doi: 10.1097/j.pain.0000000000000845] [Medline: 28106666]
26. van Beugen S, Ferwerda M, Spillekom-van Koulil S, Smit JV, Zeeuwen-Franssen ME, Kroft EB, et al. Tailored Therapist-Guided Internet-Based Cognitive Behavioral Treatment for Psoriasis: A Randomized Controlled Trial. Psychother Psychosom 2016 Aug 11;85(5):297-307 [FREE Full text] [doi: 10.1159/000447267] [Medline: 27508937]
27. Klainin-Yobas P, Oo WN, Suzanne Yew PY, Lau Y. Effects of relaxation interventions on depression and anxiety among older adults: a systematic review. Aging & Mental Health 2015 Jan 09;19(12):1043-1055. [doi: 10.1080/13607863.2014.997191]
28. Schakel L, Veldhuijzen DS, Middendorp HV, Dessel PV, Houwer JD, Bidarra R, et al. The effects of a gamified approach avoidance training and verbal suggestions on food outcomes. PLoS One 2018 Jul 26;13(7):e0201309 [FREE Full text] [doi: 10.1371/journal.pone.0201309] [Medline: 30048511]
29. Paine NJ, Bosch JA, Ring C, Drayson MT, Veldhuijzen van Zanten JJCS. Induced mild systemic inflammation is associated with impaired ability to improve cognitive task performance by practice. Psychophysiology 2015 Mar;52(3):333-341. [doi: 10.1111/psyp.12360] [Medline: 25366393]
30. van Laarhoven AI, Kraaimaat FW, Wilder-Smith OH, van de Kerkhof PC, Evers AW. Heterotopic pruritic conditioning and itch--analogous to DNIC in pain? Pain 2010 May;149(2):332-337. [doi: 10.1016/j.pain.2010.02.026] [Medline: 20226590]
31. Kirschbaum C, Pirke K, Hellhammer DH. The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004]
32. Yamakawa K, Matsunaga M, Isowa T, Kimura K, Kasugai K, Yoneda M, et al. Transient responses of inflammatory cytokines in acute stress. Biol Psychol 2009 Sep;82(1):25-32. [doi: 10.1016/j.biopsycho.2009.05.001] [Medline: 19446599]
33. Paine NJ, Ring C, Bosch JA, Drayson MT, Aldred S, Veldhuijzen van Zanten JJ. Vaccine-induced inflammation attenuates the vascular responses to mental stress. Int J Psychophysiol 2014 Sep;93(3):340-348. [doi: 10.1016/j.ijpsycho.2014.06.016] [Medline: 24998644]

XSL•FO
RenderX

34. Ryan RM, Frederick C. On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. J Pers 1997 Sep;65(3):529-565. [doi: 10.1111/j.1467-6494.1997.tb00326.x] [Medline: 9327588]
35. Vercoulen J, Alberts M. Bleijenberg G: De checklist individuele spankracht (CIS). Gedragstherapie 1999;32:131-136.
36. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. Journal of Psychosomatic Research 1994 Jul;38(5):383-392. [doi: 10.1016/0022-3999(94)90099-x]
37. Worm-Smeitink M, Gielissen M, Bloot L, van Laarhoven H, van Engelen B, van Riel P, et al. The assessment of fatigue: Psychometric qualities and norms for the Checklist individual strength. J Psychosom Res 2017 Jul;98:40-46. [doi: 10.1016/j.jpsychores.2017.05.007] [Medline: 28554371]
38. Salama-Younes M, Montazeri A, Ismaïl A, Roncin C. Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: a study from France. Health Qual Life Outcomes 2009 Mar 05;7(1):22 [FREE Full text] [doi: 10.1186/1477-7525-7-22] [Medline: 19265516]
39. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Annals of Medicine 2001;33(5):350-357. [doi: 10.3109/07853890109002089]
40. VanderZee KI, Sanderman R, Heyink JW, Haes H. Psychometric qualities of the RAND 36-item health survey 1.0: A multidimensional measure of general health status. Int. J. Behav. Med 1996 Jun;3(2):104-122. [doi: 10.1207/s15327558ijbm0302_2]
41. Pennebaker JW. The Psychology of Physical Symptoms. New York: Springer Science & Business Media; 2012.
42. Allen RP, Kosinski M, Hill-Zabala CE, Calloway MO. Psychometric evaluation and tests of validity of the Medical Outcomes Study 12-item Sleep Scale (MOS sleep). Sleep Med 2009 May;10(5):531-539. [doi: 10.1016/j.sleep.2008.06.003] [Medline: 18805054]
43. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
44. Crawford J, Henry JD. The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. British journal of clinical psychology 2004;43(3):245-265. [doi: 10.1348/0144665031752934]
45. Sjak-Shie EE. PhysioData Toolbox (Version 0.4). 2018. URL: https://physiodatatoolbox.leidenuniv.nl/
46. van Meijgaarden KE, Khatri B, Smith SG, Drittij AMFH, de Paus RA, Goeman JJ, et al. Cross-laboratory evaluation of multiplex bead assays including independent common reference standards for immunological monitoring of observational and interventional human studies. PLoS One 2018 Sep 4;13(9):e0201205 [FREE Full text] [doi: 10.1371/journal.pone.0201205] [Medline: 30180167]
47. Joosten SA, van Meijgaarden KE, Del Nonno F, Baiocchini A, Petrone L, Vanini V, et al. Patients with Tuberculosis Have a Dysfunctional Circulating B-Cell Compartment, Which Normalizes following Successful Treatment. PLoS Pathog 2016 Jun 15;12(6):e1005687 [FREE Full text] [doi: 10.1371/journal.ppat.1005687] [Medline: 27304615]
48. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 2017 Nov 3;13(11):e1005752 [FREE Full text] [doi: 10.1371/journal.pcbi.1005752] [Medline: 29099853]
49. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1):14065823-14062014. [doi: 10.18637/jss.v067.i01]
50. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Soft 2017;82(13):1. [doi: 10.18637/jss.v082.i13]
51. Wickham H. ggplot2: elegant graphics for data analysis. Switzerland: Springer; 2016.
52. Ström L, Pettersson R, Andersson G. Internet-Based Treatment for Insomnia: A Controlled Evaluation. Journal of Consulting and Clinical Psychology 2004;72(1):113-120. [doi: 10.1037/0022-006x.72.1.113]
53. Sato D, Yoshinaga N, Nagai E, Nagai K, Shimizu E. Effectiveness of Internet-Delivered Computerized Cognitive Behavioral Therapy for Patients With Insomnia Who Remain Symptomatic Following Pharmacotherapy: Randomized Controlled Exploratory Trial. J Med Internet Res 2019 Apr 11;21(4):e12686 [FREE Full text] [doi: 10.2196/12686] [Medline: 30973344]
54. Devineni T, Blanchard EB. A randomized controlled trial of an internet-based treatment for chronic headache. Behav Res Ther 2005 Mar;43(3):277-292. [doi: 10.1016/j.brat.2004.01.008] [Medline: 15680926]
55. Ström L, Pettersson R, Andersson G. A controlled trial of self-help treatment of recurrent headache conducted via the Internet. Journal of Consulting and Clinical Psychology 2000;68(4):722-727. [doi: 10.1037/0022-006x.68.4.722]
56. Pennebaker JW, Skelton JA. Psychological Parameters of Physical Symptoms. Pers Soc Psychol Bull 1978;4(4):524-530. [doi: 10.1177/014616727800400405]
57. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]
58. Evers AW, Gieler U, Hasenbring MI, van Middendorp H. Incorporating biopsychosocial characteristics into personalized healthcare: a clinical approach. Psychother Psychosom 2014;83(3):148-157. [doi: 10.1159/000358309] [Medline: 24732828]
59. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol. Med 2017 May 08;47(15):2578-2586. [doi: 10.1017/s0033291717001003]

XSL•FO
RenderX

60. Thayer JF, Ahs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012 Feb;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
61. Woody A, Hamilton K, Livitz IE, Figueroa WS, Zoccola PM. Buccal telomere length and its associations with cortisol, heart rate variability, heart rate, and blood pressure responses to an acute social evaluative stressor in college students. Stress 2017 May 29;20(3):249-257. [doi: 10.1080/10253890.2017.1328494] [Medline: 28482730]
62. McEwen B. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci 1998 May 01;840(1):33-44. [doi: 10.1111/j.1749-6632.1998.tb09546.x] [Medline: 9629234]
63. Luster AD. Chemokines--chemotactic cytokines that mediate inflammation. N Engl J Med 1998 Feb 12;338(7):436-445. [doi: 10.1056/NEJM199802123380706] [Medline: 9459648]
64. Palomino DCT, Marti LC. Chemokines and immunity. Einstein (São Paulo) 2015 Sep;13(3):469-473. [doi: 10.1590/s1679-45082015rb3438]
65. Schakel L, Veldhuijzen DS, Middendorp HV, Dessel PV, Houwer JD, Bidarra R, et al. The effects of a gamified approach avoidance training and verbal suggestions on food outcomes. PLoS One 2018 Jul 26;13(7):e0201309 [FREE Full text] [doi: 10.1371/journal.pone.0201309] [Medline: 30048511]
66. Karoly HC, Bidwell LC, Mueller RL, Hutchison KE. Investigating the Relationships Between Alcohol Consumption, Cannabis Use, and Circulating Cytokines: A Preliminary Analysis. Alcohol Clin Exp Res 2018 Mar 02;42(3):531-539 [FREE Full text] [doi: 10.1111/acer.13592] [Medline: 29286537]

Abbreviations

ANCOVA: analysis of covariance.
ANOVA: analysis of variance.
BCG: bacillus Calmette-Guérin.
CBT: cognitive behavioral therapy.
CCL: chemokine (C-C motif) ligand.
CIS-20: Checklist Individual Strength.
CXCL: chemokine (C-X-C motif) ligand.
FDR: false discovery rate.
HR: heart rate.
HRV: heart rate variability.
ICBT: internet-based cognitive behavioral therapy.
IgG: immunoglobulin G.
IL: interleukin.
LPS: lipopolysaccharide.
MOS Sleep: Medical Outcomes Study Sleep Scale.
NRS: numeric rating scale.
PANAS: Positive and Negative Affect Schedule.
PILL: Pennebaker Inventory of Limbic Languidness.
SVS: Subjective Vitality Scale.
TNF: tumor necrosis factor.

Edited by G Eysenbach; submitted 05.06.19; peer-reviewed by M Buhrman, K Matsumoto, M Vugts; comments to author 03.10.19; revised version received 05.02.20; accepted 22.02.20; published 24.07.20.

Please cite as:
Schakel L, Veldhuijzen DS, van Middendorp H, Prins C, Drittij AMHF, Vrieling F, Visser LG, Ottenhoff THM, Joosten SA, Evers AWM
An Internet-Based Psychological Intervention With a Serious Game to Improve Vitality, Psychological and Physical Condition, and Immune Function in Healthy Male Adults: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e14861
URL: https://www.jmir.org/2020/7/e14861
doi:10.2196/14861
PMID:32706667

©Lemmy Schakel, Dieuwke S Veldhuijzen, Henriët van Middendorp, Corine Prins, Anne MHF Drittij, Frank Vrieling, Leo G Visser, Tom HM Ottenhoff, Simone A Joosten, Andrea WM Evers. Originally published in the Journal of Medical Internet

XSL•FO
RenderX

Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Diabetes-Related Topics in an Online Forum for Caregivers of Individuals Living With Alzheimer Disease and Related Dementias: Qualitative Inquiry

Yan Du[1], PhD, MPH, RN; Kristi Paiva[1], MPH; Adrian Cebula[1], MS; Seon Kim[1], BS; Katrina Lopez[1], BA; Chengdong Li[1], PhD; Carole White[2], PhD, RN; Sahiti Myneni[3], PhD; Sudha Seshadri[4], MD; Jing Wang[1], PhD, MPH, RN, FAAN

[1]Center on Smart and Connected Health Technologies, School of Nursing, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[2]School of Nursing, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[3]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[4]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Corresponding Author:
Jing Wang, PhD, MPH, RN, FAAN
Center on Smart and Connected Health Technologies
School of Nursing
The University of Texas Health Science Center at San Antonio
7739 Floyd Drive
San Antonio, TX, 78229
United States
Phone: 1 210 450 8561
Email: wangj1@uthscsa.edu

Abstract

Background: Diabetes and Alzheimer disease and related dementias (ADRD) are the seventh and sixth leading causes of death in the United States, respectively, and they coexist in many older adults. Caring for a loved one with both ADRD and diabetes is challenging and burdensome.

Objective: This study aims to explore diabetes-related topics in the Alzheimer's Association ALZConnected caregiver forum by family caregivers of persons living with ADRD.

Methods: User posts on the Alzheimer's Association ALZConnected caregiver forum were extracted. A total of 528 posts related to diabetes were included in the analysis. Of the users who generated the 528 posts, approximately 96.1% (275/286) were relatives of the care recipient with ADRD (eg, child, grandchild, spouse, sibling, or unspecified relative). Two researchers analyzed the data independently using thematic analysis. Any divergence was discussed among the research team, and an agreement was reached with a senior researcher's input as deemed necessary.

Results: Thematic analysis revealed 7 key themes. The results showed that comorbidities of ADRD were common topics of discussions among family caregivers. Diabetes management in ADRD challenged family caregivers. Family caregivers might neglect their own health care because of the caring burden, and they reported poor health outcomes and reduced quality of life. The online forum provided a platform for family caregivers to seek support in their attempts to learn more about how to manage the ADRD of their care recipients and seek support for managing their own lives as caregivers.

Conclusions: The ALZConnected forum provided a platform for caregivers to seek informational and emotional support for caring for persons living with ADRD and diabetes. The overwhelming burdens with these two health conditions were apparent for both caregivers and care recipients based on discussions from the online forum. Studies are urgently needed to provide practical guidelines and interventions for diabetes management in individuals with diabetes and ADRD. Future studies to explore delivering diabetes management interventions through online communities in caregivers and their care recipients with ADRD and diabetes are warranted.

(J Med Internet Res 2020;22(7):e17851) doi:10.2196/17851

XSL•FO
RenderX

KEYWORDS

diabetes; Alzheimer disease; dementia; caregivers

Introduction

Background

Diabetes and Alzheimer disease and related dementias (ADRD) are the seventh and sixth leading causes of death in the United States, respectively, and they coexist in many older adults [1,2]. Both diseases are major public health concerns in today's aging world, and the combination of the two has further burdened individuals with the two diseases, their caregivers, and the society at large. As of 2019, 5.8 million Americans are estimated to live with Alzheimer dementia, and the population of people living with the disease is expected to triple by 2060 [3]. The total annual cost of managing and treating ADRD alone is estimated to be upward of US $215 billion and is expected to increase to US $500 billion by 2040 [3]. Likewise, diabetes affects roughly 30 million people in the United States as of 2015 and is expected to grow to affect approximately 55 million people between 2015 and 2030. Costs associated with the care and management of diabetes rose from US $245 billion in 2012 to US $327 billion in 2017 [2]. Together, the cost to care for persons with ADRD and diabetes comorbidity is higher than the cost to care for either disease alone. For example, the mean cost of diabetes care in persons with ADRD is approximately 3 times higher than the cost of diabetes care in those without ADRD [4].

A significant number of individuals living with ADRD have comorbid conditions, and diabetes is one of the most common comorbidities [5]. Increasing studies have documented that there is an association between ADRD and diabetes, and researchers have proposed that Alzheimer disease is *type 3 diabetes* because of the shared pathological mechanisms [6,7]. Moreover, the onset of diabetes may affect cognitive performance. For instance, higher levels of glycosylated hemoglobin and higher average blood glucose levels are correlated with poorer cognitive performance [8,9]. According to a longitudinal study, those who had type 2 diabetes at baseline demonstrated accelerated cognitive decline at 12-year follow-up [10]. Possible underlying mechanisms for this cognitive decline have been linked to insulin resistance, insulin-degrading enzyme, and so forth. Meanwhile, some studies suggest that there is a bidirectionality between the two diseases [11,12]. Previous studies have indicated that Alzheimer disease may affect the diabetic phenotype through behavioral changes, such as binge eating, or pathological changes, such as amyloid-beta deposition and tau protein phosphorylation [12].

Caring for Individuals With ADRD and Diabetes

Although the mechanisms connecting these two diseases are still largely unknown, the implications are important as ADRD can affect the self-management of diabetes and vice versa. Both ADRD and diabetes require a unique and sometimes extensive amount of care. The management of diabetes alone involves constant blood glucose monitoring, the discipline to adhere to dietary restrictions and physical activity guidelines, and proper antidiabetic medication use [13]. Poor glycemic control can result in neuronal damage, increased ADRD incidence, and exaggerated cognitive function decline [12]. For those living with ADRD, the complexity of these regimens is often exacerbated by disease progression. Increased forgetfulness and confusion associated with the mid to late stages of dementia can result in improper meal choices or medication mismanagement [14]. This inability to participate in diabetes management as clinically instructed can result in harmful effects on the renal, cardiovascular, and peripheral nervous systems. As a result, people with ADRD may require assistance with the management of instrumental activities of daily living, such as diabetes management, from a caregiver.

Caring for persons with ADRD is often assumed by family members and close friends, with 83% of help to older adults in the United States being provided by these unpaid caregivers (referred to as family caregivers in this study) [1,15]. Recent estimates show that about 16 million family members and friends provided over 18 billion hours of unpaid care to people with ADRD at an average of 21.9 hours of care per week per caregiver [1]. This time spent caring for a loved one often comes at the expense of a caregiver's own needs, with some sacrificing vacations, hobbies, and quality time with other family members [16]. In addition, the burden of caregiving is shown to be higher in caregivers of people with ADRD as compared with caregivers of older adults without ADRD, and the impact of this can be both physical and emotional [1,17]. For example, 40% of these family ADRD caregivers suffer from depression, and 1 in 3 family caregivers reported worsening health since assuming caregiving responsibilities [1]. Moreover, it has been shown that ADRD caregiving is associated with elevated biomarkers correlating to cardiovascular disease risk and that hormones associated with stress can have negative effects on glucose metabolism [18,19]. However, a recent study that reviewed 89 papers concluded that despite personalized, continuity, and family-centered care being urgently needed when caring for people living with ADRD and diabetes, current health care systems cannot meet the caring needs [20]. Individuals with ADRD and their caregivers constantly struggle with managing diabetes and ADRD.

To mitigate the burden associated with caring for a loved one with ADRD, some caregivers turn to social media to discuss patient care and their personal struggles as they adjust to their everchanging responsibilities as caregivers. For those living with diabetes, social media allows for additional communication and discussion outside of the limited time allotted with their physicians while encouraging participation and engagement [21]. Whether being used for the discussion of personal health or to discuss the care of a loved one, social media connects caregivers not only to like-minded individuals but also to health information and tools that may have otherwise been unavailable to them [22]. ALZConnected [23], a free online forum run by the Alzheimer's Association in the United States, offers a range of forums for people with ADRD and their caregivers or friends and family members. For some, visiting this website means not

XSL•FO
RenderX

having to explain what living with Alzheimer means, and for others, it means no longer feeling alone [23].

Social Media Data

In this digital era, social media has provided platforms for users to seek health information, receive health interventions, and anonymously share thoughts and experiences that they may not feel confident or comfortable expressing in a real-world setting [24-27]. The increasing amount of user-generated information has become a valuable resource to guide health interventions and advance our knowledge of a variety of health conditions [28]. The data can be analyzed and built upon, allowing researchers to better understand user characteristics, social and information needs, and communication ecosystems as well as to serve the demographics being studied [29-32]. Such applications include, but are not limited to, the identification of individuals at risk for depression [33], the detection of drug-drug interactions or adverse side effects [34], the examination of health behaviors in various types of cancer survivors [35,36], and the investigation of the impact of social media on decision making and recovery in prostate cancer patients [37,38]. Whether the information gathered comes from well-known social media sites such as Facebook and Twitter or from more specific online support groups such as ALZConnected, the resulting analysis may provide much-needed insight to improve health outcomes in diverse groups [39-42]. However, despite the potential of media data to complement data from regulatory, clinical, administrative, and claims data sources, the utility and effectiveness of such data remain understudied in a variety of health conditions [43].

Study Objective

This study aimed to explore diabetes-related topics in the ALZConnected caregiver forums by family caregivers of persons living with ADRD.

Methods

Setting/Study Population

In this qualitative study, we collected and analyzed posts from users of the Alzheimer's Association ALZConnected caregiver's forum in the United States [41]. The forum is designed to provide a community for ADRD caregivers and people with dementia alike. Caregivers, both professional and family, make up the majority of users in the forum and are often the spouse or child of the care recipient. Users must be registered in the ALZConnected community before they can post new threads or post in another user's thread. Users must also be registered to view a member user's profile. However, all posts can be viewed publicly without registering an account. The study was approved by the institutional review board of the University of Texas Health Science Center at San Antonio.

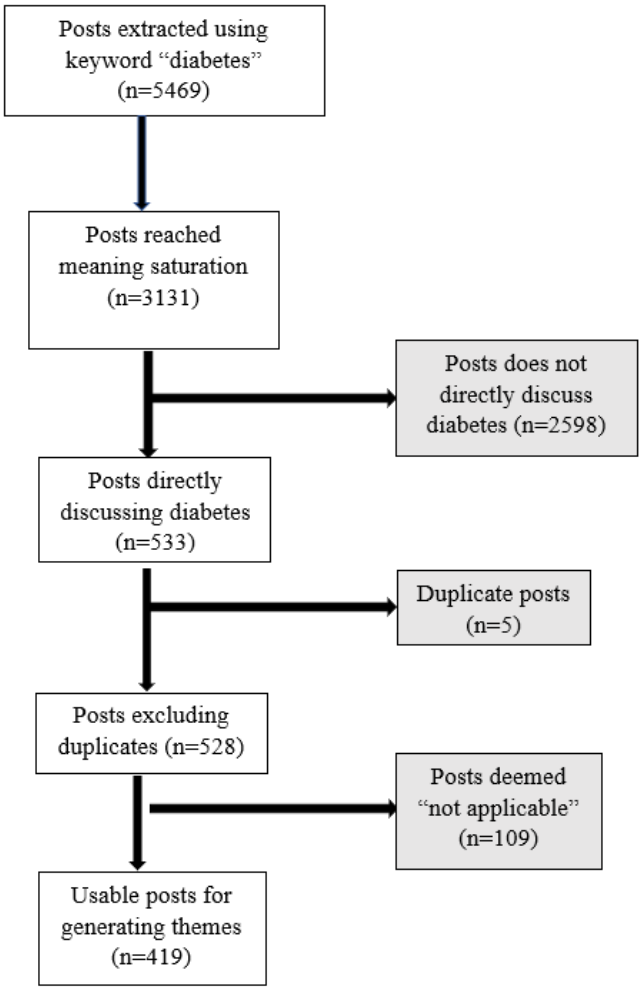
Data Collection

The earliest post in the forum was published on December 1, 2011. Posts were extracted from the forum backward in time via web scraping in January 2019. A total of 5469 user posts, posted from January 22, 2012, to January 3, 2019, were extracted using the keyword "diabetes." All extracted posts were imported into an Excel (Microsoft) file for analysis, with password protection. Due to the nature of the data and the study focus of diabetes in individuals living with ADRD in an ADRD forum, further selection of posts and preliminary data analysis were conducted simultaneously. Two trained research assistants independently reviewed posts from the beginning of the list. Codes, notes, and comments were made during this process. Saturation was reached, meaning no new ideas were presented [44], after reviewing 3131 posts. Of the 3131 posts reviewed, 2598 were excluded for not directly discussing diabetes, leaving 533 posts. An additional 5 duplicate posts were removed, leaving 528 posts for analysis. Upon further analysis and the application of codes, an additional 109 were deemed *not applicable* because of the use of diabetes for comparative purposes or without reference to a care recipient or caregiver. Posts coded as *not applicable* were excluded, leaving 419 posts included in generating themes. Figure 1 illustrates the selection scheme of the posts.

The study also obtained publicly available data on the relationship between the post generator and individuals living with ADRD, aggregately. When registering for an account in the forum, users are requested to identify their relationship with a person living with dementia (eg, the user is a person living with dementia, a type of relative, or a professional caregiver).

XSL•FO
RenderX

Figure 1. Flowchart of selection scheme of posts included in the study.



Data Analysis

"Thematic analysis is a method of identifying, analyzing, and reporting patterns (themes, topics, ideas) within a data" [45,46]. It has been widely used to identify themes or topics for web-based data (eg, blogs, web-based discussion boards, and tweets) [47-50] and data studying individuals living with ADRD [47,48,51]. Three trained research assistants analyzed the data using thematic analysis with the following steps [45,49,52]. First, data reduction and open coding were conducted during the preliminary analysis, and 2 coders developed a list of possible codes and definitions independently. The coding team and a senior qualitative scientist met to discuss and finalize the initial codes based on the feedback of all team members. Second, code validation was conducted by applying the initial codes to each of the 528 posts by 2 independent coders. Coders were instructed to code at the level of the post and were allowed to apply multiple codes to each post. Researchers coding overlapping sections of the dataset then came together in meetings to reconcile code differences in their respective analyses. They could discuss their reasoning for code applications. If in agreement, the 2 coders would combine code applications that served as the final analysis of the respective post being discussed. If the 2 coders were unable to reconcile code differences, the third coder provided insight allowing for a finalized analysis of the post. The initial codes from the first step were refined during the second step, and 26 distinct codes were generated. During this process, 109 were deemed *not applicable*, leaving 419 posts for the next step. Third, following coding reconciliation, the coding team displayed all various codes in a table and categorized them according to similarities and differences. Consideration was given to the coding processes, categorization, and noting patterns, and they came to a consensus of the key themes identified from the forum.

The coding team used descriptive statistics to obtain basic counts regarding the use of each code as well as to compare the number of instances in which 2 codes appeared together within the same thread. In addition, publicly accessible data of a user's relationship with the individual living with ADRD were assessed aggregately using descriptive statistics.

Results

Caregiver-Patient Relationship

A total of 528 posts were generated by 286 unique users posting within 399 threads. Approximately 70% (200/286) of these users were the child of an individual with ADRD for whom they were caring for or concerned, whereas only 17.1% (49/286) of the users were spouses who were caring for a loved one with ADRD. Overall, the vast majority of these users (275/286, 96.1%) were related to the care recipient with ADRD in some

XSL•FO
RenderX

way, that is, a child, grandchild, spouse, sibling, or unspecified relative. In contrast, very few users were professional caregivers (6/286, 2.1%) or someone living with ADRD (3/286, 1.0%).

Themes

Thematic analysis revealed 7 key themes: (1) disease linkages and comorbidities, (2) ADRD or diabetes symptoms without diagnosis, (3) diabetes management, (4) ADRD and diabetes complications and progress, (5) management strategies for ADRD and other comorbidities, (6) caregiver social support, and (7) caregiver health and self-care. Of these, the first 5 are predominantly focused on the individual living with ADRD and the last 2 focus on the caregiver. Table 1 shows the themes, codes, and code frequency.

Table 1. Themes, codes, and code frequency.

Themes and codes	Code frequency, n
Disease linkages and comorbidities	
ADRD[a] and diabetes comorbidity	267
ADRD and other comorbidities	123
ADRD and diabetes link	107
ADRD or diabetes symptoms without diagnosis	
Unconfirmed diagnosis of ADRD	46
Diabetes differential diagnosis	35
Diabetes management	
Diabetes medication management	124
General diabetes management	109
Diet and nutrition in diabetes management	107
Physical activity and exercise	17
ADRD and diabetes complications and progress	
Changes in ability and capacity	135
Resistance	73
Diabetes complications	55
Forgetfulness	49
Management strategies for ADRD and other comorbidities	
Obstacles to patient care coordination	200
Pharmacotherapy	196
Fiblets	31
Diet (general)	28
Alternative medicine	5
Caregiver social support	
Informational support	233
Caregiver burden	185
Emotional support	156
Caregiver health and self-care	
Caregiver self-care struggles	57
Caregivers with diabetes	40
Caregiver health outcomes	38
Caregiver positive self-care	37

[a]ADRD: Alzheimer disease and related dementias.

Disease Linkages and Comorbidities

This theme encompasses discussions on 2 or more diseases suffered by their care recipient. Within posts in this sample, it was common for caregiver to provide context for their care recipient's situation by describing the health conditions of the care recipient. Posts falling under this theme either discussed

XSL•FO
RenderX

ADRD and diabetes, ADRD and another comorbidity, or the proposed link between ADRD and diabetes. Forum participants would often describe the kind of ADRD experienced by the care recipient. Given that this sample was focused on diabetes, many individuals with ADRD also had diabetes. If other comorbidities were discussed (eg, heart disease and stroke), it was extremely likely that individuals with ADRD also had diabetes in addition to other comorbidities:

> *My dad has congestive heart failure, has had 2 heart attacks, multiple TIAs, a stroke, stage 4 kidney disease, COPD, high blood pressure, diabetes, and after we moved them in with us four years ago was diagnosed with mild vascular dementia.*

To a lesser degree but still important to the discussion of diseases related to ADRD and diabetes was the link between the 2 diseases. Frequently, this was discussed in terms of disease pathology and the common link of insulin resistance to both diseases. Users would speculate as to how strong the connection was between the 2 diseases and if one might lead to the other. Some users would also share links to articles to support their claims:

> *I myself find most compelling the connection between the top three listed and Alzheimer's type dementia: These include: insulin resistance, diabetes, and obesity; These three are very related anyway.*
>
> *Here is a high level overview of the diabetes/Alzheimer's connection... Bottom line watch that sugar closely! This is a decades long thing. https://tinyurl.com/y842evt8.*

ADRD or Diabetes Symptoms Without Diagnosis

Forum use was not limited to individuals with dementia and their caregivers. This theme encompasses discussions on a care recipient who had some symptoms of ADRD or diabetes but had not been formally diagnosed. Individuals questioning symptoms that a loved one was experiencing utilized the forum to ask users if the symptoms or behaviors could be a sign of ADRD. In some cases, a caregiver of an individual with a diagnosis of ADRD also questioned whether pre-existing conditions (eg, diabetes) might be the root cause of cognitive changes being displayed by a care recipient, rather than a diagnosis of ADRD. It was common for these forum participants to seek information about the process of diagnosis and how to work with their loved one to receive medical care:

> *Can anyone suggest a way to find a specialist who sincerely look at my mom's situation and make sure she is being correctly diagnosed with dementia. No one has really done the proper testing to rule out possible mimics and I have not gotten an exact diagnoses of the type. I have been to 2 neurologist and I feel they just do the verbal test and say she has dementia. What about b12 deficiencies or the fact that she is on blood pressure medicine.*

Respondents to these forum participants, at times, provided a differential diagnosis based on the symptoms a user describes. Given the similar pathology of ADRD and diabetes (as well as other diseases), it is possible that the symptoms a user described as possible dementia were diabetes:

> *With someone your dad's age, it is important to complete the testing because there are over 50 other illnesses that look like Alz but are not and most importantly of all, many of these illnesses can be reversed. Certainly mismanagement of diabetes can cause cognitive changes, even in young people.*

Diabetes Management

Within this theme, caregivers discussed the challenges they face (eg, balancing diabetes management and quality of life) and shared some practical strategies they use when helping an individual living with ADRD to manage his or her diabetes. Overall, there were frequent discussions of diet and medications used to manage diabetes. Although some of these posts focused on individuals who had good management of diabetes, it was common for caregivers to question whether adherence to diabetes management practices was necessary for those with advanced ADRD. Caregivers expressed feelings that the overall quality of life was more important than strict adherence; however, they still maintained that diabetes management was important. Thus, caregivers commonly discussed how best to maintain a good quality of life for the care recipient, first and foremost, with a more relaxed approach to managing diabetes:

> *Mom, a type 2 diabetic, loves sweets...At some point in the last few months I decided that if having access to colas would give mom a 5 star day instead of a 3 star day and she should have the cola...I want to balance between having a good quality of life and a (somewhat) healthy life.*

It was also frequent for forum respondents to discuss the lack of diabetes management and high or variable blood sugar readings, which an individual with ADRD might experience:

> *His blood sugar was over 1000! The hospital got it under control and recommended that he have his blood sugars checked four times a day and given oral medicine and injected insulin to control it. The problem is that he doesn't want to get up in the morning or the afternoon and he is refusing to eat frequently which he must do to take the insulin.*

In contrast, very few caregivers discussed physical activity or exercise as a part of diabetes management for their care recipient. When they did, the discussion often centered around a concern for a lack of physical activity on the part of the care recipient:

> *I would like to ask for some advice on managing my father's repetitive behaviors. For example, he will sit idle for HOURS, sometimes ENTIRE days just staring off into space in the dark...My concern is the sitting idle for hours. He has diabetes, is in 3rd stage kidney failure, has poor circulation, and muscles that are getting VERY weak from being sedentary.*

Given the complexity of managing diabetes in an ADRD care recipient, caregivers frequently discussed management strategies for their care recipients as well. For example, some caregivers

XSL•FO
RenderX

helped care recipients to manage their diabetes through a diet with inventive approaches such as halving juice with water, providing Meals on Wheels subscriptions, or slowly exchanging poor food choices with more healthy alternatives. Caregivers often noted that these changes had to be made slowly, as abrupt changes often resulted in resistance from their care recipients:

> *My mom would report cereal with OJ for breakfast...When I got over there, it turned out that was true, so I started a Meals on Wheels subscription for her and dad. One meal was diabetes health labelled with her name and one was heart healthy labeled with his name...With orange juice, what we did was save one of the empty containers of orange juice and then pour half into the empty container and dilute the OJ by half. Mom never noticed the difference.*

However, some caregivers felt that a more direct approach better served their care recipients and, instead, opted to obtain the medical power of attorney or hire a home health professional to manage the care recipient's diabetes themselves. Whether their strategies were direct or indirect, caregivers frequently discussed practical ways to encourage or enforce adherence to diabetes management regimens through medication or diet:

> *Auditory memory is poor, she's repetitious and we are concerned about her continuing to live alone because she is an insulin dependent diabetic... I fill her pillbox and monitor it as much as possible, but I know that there are days when she doesn't take some of her pills...I have POA for healthcare and other POA is in the works.*

ADRD and Diabetes Complications and Progress

This theme encompasses discussions on how an ADRD care recipient's resistance or forgetfulness complicates care management. Many caregivers came to the forum to discuss disease progress and disease complications in their care recipients. Often forum users discussed the changes in abilities or capacities that an individual living with ADRD experienced and how this complicated overall care. Many of these posts expressed a worsening of disease state, which ultimately led to a change in the caregiver's caregiving experience:

> *My mother has been having real issues with her memory for well over a year now. It has gotten progressively worse...We have been lengthening her visits over the past year to help her adjust to the idea of living with us.*

When discussing complications and progress, some caregivers spoke about how resistance to diagnosis and forgetfulness affected the ability of the caregiver to provide appropriate care to care recipients for diabetes and ADRD, as it complicated how they provided care or how a care recipient responded to care:

> *I think my biggest concern about my mom continuing alone is the impact that her short term and auditory memory issues have on her medication management, more than anything else. My mom doesn't chart her blood sugars, never has. That's just who she is. If I ask her if she ate or took insulin, she doesn't remember.*
>
> *He has declined rapidly by his own choices not to follow the doctors orders from the very beginning. He won't monitor his blood sugars properly and rarely administers his insulin because he doesn't like the pricking of his finger or the pain of the shot. He eats whatever he wants and it consists entirely of carbs, high fats and sugar.*

Although less frequent, some of the users described complications of diabetes as one of the many concerns that their care recipient was experiencing:

> *The consequences of untreated diabetes are not fun; things like blindness and amputations of toes, then feet, then below the knee...I am sure your LO [loved one] does not want those, and yet with dementia they cannot make rational choices based on future risk.*

Management Strategies for ADRD and Other Comorbidities

Although this sample focused on diabetes, caregivers came to this forum, most importantly, to discuss ADRD. Within this theme, caregivers discussed how they helped an individual living with ADRD and with the management of comorbidities other than diabetes. As a result of the complexities caregivers experienced in caring for an individual with ADRD, caregivers frequently had questions or provided advice about how to navigate the housing, medical, legal, and day-to-day needs of their loved one:

> *After much discussion with my grandmother we have decided that my grandfather should be put in a nursing home. I unfortunately don't know a lick of anything related to senior care. I have been doing nonstop research and am so confused...I know some nursing homes take medicare but I am extremely weary.*
>
> *It is time for you to step in and take control of the situation. You must protect her. I am hoping that you are already your mother's financial and medical POA. If not, you need to contact your family attorney. If your family doesn't have an attorney, then look for an eldercare attorney in your area.*

A common management technique was using medication to help manage ADRD or other comorbidities that the individual with ADRD might be experiencing:

> *My mother was on Namenda 10mg once a day for two and a half years. I made the decision to stop the meds last June after many discussions (over time) with her PCP...My mom is currently taking Risperidone to manage anger and agitation and has been for two years. She also takes oral meds to manage Type II Diabetes and high blood pressure.*

However, caregivers also had unconventional techniques to help them manage the complexities of multiple conditions for an individual living with ADRD. For example, some caregivers

XSL•FO
RenderX

would use a *fiblet* or a therapeutic lie to help a patient to adhere to a medical regimen, a diet change, or new living arrangements:

> *In the beginning we did tell mom and she was always shocked: "I have Alzheimers?!?!" Now I just say her blood sugars levels are too high (she's diabetic) and once we can get control on them, she'll feel better. It's a lie, I know, but I don't see the point in continually telling her.*

Occasionally, caregivers discussed the use of special diets, such as the Mediterranean diet, to prevent the worsening of ADRD. This was distinct from the use of diet to manage diabetes and blood glucose levels:

> *...The "grain brain" as a cause of dementia may be somewhat off because many grains also contain polyphenols that help protect against Alzheimer's disease such as ferulic acid in rice bran or flaxseed in some breads. That is why a Mediterranean diet relatively high in carbohydrates is partially protective against Alzheimer's disease.*

Caregiver Social Support

A prominent theme within the forum was social support. Individuals frequently used the forum to seek support related to providing care or to provide support to other individuals in need of information or encouragement. Many users discussed the burden of taking care of a patient with ADRD and would use the forum to discuss frustrations or stress associated with caregiving as well as to seek support:

> *I struggle with the what if's even though everything says AD. Being her daughter, I want to throw her in the vehicle, drive her to the doctor and have them check her from head to toe everyday. But at what expense? Her dignity? Her comfort? I do everything I know to do and well...I just wish I guess...I feel like I'm rambling out of exhaustion.*

Some users also provided emotional support to help encourage a user who was struggling:

> *I'm sorry to hear about what you are going through with your family. For myself, as someone on the outside looking in, the solutions seem easy, but it isn't.*

In addition, users posed questions and gave advice that was strictly informational in nature. Some users provided guidance and suggestions derived from their personal experiences, and others provided links, articles, and phone numbers for resources that they believed would be valuable to the individual with the question:

> *To answer your main question - how do we convince our LO to do certain things that will help them? Well, that usually doesn't work because even if they are agreeable to the idea, their brains don't retain new information or learn new habits very well if at all. As far as my mom, for a long time we went back and forth with the idea of her living with me...mom had POA in order but had "capacity" (a psychiatrist evaluated her and found her able to make her own decisions) so I couldn't force anything.*

> *I would like to invite you to contact the Helpline at the Alzheimer's Assn. which can be reached at: (800) 272-3900…Consultants are highly educated Social Workers who specialize in dementia and they can be wonderful support, provide education materials, have contact numbers for helpful community entities and who can assist with problem solving and more.*

Caregiver Health and Self-Care

In contrast to social support, this theme was focused on discussions on caregivers' own physical and mental health and self-care. However, some caregivers did discuss having their own diagnosis of diabetes or another health concern:

> *Anxiety has now joined it, and Dad is now in the nh [nursing home]. But caring for Dad here at home for 12 years triggered it - along with hypertension and prediabetes. These are all stress related health issues.*

In discussions of their own health outcomes, some caregivers discussed their inability to participate in self-care activities. This struggle to maintain a healthy lifestyle, whether physically or mentally, was often attributed to a lack of free time because of the sheer amount of time spent caregiving:

> *There are times, I am so exhausted that I just want to "drop dead". I am too tired of being tired...too tired to rest. But I keep moving forward.*

> *I have no life at all, haven't seen friends in years, can't establish any new friends, can't set up any type of routine for my own enjoyment or benefit (school, hobbies, gym, martial arts) or work on my own life that I walked away from to take care of her and I am always exhausted these days.*

In contrast, others were concerned about developing chronic diseases, such as dementia and diabetes, and participated in and promoted positive self-care behaviors to reduce their risk of developing diseases:

> *I say this as a caregiver who also has trouble taking care of myself sometimes, but life is forcing me to learn. In fact, right now, I need to get off the computer and get outside for some exercise, which is my daily medicine against encroaching diabetes, heart disease, and mental illness from all the stress.*

> *My next health challenge is to NOT get dx with diabetes, so over the past few months I've stopped stress-eating chocolate foods, and have finally started exercising daily after the Mom-details, which took many months to sort out, are sorted.*

Discussion

Principal Findings

This is, to the best of our knowledge, the first study to examine diabetes-related topics in individuals with ADRD from the perspectives of family caregivers participating in an online support community. Our findings advance knowledge regarding the role of social media in health management. The findings are consistent with previous findings that health forums provide support to forum users (eg, informational and emotional support)

XSL•FO
RenderX

[53,54] and offer platforms for users in communication of diseases in general, disease symptoms and treatments, and user opinions [55]. However, unlike other health forum users, most of whom are patients themselves [27,56], the ALZConnected forum was designed for caregivers. The study findings demonstrated how family caregivers used this online forum to obtain resources for informational and emotional support and to seek ways to relieve caregiving burden stress. As this study focused on diabetes-related topics, discussions of the analyzed posts were always related to (1) health conditions of care recipients, symptoms, and management of diabetes and ADRD; (2) the daily burden and struggles of diabetes management in individuals with both ADRD and diabetes, which might not be captured by traditional data collection; and (3) how constant caregiver burden and daily struggles affect their self-care and result in poorer physical and mental health in family caregivers.

Forum participants extensively discussed information related to the care recipients' health conditions with comorbidities, links between ADRD and comorbidities, and uncertainty of the root causes of some symptoms and behaviors. This may reflect the complex interactions of ADRD and its comorbidities [57] and the underlying causes of family caregivers' struggles. Older adults without dementia have 2 comorbidities on average, whereas those living with dementia have an average of 4 comorbidities, and almost 9 out of 10 individuals living with dementia have at least one comorbidity [58]. A scoping review found that little is known about the care of comorbidities for people living with dementia and comorbidities, especially from the perspectives of those with dementia and their family caregivers [59]. Our study further echoes and reinforces the importance of understanding the management of ADRD comorbidities, especially diabetes, from the experience of family caregivers. This line of discussion in the forum could serve as an outlet for forum users to relieve their frustrations and anxiety about the uncertainty of the disease progression in their care recipients living with multiple chronic comorbidities and how they can better care for their loved ones. Moreover, the information obtained from the forums by caregivers might also prove useful in supporting communication with their health providers, both for themselves and their care recipients.

Given the focus on diabetes in this study, we found that family caregivers commonly discussed challenges associated with diabetes management in their loved ones living with ADRD, regardless of type 1 or type 2 diabetes. For example, many challenges faced by caregivers centered on diet, exercise, and medication management in their care recipients. This is unsurprising, as diabetes is largely a self-managed disease that necessitates high-level cognitive capabilities for patients to properly adhere to medication and lifestyle regimens. However, memory loss, impaired problem-solving, and other ADRD symptoms make diabetes management extremely difficult for people living with ADRD and their family caregivers [14,60]. Previous recommendations and studies have suggested that for people with cognitive impairment, diabetes care should be individualized and patient-centered [60]. The glycemic target (eg, hemoglobin A_{1c}, fasting glucose, and postprandial glucose) should also be relaxed [14]. However, the incorporation of these recommendations into daily care regimens has rarely been studied in individuals living with ADRD and diabetes. In fact, studies examining diabetes self-management often exclude individuals with cognitive impairment. In addition, despite the well-recognized importance of including family caregivers in the development of self-management plans for ADRD, few studies have evaluated structured interventions to provide education and support for family caregivers of ADRD [60], and even less studies have focused on diabetes management. Studies involving people living with ADRD and their family caregivers are urgently needed to address diabetes *self-management* in home settings for this population living with both ADRD and diabetes.

For challenges of diabetes management, the findings also highlight the difficulties caregivers face when trying to balance quality of life and diabetes control in their ADRD care recipients. Much of this is associated with the responsibility of providing proper health care to their loved ones that caregivers feel while still making sure their loved one's overall quality of life is high. Regardless of a caregiver's best efforts, uncontrolled blood glucose levels and subsequent hospital admission were often discussed by family caregivers. Balancing glycemic control and quality of life in ADRD care recipients warrants further exploration. Despite how challenges were frequently discussed, in the forum, some personal practical strategies for diabetes management were also shared. For example, some caregivers helped care recipients manage their diabetes through gradual diet changes or with the use of inventive approaches, such as halving juice with water. Whether the shared strategies were direct or indirect, caregivers frequently discussed practical ways by which they encouraged and enforced adherence to diabetes management regimens, whether through medication or diet. The various strategies shared may serve as examples for forum users looking to tailor the approaches to their own loved ones and situations. In addition, these suggestions might provide helpful information and examples for developing diabetes management guidelines for people living with ADRD and their family caregivers. This further reinforces the importance of involving caregivers and their loved ones in studies of developing diabetes management guides.

Our study revealed that forum users, who take care of individuals with ADRD and diabetes, frequently reported increased distress and neglect of self-care, both of which may contribute to a decrease in health status and quality of life for both the care recipient and family caregivers [61]. Previous studies reported that the time spent on daily activities and supervision by caregivers was higher when caring for individuals with ADRD and diabetes compared with caring for those with ADRD without diabetes [62]. Consequently, the level of the caregiver's perceived burden increases [63-66]. ALZConnected has provided a platform for family caregivers to seek information and emotional support in their attempts to learn more about how to manage ADRD of their care recipients and seek support for how to manage their own health and well-being as a caregiver. The study findings of challenges and struggles that caregivers constantly face have implications not just on their care recipients but also on their own health and well-being. We call for more practical diabetes management guidelines for individuals living with ADRD and their caregivers to ease the

XSL•FO
RenderX

burden of caregivers and improve the health and quality of life for both.

Limitations

This study has several limitations. First, the study used publicly available forum data, and as a result, the data collection process was out of the control of the study team. However, this could also be considered a strength as there was no interference with participants' opinions, meaning the posts reviewed truly reflect the caregivers' struggles and concerns. Second, the posts extracted for this study only represent a portion of posts made in the forum and may not be applicable to all discussions regarding diabetes. However, our analysis did reach saturation, which might minimize this limitation. Third, we attempted to explore discussion topics related to diabetes in this forum and used thematic analysis to analyze the data [67]. Thus, there is no theoretical framework to support this study. Fourth, this study only used data from one social media platform, and the discussions in the forum are largely user-driven. In the future, using multiple social media data sources may further extend our knowledge of the study topic. Finally, a majority of the posts included in this study were from adult children of ADRD care recipients, which might reflect their proportion of family care or that younger persons are more likely to use social media/online forums. As a result, the generalizability of the results to all family caregivers might be limited.

Conclusions

In summary, the overwhelming burdens of diabetes management in individuals living with ADRD and diabetes were apparent for both caregivers and care recipients based on discussions from the ALZConnected forum. Research involving both care recipients and their caregivers in developing diabetes management guidelines and interventions for family caregivers of individuals with diabetes and ADRD would be important. In addition, the ALZConnected forum provided a platform for caregivers to seek informational and emotional support for caring for persons living with ADRD and diabetes. Studies are urgently needed to provide practical guidelines or tools available in online support communities as new ways to support daily diabetes management for individuals living with ADRD coexisting with diabetes and their caregivers. Future studies to explore delivering diabetes management interventions through online communities in caregivers and their care recipients with ADRD and diabetes are warranted.

Acknowledgments

The authors would like to thank the study participants who shared their data on ALZConnected. The study was supported by the Center on Smart and Connected Health Technologies at the University of Texas Health Science Center at San Antonio. YD is supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant TL1 TR002647.

Authors' Contributions

YD, JW, KP, CW, SM, and SS designed the study. CL extracted and cleaned the data. YD, KP, AC, SK, and JW analyzed the data. YD, KP, AC, SK, JW, CW, SM, and SS reviewed the results. YD, KP, AC, and KL wrote the manuscript. All authors reviewed, edited, and approved the manuscript.

Conflicts of Interest

None declared.

References

1. 2019 Alzheimer's Disease Facts and Figures. Alzheimer's Association: Alzheimer's Disease & Dementia Help. 2019 Mar 1. URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf [accessed 2020-05-14]
2. National Diabetes Statistics Report 2020. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2020-06-22]
3. Alzheimer's Disease. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/aging/aginginfo/alzheimers.htm [accessed 2020-01-10]
4. Salber PR, Selecky CE, Soenksen D, Wilson T. Impact of dementia on costs of modifiable comorbid conditions. Am J Manag Care 2018 Nov 1;24(11):e344-e351 [FREE Full text] [Medline: 30452202]
5. Bunn F, Burn AM, Goodman C, Robinson L, Rait G, Norton S, et al. Comorbidity and dementia: a mixed-method study on improving health care for people with dementia (CoDem). Health Serv Delivery Res 2016. [doi: 10.3310/hsdr04080] [Medline: 26937541]
6. Kandimalla R, Thirumala V, Reddy PH. Is Alzheimer's disease a type 3 diabetes? A critical appraisal. Biochim Biophys Acta Mol Basis Dis 2017 May;1863(5):1078-1089 [FREE Full text] [doi: 10.1016/j.bbadis.2016.08.018] [Medline: 27567931]
7. Mittal K, Mani RJ, Katare DP. Type 3 diabetes: cross talk between differentially regulated proteins of type 2 diabetes mellitus and Alzheimer's disease. Sci Rep 2016 May 6;6:25589 [FREE Full text] [doi: 10.1038/srep25589] [Medline: 27151376]
8. Ravona-Springer R, Heymann A, Schmeidler J, Moshier E, Godbold J, Sano M, et al. Trajectories in glycemic control over time are associated with cognitive performance in elderly subjects with type 2 diabetes. PLoS One 2014;9(6):e97384 [FREE Full text] [doi: 10.1371/journal.pone.0097384] [Medline: 24887092]

XSL•FO
RenderX

9. Cukierman-Yaffe T, Gerstein HC, Williamson JD, Lazar RM, Lovato L, Miller ME, Action to Control Cardiovascular Risk in Diabetes-Memory in Diabetes (ACCORD-MIND) Investigators. Relationship between baseline glycemic control and cognitive function in individuals with type 2 diabetes and other cardiovascular risk factors: the action to control cardiovascular risk in diabetes-memory in diabetes (ACCORD-MIND) trial. Diabetes Care 2009 Feb;32(2):221-226 [FREE Full text] [doi: 10.2337/dc08-1153] [Medline: 19171735]
10. Spauwen PJ, Köhler S, Verhey FR, Stehouwer CD, van Boxtel MP. Effects of type 2 diabetes on 12-year cognitive change: results from the Maastricht aging study. Diabetes Care 2013 Jun;36(6):1554-1561 [FREE Full text] [doi: 10.2337/dc12-0746] [Medline: 23275366]
11. Roberts RO, Knopman DS, Przybelski SA, Mielke MM, Kantarci K, Preboske GM, et al. Association of type 2 diabetes with brain atrophy and cognitive impairment. Neurology 2014 Apr 1;82(13):1132-1141 [FREE Full text] [doi: 10.1212/WNL.0000000000000269] [Medline: 24647028]
12. Shinohara M, Sato N. Bidirectional interactions between diabetes and Alzheimer's disease. Neurochem Int 2017 Sep;108:296-302. [doi: 10.1016/j.neuint.2017.04.020] [Medline: 28551028]
13. American Diabetes Association. Standards of medical care for patients with diabetes mellitus. Diabetes Care 2003 Jan;26(Suppl 1):S33-S50. [doi: 10.2337/diacare.26.2007.s33] [Medline: 12502618]
14. Puttanna A, Padinjakara NK. Management of diabetes and dementia. Br J Diabetes 2017 Sep 10;17(3):93. [doi: 10.15277/bjd.2017.139]
15. Prince M, Albanese E, Guerchet M, Prina M. Dementia and Risk Reduction: An Analysis of Protective and Modifiable Factors. Alzheimer's Disease International. 2014. URL: https://www.alz.co.uk/research/WorldAlzheimerReport2014.pdf [accessed 2020-06-24]
16. Schulz R, Martire LM. Family caregiving of persons with dementia: prevalence, health effects, and support strategies. Am J Geriatr Psychiatry 2004;12(3):240-249. [Medline: 15126224]
17. Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. Am J Nurs 2008 Sep;108(9 Suppl):23-7; quiz 27 [FREE Full text] [doi: 10.1097/01.NAJ.0000336406.45248.4c] [Medline: 18797217]
18. von Känel R, Mills PJ, Mausbach BT, Dimsdale JE, Patterson TL, Ziegler MG, et al. Effect of Alzheimer caregiving on circulating levels of C-reactive protein and other biomarkers relevant to cardiovascular disease risk: a longitudinal study. Gerontology 2012;58(4):354-365 [FREE Full text] [doi: 10.1159/000334219] [Medline: 22133914]
19. Surwit RS, Schneider MS. Role of stress in the etiology and treatment of diabetes mellitus. Psychosom Med 1993;55(4):380-393. [doi: 10.1097/00006842-199307000-00005] [Medline: 8105502]
20. Bunn F, Goodman C, Jones PR, Russell B, Trivedi D, Sinclair A, et al. Managing diabetes in people with dementia: a realist review. Health Technol Assess 2017 Dec;21(75):1-140 [FREE Full text] [doi: 10.3310/hta21750] [Medline: 29235986]
21. Gabarron E, Bradway M, Fernandez-Luque L, Chomutare T, Hansen AH, Wynn R, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 5;18(1):414 [FREE Full text] [doi: 10.1186/s12913-018-3178-7] [Medline: 29871675]
22. Anderson JG, Hundt E, Dean M, Keim-Malpass J, Lopez RP. 'The church of online support': examining the use of blogs among family caregivers of persons with dementia. J Fam Nurs 2017 Feb;23(1):34-54. [doi: 10.1177/1074840716681289] [Medline: 27920340]
23. ALZConnected. 2014. URL: https://www.alzconnected.org/ [accessed 2020-01-10]
24. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
25. Welch V, Petkovic J, Pardo Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
26. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
27. Sinha A, Porter T, Wilson A. The use of online health forums by patients with chronic cough: qualitative study. J Med Internet Res 2018 Jan 24;20(1):e19 [FREE Full text] [doi: 10.2196/jmir.7975] [Medline: 29367181]
28. Martí P, Serrano-Estrada L, Nolasco-Cirugeda A. Social media data: challenges, opportunities and limitations in urban studies. Comput Environ Urban 2019 Mar;74:161-174. [doi: 10.1016/j.compenvurbsys.2018.11.001]
29. Liu D, Yamashita T, Burston B. Identifying consumers who search for long-term care on the web: latent class analysis. JMIR Aging 2018 Nov 2;1(2):e10763 [FREE Full text] [doi: 10.2196/10763] [Medline: 31518237]
30. Mikal JP, Grande SW, Beckstrand MJ. Codifying online social support for breast cancer patients: retrospective qualitative assessment. J Med Internet Res 2019 Oct 24;21(10):e12880 [FREE Full text] [doi: 10.2196/12880] [Medline: 31651404]
31. Shang L, Zuo M, Ma D, Yu Q. The antecedents and consequences of health care professional-patient online interactions: systematic review. J Med Internet Res 2019 Sep 25;21(9):e13940 [FREE Full text] [doi: 10.2196/13940] [Medline: 31573908]
32. Signorelli GR, Lehocki F, Fernández MM, O'Neill G, O'Connor D, Brennan L, et al. A research roadmap: connected health as an enabler of cancer patient support. J Med Internet Res 2019 Oct 29;21(10):e14360 [FREE Full text] [doi: 10.2196/14360] [Medline: 31663861]

XSL•FO
RenderX

33. Karmen C, Hsiung RC, Wetter T. Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. Comput Methods Programs Biomed 2015 Jun;120(1):27-36. [doi: 10.1016/j.cmpb.2015.03.008] [Medline: 25891366]
34. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. Brief Bioinform 2018 Sep 28;19(5):863-877 [FREE Full text] [doi: 10.1093/bib/bbx010] [Medline: 28334070]
35. Olsen A, Keogh J, Sargeant S. Investigating how bowel cancer survivors discuss exercise and physical activity within web-based discussion forums: qualitative analysis. J Med Internet Res 2019 Dec 16;21(12):e13929 [FREE Full text] [doi: 10.2196/13929] [Medline: 31841117]
36. Lognos B, Carbonnel F, Launay IB, Bringay S, Guerdoux-Ninot E, Mollevi C, et al. Complementary and alternative medicine in patients with breast cancer: exploratory study of social network forum data. JMIR Cancer 2019 Nov 27;5(2):e12536 [FREE Full text] [doi: 10.2196/12536] [Medline: 31774404]
37. de Silva D, Ranasinghe W, Bandaragoda T, Adikari A, Mills N, Iddamalgoda L, et al. Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. PLoS One 2018;13(10):e0205855 [FREE Full text] [doi: 10.1371/journal.pone.0205855] [Medline: 30335805]
38. Johansson P, Westas M, Andersson G, Alehagen U, Broström A, Jaarsma T, et al. An internet-based cognitive behavioral therapy program adapted to patients with cardiovascular disease and depression: randomized controlled trial. JMIR Ment Health 2019 Oct 3;6(10):e14648 [FREE Full text] [doi: 10.2196/14648] [Medline: 31584000]
39. Eliya Y, Pellegrini D, Gevaert A, Code J, van Spall HG. Social media in heart failure: a mixed methods systematic review. Curr Cardiol Rev 2019 Dec 10:- epub ahead of print. [doi: 10.2174/1573403X15666191210143657] [Medline: 31820703]
40. Fisher S, Jehassi A, Ziv M. Hidradenitis suppurativa on Facebook: thematic and content analyses of patient support group. Arch Dermatol Res 2019 Dec 16:- epub ahead of print. [doi: 10.1007/s00403-019-02027-3] [Medline: 31845011]
41. Discussion Board: Caregivers Forum. ALZ Connected. 2019. URL: https://www.alzconnected.org/discussion.aspx?g=topics&f=151 [accessed 2019-12-01]
42. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
43. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak 2018 Jun 14;18(1):38 [FREE Full text] [doi: 10.1186/s12911-018-0621-y] [Medline: 29898743]
44. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
45. Castleberry A, Nolen A. Thematic analysis of qualitative research data: is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Cheng TY, Liu L, Woo BK. Analyzing twitter as a platform for Alzheimer-related dementia awareness: thematic analyses of tweets. JMIR Aging 2018 Dec 10;1(2):e11542 [FREE Full text] [doi: 10.2196/11542] [Medline: 31518232]
48. Clark IN, Tamplin JD, Baker FA. Community-dwelling people living with dementia and their family caregivers experience enhanced relationships and feelings of well-being following therapeutic group singing: a qualitative thematic analysis. Front Psychol 2018;9:1332 [FREE Full text] [doi: 10.3389/fpsyg.2018.01332] [Medline: 30104994]
49. Morris MT, Daluiski A, Dy CJ. A thematic analysis of online discussion boards for brachial plexus injury. J Hand Surg Am 2016 Aug;41(8):813-818. [doi: 10.1016/j.jhsa.2016.05.007] [Medline: 27311862]
50. Sullivan CF. Gendered cybersupport: a thematic analysis of two online cancer support groups. J Health Psychol 2003 Jan;8(1):83-104. [doi: 10.1177/1359105303008001446] [Medline: 22113903]
51. Greenwood N, Smith R, Akhtar F, Richardson A. A qualitative study of carers' experiences of dementia cafés: a place to feel supported and be yourself. BMC Geriatr 2017 Jul 25;17(1):164 [FREE Full text] [doi: 10.1186/s12877-017-0559-4] [Medline: 28743253]
52. Alhojailan M. Thematic analysis: a critical review of its process and evaluation. West East J Soc Sci 2012;1(1):-.
53. Ito T, Meguro K, Akanuma K, Ishii H, Mori E. A randomized controlled trial of the group reminiscence approach in patients with vascular dementia. Dement Geriatr Cogn Disord 2007;24(1):48-54. [doi: 10.1159/000103631] [Medline: 17565213]
54. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: a content analysis. Patient Educ Couns 2012 Jun;87(3):405-410. [doi: 10.1016/j.pec.2011.09.011] [Medline: 22019021]
55. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]
56. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]

XSL•FO
RenderX

57. Subramaniam H. Co-morbidities in dementia: time to focus more on assessing and managing co-morbidities. Age Ageing 2019 May 1;48(3):314-315. [doi: 10.1093/ageing/afz007] [Medline: 31063580]
58. Poblador-Plou B, Calderón-Larrañaga A, Marta-Moreno J, Hancco-Saavedra J, Sicras-Mainar A, Soljak M, et al. Comorbidity of dementia: a cross-sectional study of primary care older patients. BMC Psychiatry 2014 Mar 20;14:84 [FREE Full text] [doi: 10.1186/1471-244X-14-84] [Medline: 24645776]
59. Bunn F, Burn A, Goodman C, Rait G, Norton S, Robinson L, et al. Comorbidity and dementia: a scoping review of the literature. BMC Med 2014 Oct 31;12:192 [FREE Full text] [doi: 10.1186/s12916-014-0192-4] [Medline: 25358236]
60. Bunn F, Goodman C, Reece Jones P, Russell B, Trivedi D, Sinclair A, et al. What works for whom in the management of diabetes in people living with dementia: a realist review. BMC Med 2017 Jul 28;15(1):141 [FREE Full text] [doi: 10.1186/s12916-017-0909-2] [Medline: 28750628]
61. Varela G, Varona L, Anderson K, Sansoni J. Alzheimer's care at home: a focus on caregivers strain. Prof Inferm 2011;64(2):113-117 [FREE Full text] [Medline: 21843435]
62. Lebrec J, Ascher-Svanum H, Chen Y, Reed C, Kahle-Wrobleski K, Hake AM, et al. Effect of diabetes on caregiver burden in an observational study of individuals with Alzheimer's disease. BMC Geriatr 2016 May 3;16:93 [FREE Full text] [doi: 10.1186/s12877-016-0264-8] [Medline: 27142529]
63. Cheng S. Dementia caregiver burden: a research update and critical analysis. Curr Psychiatry Rep 2017 Aug 10;19(9):64 [FREE Full text] [doi: 10.1007/s11920-017-0818-2] [Medline: 28795386]
64. Chiao C, Wu H, Hsiao C. Caregiver burden for informal caregivers of patients with dementia: a systematic review. Int Nurs Rev 2015 Sep;62(3):340-350. [doi: 10.1111/inr.12194] [Medline: 26058542]
65. Haro JM, Kahle-Wrobleski K, Bruno G, Belger M, Dell'Agnello G, Dodel R, et al. Analysis of burden in caregivers of people with Alzheimer's disease using self-report and supervision hours. J Nutr Health Aging 2014 Jul;18(7):677-684. [doi: 10.1007/s12603-014-0500-x] [Medline: 25226106]
66. Mioshi E, Foxe D, Leslie F, Savage S, Hsieh S, Miller L, et al. The impact of dementia severity on caregiver burden in frontotemporal dementia and Alzheimer disease. Alzheimer Dis Assoc Disord 2013;27(1):68-73. [doi: 10.1097/WAD.0b013e318247a0bc] [Medline: 22314247]
67. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]

Abbreviations

ADRD: Alzheimer disease and related dementias

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by F Bunn, D Liu, C Peterson, G Signorelli; comments to author 10.02.20; revised version received 07.04.20; accepted 03.06.20; published 06.07.20.

Please cite as:
Du Y, Paiva K, Cebula A, Kim S, Lopez K, Li C, White C, Myneni S, Seshadri S, Wang J
Diabetes-Related Topics in an Online Forum for Caregivers of Individuals Living With Alzheimer Disease and Related Dementias: Qualitative Inquiry
J Med Internet Res 2020;22(7):e17851
URL: http://www.jmir.org/2020/7/e17851/
doi:10.2196/17851
PMID:32628119

©Yan Du, Kristi Paiva, Adrian Cebula, Seon Kim, Katrina Lopez, Chengdong Li, Carole White, Sahiti Myneni, Sudha Seshadri, Jing Wang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessment of Adaptive Engagement and Support Model for People With Chronic Health Conditions in Online Health Communities: Combined Content Analysis

Brian M Green[1*], MS; Katelyn Tente Van Horn[1*], MSW, MPH; Ketki Gupte[1*], MA; Megan Evans[2*], MS; Sara Hayes[1*], MPH; Amrita Bhowmick[1,2*], MPH, MBA

[1]Health Union, LLC, Philadelphia, PA, United States

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

[*]all authors contributed equally

Corresponding Author:
Brian M Green, MS
Health Union, LLC
1217 Sansom Street
2nd Floor
Philadelphia, PA, 19107
United States
Phone: 1 2673072525
Fax: 1 4844502513
Email: brian.m.green@health-union.com

Abstract

Background: With the pervasiveness of social media, online health communities (OHCs) are an important tool for facilitating information sharing and support among people with chronic health conditions. Importantly, OHCs offer insight into conversations about the lived experiences of people with particular health conditions. Little is known about the aspects of OHCs that are important to maintain safe and productive conversations that support health.

Objective: This study aimed to assess the provision of social support and the role of active moderation in OHCs developed in accordance with and managed by an adaptive engagement model. This study also aimed to identify key elements of the model that are central to the development, maintenance, and adaptation of OHCs for people with chronic health conditions.

Methods: This study used combined content analysis, a mixed methods approach, to analyze sampled Facebook post comments from 6 OHCs to understand how key aspects of the adaptive engagement model facilitate different types of social support. OHCs included in this study are for people living with multiple sclerosis, migraine, irritable bowel syndrome, rheumatoid arthritis, lung cancer, and prostate cancer. An exploratory approach was used in the analysis, and initial codes were grouped into thematic categories and then confirmed through thematic network analysis using the Dedoose qualitative analysis software tool. Thematic categories were compared for similarities and differences for each of the 6 OHCs and by topic discussed.

Results: Data on the reach and engagement of the Facebook posts and the analysis of the sample of 5881 comments demonstrate that people with chronic health conditions want to engage on the web and find value in supporting and sharing their experiences with others. Most comments made in these Facebook posts were expressions of social support for others living with the same health condition (3405/5881, 57.89%). Among the comments with an element of support, those where community members validated the knowledge or experiences of others were most frequent (1587/3405, 46.61%), followed by the expression of empathy and understanding (1089/3405, 31.98%). Even among posts with more factual content, such as insurance coverage issues, user comments still had frequent expressions of support for others (80/213, 37.5%).

Conclusions: The analysis of this OHC adaptive engagement model in action shows that the foundational elements—social support, engagement, and moderation—can effectively be used to provide a rich and dynamic community experience for individuals with chronic health conditions. Social support is demonstrated in a variety of ways, including sharing information or validating information shared by others, expressions of empathy, and sharing encouraging statements with others.

(J Med Internet Res 2020;22(7):e17338) doi:10.2196/17338

XSL•FO
RenderX

KEYWORDS

social media; social support; health education; qualitative research; patient empowerment

Introduction

Overview of the Role of Social Support in Health

Social relationships and social support are purported to promote good physical and mental health for people living with a variety of chronic medical conditions [1-3]. Social support can be defined as "verbal and nonverbal communication between recipients and providers that reduces uncertainty about the situation, the self, the other, or the relationship, and functions to enhance a perception of personal control in one's experience" [4].

Social relationships and social support can be beneficial to health in many ways. Numerous studies have examined the connection between an individual's social relationships or networks and their physical health. Research demonstrates that, compared with individuals who are isolated, those who have a social network live longer, experience fewer physical symptoms of illness, and specifically have lower blood pressure [2,5,6]. Poor-quality or low-quality social ties have been associated with the development and progression of certain conditions, including high blood pressure, cancer, cardiovascular disease, autonomic dysregulation, and mortality [2,5,6].

Previous research has also shown an increased likelihood of survival in people who have stronger social ties or social relationships compared with those who have weaker ties or relationships [7]. A review of 81 studies revealed that social support was reliably related to beneficial effects on cardiovascular, endocrine, and immune health [8]. Additionally, social support is associated with increased health-promoting behaviors such as medication adherence [9,10], smoking cessation [11], and weight loss [12,13]. One study in particular showed that providing support was also related to lower measures of systolic and diastolic blood pressure among those who gave support [14].

Use of the Internet and Social Media for Social Support for Health

The use of the internet has changed how patients, caregivers, friends, family members, and interested citizen-scientists learn about and seek support for specific health conditions. In the United States, 61% of adults search on the web and 39% use social media to find health information [15]. Additionally, 1 in 4 internet users with a chronic illness has gone on the web to find others with similar health problems [16]. It is clear that patients rely on web-based resources, particularly when encountering specific situations, and being diagnosed with a chronic health condition is one such situation.

In a web-based *Health Information Experiences* survey of >2200 patients across 7 health conditions, the authors found that people used web-based resources at some points along their health care journey [17]. For example, 67% of the people used web-based resources when starting a new medication or treatment, 67% when experiencing a new side effect or symptom, 61% when making a medication treatment decision or change, and 57% when needing emotional support [17]. In the same survey, patients were asked about the web-based resources they used; 71% used condition-specific websites and 65% used Facebook for health information and support [17].

Diverse population groups across different races, ethnicities, and education levels access health-related information using social media [18,19]. Social media platforms, such as Facebook, can facilitate dialogue between patients as well as between patients and providers [18].

Using a definition similar to that found in other recent research, online communities are defined in this study as the use of an internet-based platform with the purpose of bringing together a group of individuals with a shared interest or goal [20]. As online communities are hosted on the internet, the geographic location of the member is not a vital component of *membership* in the community or how the community functions. Online support groups and communities have been in existence since the 1990s in various forms. The more recent development of online communities focused on health has been especially beneficial for those living with or caring for someone with a chronic health condition [20,21].

Online health communities (OHCs) are internet-based platforms that allow people to connect over shared experiences and communicate about health [20-22]. OHCs are disrupting traditional hierarchies in the health care industry by offering what traditional health care cannot offer—conversations about the lived experiences of people with particular health conditions. Individuals living with specific health conditions and their caregivers or care partners can browse and search on the web. They can also control the content and flow of information available to them, and, in the case of social media, produce content, rather than just receiving information directly from a health professional in a face-to-face encounter [23]. Although there is criticism surrounding the role of technology in increasing rates of loneliness and social isolation in adults, many view social media as a means to connect with others [24,25].

There are different types of OHCs, both open and closed, with the early versions offering *bulletin board*–style forums where individuals could ask questions and receive responses from other participants. In a closed OHC, a prospective member must opt in to join and be accepted by the site administrators to access site resources and engage with other members, whereas open communities are accessible to all people without having to sign in or provide contact information [20]. Over time, OHCs have evolved to allow for more robust offerings in the ways in which community members can engage through the platform as well as with other community members. OHCs are now more complex ecosystems, capable of being programmed to be better able to adapt and respond to changes in health care and patient experience.

XSL•FO
RenderX

The Value of Online Communities for Health Promotion

A vital aspect of using OHCs and social media as avenues for health communication is that they can provide valuable peer, social, and emotional support [18,26]. In studies on online communities, 4 types of social support are demonstrated: informational support [27], emotional support [28], instrumental support [29], and appraisal support [28].

An additional type of support that can be garnered by participation in an online community is network support, which is the feeling of being part of a group that shares common interests, which in turn can widen an individual's social network [29].

Web-based communication and interaction have been shown to provide meaningful support to people seeking health information. This is the case, whether that support is more transactional and provides information support that may help an individual obtain a desired service or make a decision or emotional support where the individual receives validation of their feelings or experiences [30]. Previous research has demonstrated that social support is important for an array of health outcomes and can be provided in various ways. There is evidence that social support can flourish when provided specifically through a web-based environment and can even empower individuals [27,31,32].

Social media offers additional space for dialogue around health information and experiences. Although many individuals find comfort in openly sharing aspects of their health experiences on the web, other people are more passive in how they participate in online communities. Research has identified 2 main types of participation in online communities: active participation, which includes posting and sharing, and passive participation, defined as browsing and reading [33]. Whether participants in an online community are more active or passive, both groups can find the level of support they are seeking [34].

Social support through a web-based medium can offer similar depth, breadth, and level of intimacy that occurs in private offline conversations [21,35]. Face-to-face social support may not always be available or may be difficult for people who have a chronic health condition or are disabled; therefore, OHCs can provide the vehicle through which ongoing and regular interaction is available [29].

However, the effects of web-based social support on chronic disease management are less clear from the available literature [26]. Research examining online cancer support groups found that participants in these groups had decreased levels of depression as well as decreased reactions to pain [36]. Previous research has also examined perceived empathy as a key factor in patient health outcomes [37]. Recent research has shown that users of an online diabetes management community have higher levels of diabetes self-care and health-related quality of life [38]. Despite the value of online communities for engaging people in a variety of health issues, research on the use of social media platforms, such as Facebook, to help people manage chronic health conditions is in its relative infancy [39].

Purpose of This Study

This study aimed to assess the provision of social support and the role of moderation in OHCs developed in accordance with and managed by the adaptive engagement model described herein and created by Health Union. Health Union cultivates OHCs that are designed to have a positive impact on the lives of people living with chronic conditions by providing the information and support they are seeking. Health Union currently has 27 active condition-specific OHCs and corresponding Facebook pages that serve to extend the reach of the adaptive engagement model.

In this study, the authors, who also serve as OHC community managers, first demonstrate how the adaptive engagement model works to maintain safe web-based spaces for the provision of social support for people with chronic health conditions. In the study model, moderation includes responding to comments; thanking and welcoming comments; providing validation; linking to resources; and maintaining a respectful, inclusive tone. In this paper, the authors elucidate key elements of the model that are central to the development, maintenance, and adaptation of OHCs for people with chronic health conditions. For this study, the model was adapted using the principles of sharing content and supporting engagement on dedicated public Facebook pages to demonstrate the utility of the OHC model.

The Health Union OHC model encourages people to take an active role in their health by providing content that aligns with their needs and interests and by cultivating a safe environment where communication, understanding, and meaningful relationships can thrive. This original content can be accessed on condition-specific websites and their corresponding social media pages. In this study, we refer to this model as the "Health Union Online Health Community Adaptive Engagement Model," and in previous publications have referred to this as the "meeting people where they are at" model [40].

The OHC adaptive engagement model is developed with the belief that OHCs can be defined and have discernible features. First, similar to in-person communities, OHCs are situational, contextual, and involve a community of choice. There are 3 structural elements of the OHC adaptive engagement model: social support, adaptive engagement, and active moderation. Some features of OHCs can be categorized into more than one of the structural elements as there is often an overlap. These features include the following:

1. Social identity: OHCs, by definition, are for people with a health condition, caregivers, or others who have an interest in the health condition. People who live with a given chronic health condition may choose to identify as a person living with the condition or not, but regardless, being part of an OHC confers a social identity.
2. Social relationships: Relationships with others in OHCs are social relationships. This means they are imbued with a range of characteristics, including reciprocity, mutual support, and emotional connection, with the potential for shared meaning and conflict [41].
3. Interdependence: Being part of a social group means that there is a connection or relationship with others. A relationship entails interdependence in terms of

XSL•FO
RenderX

communication, expectations, behaviors, and other aspects of the relationship [41].

4. Experience of societal stigma: Chronic health conditions, because of their very definition as a state of ill health and the fear of contagion that underlies notions of illness in our society, bear a level of social stigma. Some health conditions bear more stigma than others, and individuals living with these health conditions may feel highly stigmatized if there are visible markers of their health condition or when their diagnosis is known to others [42,43]. The experience of stigma and people sharing how they confront or cope with it is an important aspect of community bonding.
5. Boundary marking: The community is defined by the diagnosis or experience of a health condition. Although there are usually clear definitions and an understanding of who is "living with the health condition," these are based on self-disclosure. In addition, these boundaries are porous and let in new members [44].
6. Social norms: Norms that define interactions of the larger society are represented within OHCs. Often, these social norms are communicated in the form of "community rules," and OHC participants may need reminders of the rules from time to time. There may be formal (warning and banning from the group) and informal (encouragement to use respectful language) sanctions for those who violate these community rules.

Methods

Overview of Methods

The Health Union OHC adaptive engagement model was developed to connect people with information that meets their health needs and as a means of providing social support [32,45]. This model includes a paid community management team encompassing dedicated community leads as well as peer moderators and content contributors who are living with a chronic health condition, caregivers, or health care professionals. Community managers also serve as moderators and may contribute to content [46].

OHCs included in this study are Health Union online communities created for people living with multiple sclerosis, migraine, irritable bowel syndrome, rheumatoid arthritis, lung cancer, and prostate cancer. Each OHC has a website with a dedicated URL and corresponding linked social media channels, including a condition-specific public Facebook page. These OHCs were launched in the following order: Migraine, December 2010 [47]; Multiple Sclerosis, March 2013 [48]; Rheumatoid Arthritis, July 2013 [49]; Irritable Bowel Syndrome, June 2016 [50]; Lung Cancer, January 2017 [51]; and Prostate Cancer, December 2017 [52].

Original content is published on each of the community websites daily and is shared via the corresponding social media pages, including Facebook. All Facebook posts include a link to the original article that was first published on the corresponding community website. These Facebook posts include a customized clickable image with a short engaging message (approximately 80 characters, hereafter referred to as Facebook post copy) that links to an article published on the condition-specific community website. For this study, comments on Facebook posts were the primary unit of analysis. Any comments that were made directly on the condition-specific websites were excluded from this analysis.

Facebook analytics data were used to determine the reach and engagement of the 6 OHCs. The mixed methods combined content analysis (CCA) [53] was used to analyze the comments on 39 sampled Facebook posts from the 6 OHCs identified above to identify themes that aided the authors in illustrating how key aspects of the adaptive engagement model facilitate the different types of social support and other constructs of the model. Additionally, this analysis examined how the adaptive nature of this model can be used to support people regardless of their chronic health condition.

Facebook Reach and Engagement Data

Reach and engagement data were calculated using Facebook's analytic tool, Facebook Insights. In this study, the authors used the Facebook definition of engagement as users reacting to (ie, clicking a reaction button such as "like"), sharing, commenting, or clicking on any posted content; in this case, a link to an original article on the corresponding community website. Reach refers to the number of people who saw the Facebook post, image, and short copy either in their newsfeed or by directly visiting the Facebook page. The reported gender and age cohort of those who reached and engaged were also collected. Weekly averages of reach and engagement were calculated for posts spanning the period from January 1 to June 30, 2018 (Table 1).

XSL•FO
RenderX

Table 1. Average weekly organic reach and engaged users per Facebook community (Facebook insights data from RheumatoidArthritis.net, ProstateCancer.net, MultipleSclerosis.net, Migraine.com, LungCancer.net, IrritableBowelSyndrome.net Pages, January 1 to June 30, 2018).

Facebook insights data	Rheumatoid arthritis	Prostate cancer	Multiple sclerosis	Migraine	Lung cancer	Irritable bowel syndrome
Weekly reach[a] (unique users), n	104,874	6617	160,906	161,958	9328	11,692
Engaged[b], n (%)	9176 (8.75)	484 (7.31)	12486 (7.76)	11791 (7.28)	561 (6.02)	1073 (9.18)
Gender[c], n (%)						
Female	7800 (85.00)	131 (27.1)	10613 (84.99)	10730 (91.00)	471 (83.9)	987 (91.98)
Male	1293 (14.09)	353 (72.9)	1786 (14.30)	987 (8.37)	86 (15.3)	78 (7.23)
Age cohort[c] (years), n (%)						
18-34	403 (4.39)	9 (1.9)	1024 (8.20)	2052 (17.40)	18 (3.2)	95 (8.85)
35-44	890 (9.69)	15 (3.1)	1872 (14.99)	2712 (23.00)	26 (4.6)	129 (12.02)
45-54	2019 (22.00)	48 (9.9)	2997 (24.00)	2948 (25.00)	56 (9.9)	182 (16.96)
55-64	3028 (32.99)	136 (28.1)	3371 (26.99)	2122 (17.99)	180 (32.1)	278 (26.00)
>65	2753 (30.00)	266 (54.9)	3122 (25.00)	1651 (14.00)	275 (49.0)	365 (34.02)
Total Facebook followers[d], n	115,232	8579	125,250	170,219	19,461	25,059

[a]Facebook Insights data, average weekly organic reach (unique users).

[b]Facebook Insights data, average weekly engaged users as the proportion of weekly organic reach.

[c]Gender and age for Facebook-defined engaged users; data reported from user profiles.

[d]Facebook Insights data; total number of followers as of June 30, 2018.

Sampling and Data Collection

A diverse range of Facebook posts linking to original content from the 6 OHCs was purposefully sampled, resulting in a sample of 39 Facebook posts and their respective comment threads for a total sample of 5881 comments. Each selected post received comments from community members on Facebook. The content of the posts and comments included a broad range of topics, including impact on relationships, coping, quality of life, and treatment experience, all of which comprise the patient journey for people with a chronic health condition.

Using the principles of CCA, it is important to distinguish between the sampling unit, contextual unit, and coding unit as part of the methodology and the resultant analysis [53]. In this study, the Facebook posts with their corresponding extracted comments are the *sampling units*; the Facebook post copy along with the linked article, comments, and/or comment threads are the *contextual units*; and a sentence or phrase within an individual comment (as comments can range from a single word or phrase to a paragraph) are the *coding units*.

A purposeful sampling of the original articles and corresponding Facebook posts was drawn based on the following criteria: a range of content reflecting shared experience and sense of belonging of community members, stigma, and other challenges of living with a chronic health condition. An additional selection criterion for purposeful sampling was that the post needed to have a minimum of 20 comments to be included in the sample.

The Facebook posts that were analyzed for this study (Multimedia Appendix 1) [54-92] were originally posted on one of the OHCs' Facebook pages between February 11, 2017, and June 5, 2018. The comments from these Facebook posts were extracted into a spreadsheet and imported into a qualitative and mixed method research data web application, Dedoose (Version 8.0.35; SocioCultural Research Consultants, LLC) [93]. Comment transcripts were labeled by community website and Facebook post ID (eg, PC 6 is used to label a specific Facebook post about PC on a specific date/time and all comments associated with that post). A total sample of 5881 comments were distributed as follows: PC, 1576; RA, 1417; migraine, 876; LC, 735; MS, 701; and IBS, 576.

Combined Content Analysis of Facebook Comments

CCA is a mixed methods approach used to analyze data from social media platforms [53]. This approach allows for the combination of qualitative and quantitative methods, inductive and deductive coding procedures, and manual and automated analytic modes [53]. As stated above, the authors used CCA as a methodological approach for deriving and analyzing a sample of Facebook post comments.

Virtual ethnography methods were used to develop a coding schema using an iterative deductive process [94]. As the authors serve as community managers and moderators, they are immersed in these communities on a day-to-day basis and have the ability to understand and tease apart some nuances in these conversations.

The analysis was carried out in 3 stages. In the first stage, the authors used a randomly selected subset of the sampled comments to identify initial thematic codes using content analysis. Using a concept-mapping process aided by Dedoose, the authors developed a conceptual framework to organize and further analyze the data to achieve theoretical saturation of themes rather than generalizability.

XSL•FO
RenderX

These initial coding samples were each tested by pairing the authors to confirm the interrater reliability (IRR) of the coding schema. The IRR of initial coding was calculated using Pearson correlation coefficient, and coding testing was repeated until the IRR achieved 90% agreement.

The second stage consisted of an in-depth review of these codes by the team, combining and collapsing codes when appropriate and developing the final coding schema with examples to generate the full codebook (Multimedia Appendix 2).

In the final round of coding, the full sample of Facebook post comments was coded by the authors, and an analytical set of thematic categories was identified throughout the full data set using Dedoose. As stated above, comments can be of variable length, from a phrase to several sentences. Code applications to a comment were not mutually exclusive (eg, a given comment could have multiple codes applied). The final set of codes was then juxtaposed against the thematic categories in the analysis to determine their relationship with the structural elements of the OHC model.

The qualitative analysis of the coded comments and patterns of code distributions were used to identify thematic categories that were further compared for similarity and differences across each of the 6 OHCs and by content descriptive category (eg, a comment about symptoms or treatment).

Results

Facebook Reach and Engagement

Facebook Insights data, which show the average organic reach (unique users) and engaged users per week for each of these OHCs, is presented in Table 1. Overall, the 6 OHCs reached an average of 455,375 people per week and engaged approximately 6% to 9% of them, which is considered a high level of engagement on a Facebook post [95]. Individual OHCs reached a range of users: from 6617 people in the PC OHC to 161,958 in the migraine OHC. The gender and age cohort of engaged users, as self-reported in individual user Facebook profiles, is presented to show the demographics of the participants. Except for the PC OHC, females made up a large majority of OHC users.

Facebook Post Comments and Model Elements

Qualitative thematic network analysis of comments from the OHCs corresponds to the structural elements of the Health Union OHC adaptive engagement model—social support, adaptive engagement, and active moderation. Examples of comments that typify each of the structural elements of the model as well as the frequency with which each theme occurred in the comments are presented in Table 2. Overall, the majority (3405/5881, 57.89%) of the comments included one or more codes that represent an element of social support. In addition, almost one-third (1758/5881, 29.89%) of the comments included codes that represent adaptive engagement, and 12.21% (718/5881) reflected an element of active moderation.

Social support was demonstrated in a variety of ways, including giving advice, providing encouragement, and expressing empathy. Sharing knowledge and experience is one such manner of conveying support, with one commenter stating, “There’s many days when these posts remind me I’m not the only one having to deal with these same issues.” This is especially important as people suffering from chronic illness often feel that there are few people who truly understand what they are going through. For example, the previous commenter went on to post:

> *I constantly get looks and attitudes of ‘you look too young to be so sick, you don’t look sick.’ Even from naïve Dr’s [sic] on occasion and often have people attempt to tell me such. It’s infuriating.*

Having a shared experience helps community members to feel understood and less alone, with another commenter stating:

> *Unfortunately, what I have learned is that unless you are someone who struggles with it, the rest of the world (including some family and health professionals) just doesn’t care. There will be no support nor empathy. Hence why I am thankful for this site.*

Community participants share encouragement and express their willingness to be there for each other through hard times. The simple existence of the community, the daily content posted, and the experiences and support shared by users can help those who suffer from chronic illnesses feel less alone. This is shown by comments such as this one:

> *Your articles seem to come out at the perfect time and speak right to me. It’s as if I unknowingly wrote them myself. Thank you for being the biggest support system I have outside of my family.*

Effective moderation by community managers helps to keep the discussion respectful and can help ameliorate some of the harmful effects that are often encountered on the internet, such as misleading medical advice, conspiracy theories, and the like.

XSL•FO
RenderX

Table 2. Exemplary quotes demonstrating qualitative themes and their frequency distribution.

Thematic primary code Subcode	Exemplary quote	Frequency distribution, n (%)
Social support (n=3405)		
Empathy and understanding	"I'm glad this post came up. Most people who have never had a Migraine cannot understand the degree of pain and the strange feeling."	1089 (31.98)
Knowledge and shared experience from community	"Sometimes I see some of your posts and I think YAY it's not just me…and your column makes me feel like it is not only just me, but it's me and 5 million of my closes [sic] friends."	1587 (46.61)
Express stigma	"A lot of people don't think I have a "real" illness, and that includes healthcare professionals. It is probably the most frustrating part of this illness…sometimes I am treated like a drug-seeker, and sometimes I am treated like a nutcase…The only battle worse that getting healthcare professionals to accept my illness and treat it, is getting my family members to accept and understand it as real, and not blame me for it."	381 (11.19)
Giving advice	"The monster is never very far away. Write a journal, buy a dog, start doing your bucket list. The busier you stay the better."	283 (8.31)
Advocacy and awareness	"It has helped me research, it has 'pushed out' things that I need to learn, it has provided a community that I can engage with and be a part of. It has saved me from feeling so alone."	133 (3.91)
Sense of belonging/group membership	"I do get much comfort coming here knowing I am not alone in this world."	378 (11.10)
Encouragement and motivation	"I applaud your courage and determination to have stayed the course… I plan to stay the course as well."	579 (17.00)
Caregiver perspective	"I am caregiver for my husband who is in advance stages. It is becoming more apparent that he will lose the battle due to respiratory issues…Reading your entries remind me that others have gone before me as caregivers for this disease that sucks."	180 (5.28)
Adaptive engagement (n=1758)		
General (tag person, agreement, and emoji)	"[tagged person] I'm glad this post came up…When you have the time, please read some of what people say in this post."	432 (24.57)
Information seeking	"My question is with 15-16 migraines a month, if not more, how [do] you work through them… Any tips would be greatly appreciated!"	332 (18.88)
Information giving	"I have learned there's new technology they do in FL and NY called Sperling Prostate Cancer Treatment Laser…"	272 (15.47)
Conflicts/difference of opinion	"I'm sorry, but I think your outlook is incorrect. You should not avoid asking an important question in order to spare someone's feelings. The goal should be to remove the negativity toward people that do/did smoke, instead of simply not asking the question."	51 (2.90)
Community norms and rules	"Thank you for voicing your concern…I did want to take a moment to clarify that migraine.com is not owned by a pharmaceutical company. You can visit the site for further information about who we are and what we do: https://health-union.com/."	49 (2.79)
Boundary marking	"Hopefully you'll be able to enjoy some support and kinship here, with lots of folks who have similar experiences and can relate to what you've said."	51 (2.90)
Active moderation (n=718)		
Empathy and understanding	"I value you and am grateful to have you as part of our community…we can all relate to one another…and it is comforting to hear a kind response from you and others."	432 (60.2)
Share knowledge and resources	"When you have the time, please read some of what people say in this post. I hope the stories help to make those people without migraines understand that it is different for all people and … try to understand odd behavior and debilitating pain."	298 (41.5)
Share patient/caregiver experiences	"I am so sorry to hear about the mental health struggles you have alongside your migraine. As we speak, I am 1-2 hours into a sudden and dramatic downturn in mood, and it has me worried about a potential bad migraine."	207 (28.8)
Resolve conflicts	"It's never our intention to make anyone feel that they are difficult to love- I believe the article was trying to address how difficult chronic pain can be in the relationship dynamic, but I hear you, and apologize that was the message that came across. Thanks for your feedback and for being part of our community."	25 (3.5)

XSL•FO
RenderX

Thematic primary code Subcode	Exemplary quote	Frequency distribution, n (%)
Maintain community rules	"Since the community is bound to disagree at some points when discussing politics, we wanted to post a friendly reminder that posts of all opinions are welcome as long as they are respectful of each other and not inflammatory."	24 (3.3)

As shown in Table 2, social support is most frequently demonstrated by sharing knowledge of their health condition and experience among community members (1587/3405, 46.61%) and expressing empathy and understanding (1089/3405, 31.98%). Conversely, very rarely do members provide social support by giving advice (283/3405, 8.31%). Most likely, this is because the OHC community rules explicitly state that we do not provide medical advice [96]. Community managers mirror the behavior of discouraging giving advice and reminding participants that people may use a range of treatment options and have different treatment experiences.

From the perspective of the content featured in the Facebook posts and their respective linked articles, the topics discussed in the comments were a broad range. They included caregiver experiences, complications and comorbidities, complementary therapies, coping with the health condition, diagnosis, emotional impact, experiences with health care provider(s), insurance coverage or disability issues, laboratory and other tests, life impact of chronic illness, lifestyle measures, mental health, patient journey, relationship issues, sexual performance and sexual health, and treatment or specific medication discussions (Table 3).

Table 3. Frequency distribution of model elements by topic discussed.

Topic discussed	Number of comments per topic, N	Social support comments, n (%)	Adaptive engagement comments, n (%)	Active moderation comments, n (%)
Caregiver experience	128	102 (79.6)	18 (13.8)	8 (6.9)
Complications and comorbidities	265	173 (65.3)	81 (30.7)	11 (4.0)
Complimentary alternative medicine	142	78 (55.0)	50 (35.0)	14 (10.0)
Coping with condition	358	275 (76.8)	60 (16.8)	23 (6.4)
Diagnosis	181	125 (69.1)	32 (17.7)	6 (3.1)
Emotional impact	425	282 (66.4)	102 (24.0)	41 (9.6)
Health care provider experience	295	139 (47.1)	111 (37.5)	45 (15.3)
Insurance coverage and disability	213	80 (37.5)	86 (40.4)	47 (22.0)
Laboratory and other tests	326	202 (61.9)	49 (15.0)	75 (23.0)
Life impact of chronic illness	458	332 (72.4)	112 (24.2)	14 (3.0)
Lifestyle measures	116	64 (55.2)	28 (24.1)	24 (20.1)
Mental health	110	76 (69.1)	27 (24.5)	7 (6.4)
Patient journey	527	384 (72.9)	91 (17.3)	52 (9.9)
Relationship issues	303	199 (65.7)	87 (28.7)	17 (5.6)
Sexual performance and sexual health	122	77 (63.1)	21 (17.2)	24 (19.7)
Symptoms	725	460 (63.4)	234 (32.3)	31 (4.3)
Treatment discussion	1187	570 (48.02)	472 (39.76)	145 (12.21)

As seen in Table 3, the original Facebook post content topic is often associated with different levels of engagement, with some content topics requiring more moderation than others. In addition, the social support provided or solicited within the context of content topics can differ greatly. For example, 72.4% (332/458) of comments where the life impact of the health condition is discussed also include asking or giving social support, whereas only 48.02% (570/1187) of comments where treatment is discussed include a social support message. However, it is notable that regardless of the content topic, there are still high levels of social support for others expressed. Even for a *dry* topic like insurance coverage, 37.5% (80/213) of the comments were an expression of some type of support.

Content topics that have the highest proportion of comments that needed moderation of some sort, not surprisingly, are topics that are the most medical and/or technically scientific (Table 3). These are comment topics that also include insurance coverage and disability issues 22.0% (47/213), laboratory and other tests 23.0% (75/326), health care provider experiences 15.3% (45/295), and treatment discussions 12.21% (145/1187). These are also content topics with the corresponding lowest percentages of social support codes. These comment topics are more factual and may not elicit the same needs to provide

XSL•FO
RenderX

validation or seek empathy and understanding compared with other topics. This may reflect the perception that these topics are more centered on individual experiences, and other community members may perceive that they require less social support than other content topics.

In addition to differing across content topics, the distribution of codes falling into each of the model's thematic categories also differs depending on the OHC from which the comments arose. As shown in Table 4, social support was highest in the RA and migraine communities and lowest in the PC and IBS communities, although it should be noted that in these groups as well, social support was reflected in 49.49% (780/1576) and 40.3% (232/576) of the comments, respectively.

Table 4. Comment counts and frequency distribution of model elements by community Facebook page.

Facebook page	Comment count, N	Social support comments, n (%)	Adaptive engagement comments, n (%)	Active moderation comments, n (%)
All community FB[a] pages	5881	3405 (57.89)	1758 (29.89)	718 (12.21)
RheumatoidArthritis.net (FB)	1417	860 (60.69)	517 (36.48)	40 (2.82)
ProstateCancer.net (FB)	1576	780 (49.49)	440 (27.92)	356 (22.60)
MultipleSclerosis.net (FB)	701	377 (53.8)	296 (42.2)	28 (3.9)
Migraine.com (FB)	876	479 (54.7)	309 (35.3)	88 (10.0)
LungCancer.net (FB)	735	397 (54.0)	255 (34.7)	83 (11.3)
IrritableBowelSyndrome.net (FB)	576	232 (40.3)	199 (34.5)	145 (25.2)

[a]FB: Facebook.

Discussion

Principal Findings

Results from the reach and engagement data on Facebook demonstrate that people with chronic health conditions want to engage on the web, particularly using social media. Overall, the 6 OHCs reached an average of 455,375 people per week and engaged approximately 6% to 9%, which is considered a high level of engagement on a Facebook post [95]. In addition to the benefits for those who actively engage with the content, previous research has shown that Facebook support groups provide group members who are less active with an important support network in the form of emotional support, informational support, and special companionship [34].

This highlights the importance of creating web-based spaces for the dissemination of information and facilitation of discussion regarding chronic disease. For some health conditions, people may suffer in silence because of social isolation, and the stigma of the condition may mean that online communities provide the only outlet for people to find others that are like them. In addition, people with more severe symptoms, such as those with severe MS, may look to share their experiences with others and provide support to those at earlier stages in their disease journey.

Social support is demonstrated in a variety of ways, including through information sharing, expressions of empathy, and sharing encouraging statements with others [21]. The most common manner of providing social support through Facebook comments was through the sharing of knowledge and experiences unique to people living with the chronic condition (47% of comments had this code). This shared lived experience helps members of the community, or people who visit the Facebook page, feel a bond with others living with the same health condition. This type of online community can then provide an open space for sharing things that are not necessarily understood by users' in-person confidantes.

Previous research has shown that information seeking from people with similar experiences leads to greater perceived empathy among members of OHCs [37]. Allowing space for the expression of feelings and experiences that are not well understood by people without the condition likely helps to combat the stigma faced by people with chronic health conditions in their everyday lives. Importantly, as described above, it was much less common for these OHC members to give explicit advice to others (only 8% of comments had this code). This is important to mention as one of the common criticisms or concerns about social and peer support for health conditions is the possibility of posters sharing misinformation or pushing one's own treatment experience on others [18]. Community guidelines that discourage giving explicit medical advice, as well as active moderation that reminds people of this rule and provides for fair balance in information sharing, are important features of this model. These OHC moderation features allow for the sharing of knowledge and experience without the downside of inaccurate and potentially misleading or dangerous unqualified medical advice.

Knowing when to let community members engage in an exchange of comments without intervening as a moderator is as important as knowing when to apply specific moderation principles. Moderation sometimes involves enforcing community rules or removing comments that violate these rules. Maintaining a safe community environment that presents balanced and medically accurate information is essential to OHCs. When needed, moderators intervene to guide the conversation away from explicit giving of advice and emphasizing that all are welcome to share their experiences, but to be mindful that everyone has different experiences with certain treatments or procedures and there is most often no one right approach or answer.

XSL•FO
RenderX

Previous research has shown that moderators can serve several functions in OHCs: provide clinical expertise, suggest users talk to their physician about their specific question, point users to outside resources, provide community expertise, build rapport, and provide technical help [97]. These results add to the finding that moderation is needed more frequently when discussing scientific and technical subjects, such as labs, diagnostic tests, insurance, and disability, than when the focus is on sharing experiences such as those related to coping or relationship issues.

The skillful use of moderation helps these OHCs function smoothly and deal with any concerning issues as they arise, ensuring that the community is not endorsing misinformation or one specific treatment approach. Additionally, moderation needs may change over time as OHCs gain maturity. The frequency of codes that reflect the need for active moderation were lowest in the migraine, MS, and RA communities. These were launched by Health Union several years before the others and represent more *mature communities* from the perspective of community management. Several factors, including the length of time a community has been active, the number of active community members, and the frequency of member engagement are associated with more mature OHCs.

Analysis of the Health Union OHC adaptive engagement model in action demonstrates that the foundational elements (social support, adaptive engagement, and active moderation) are skillfully used to provide a rich and dynamic community experience for individuals with chronic health conditions. In addition, this analysis indicates that in an open community, social support can be provided and received through the use of active moderation and adaptive engagement. This standardized structure underlying the OHCs allows for adaptation to fit the specific needs of each particular community and provides a basis to facilitate the effective and beneficial exchange of social support.

Although the authors examine a proprietary OHC model, the features of this model, as outlined in this study, are part of other OHCs as well as social media apps and proprietary health apps. Thus, these findings may be extrapolated to these other platforms as well. How consumers or people living with chronic health conditions seek, find, and use information and support is important to a range of consumer health research and applications. Patterns of use and behavior identified herein can also be used as social media apps try to adapt their most notable features to better serve OHCs.

Another implication of this research could be applied to stand-alone health apps and the need for those apps to incorporate the ability to provide moderation to better mirror the experience of OHCs. Pharmaceutical and biotechnology companies interested in social media as a means of engaging communities would be well served to consider the implications of adverse events reporting and the role of active community moderators in maintaining safe spaces for engaging communities of people with chronic health conditions.

Limitations

The Facebook analytics data and the qualitative content of the Facebook comments presented in this study are descriptive and exploratory and are not designed for formal hypothesis testing. This study was initiated to analyze the features of support and engagement that are found organically within OHCs and enhanced through active moderation. We sought to elucidate these themes and understand whether they were consistent with the constructs of the theoretical model.

We note that a central feature of the model, that it is adaptive, is also a limitation of this research design. There is no control group to validate the extent to which social support is an inherent characteristic of OHCs or the additional effect added by features that are specific to the Health Union OHC adaptive engagement model, such as paid moderators and community rules. However, the fact that a high degree of *self-moderation* is observed among community members reacting to each other's comments in a mutually supportive manner in the more mature communities is noteworthy.

As noted in the methods section, given the desire to demonstrate the adaptability of the OHC adaptive engagement model to a social media platform, the authors intentionally only used comments from the OHC Facebook posts for this study and did not include any comments from the linked OHC native domains. As such, the findings from online Facebook communities may not be generalizable to online communities that do not have the same interaction features as this social media platform. However, like Facebook, the trend is for OHCs, in general, to incorporate both passive and active engagement features [35].

To develop a manageable and meaningful sampling frame for this study, the random selection of posts was not feasible, which poses an additional limitation. The sampling frame was driven by the need to ensure representative content types and a large enough sample of community participant comments to determine thematic relevance and redundancy.

A final limitation of this study is the lack of longitudinal data for individual community participants. This study and other research completed by the authors to date present analyses of aggregate data from OHCs, and we have replicated the model several times with success [98]. Future research will be needed to follow distinct cohorts of OHC participants over time to assess the long-term impact of community participation on health-related behaviors and outcomes.

Conclusions

OHCs, particularly those with clear community rules and active moderation, can offer a safe and supportive environment for people living with chronic health conditions. Despite the inherent limitations of the internet and social media, OHCs are dynamic communities of people engaging in social relationships. The dynamic nature of OHCs may make them challenging to research, but it is possible to examine the structural elements of a model in action as well as explore qualitatively and quantitatively the aspects of interactions [99].

Reaching people on the web is a promising approach to facilitate the exchange of social support among people with chronic health

XSL•FO
RenderX

conditions. Health care messages bombard patients on a daily or even hourly basis. Using social media to deliver a message and provide a space for people to engage with others of similar condition-specific interests is a valuable way to cut through that noise.

OHCs such as those analyzed herein allow people with chronic health conditions to learn and engage with others on their own time and at their own pace. However, it is important to design and maintain such OHCs and their associated social media sites as safe spaces, and a structured model with essential features such as those described in this study facilitates their success. The use of a well-designed model helped the OHCs studied to achieve their primary aim: to meet people where they are at with the health information and social support they seek.

Community-responsive content, or content tailored to the needs of the community, is the foundation of an OHC. However, content alone does not provide sufficient experience for an engaging OHC. It is important to provide both active and passive opportunities for people to engage with other community members. In addition, active moderation is needed to ensure a balance of views, appropriate tone, and recognition of community rules and values.

Future research is needed to identify segments of OHCs that have different needs and different patterns of interactions. A comparative analysis of these community segments may be used to help design the features of online communities that provide for more real-time interaction, engagement features, and personalized content.

Focusing on the ongoing relationship with community members, which is enhanced through active moderation and adaptive engagement, provides a community experience that is mutually supportive and results in a healthy online community that can thrive and mature. This study has practical significance as it helps to demonstrate the value of OHCs for people living with chronic health conditions and provides support for the use of an underlying structured model to guide community interactions.

Acknowledgments

The authors would like to thank Michael Booth, PhD candidate, University of Michigan, and RheumatoidArthritis.net community advocate, for reviewing the previous versions of the Health Union OHC adaptive engagement model. The authors further recognize the Health Union community management team, user experience/creative team, and other Health Union business teams for the infrastructure that sustains the OHCs. The authors are indebted to Tim Armand and Olivier Chateau for founding Health Union, LLC, and contributing to the vision that inspired the OHC adaptive engagement model. Finally, the authors thank all our community members who contribute content, engage daily, and provide ongoing support to each other.

Conflicts of Interest

BMG, KTVH, KG, SH, and AB are employees of Health Union, LLC, whose OHCs and OHC adaptive engagement model are discussed and evaluated herein. The authors do not financially benefit from this publication.

Multimedia Appendix 1
Facebook posts linked articles.
[DOCX File , 20 KB - jmir_v22i7e17338_app1.docx]

Multimedia Appendix 2
Codebook.
[DOCX File , 17 KB - jmir_v22i7e17338_app2.docx]

References

1. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003 Apr;30(2):170-195. [doi: 10.1177/1090198102251030] [Medline: 12693522]
2. Uchino BN. Social support and health: a review of physiological processes potentially underlying links to disease outcomes. J Behav Med 2006 Aug;29(4):377-387. [doi: 10.1007/s10865-006-9056-5] [Medline: 16758315]
3. Heaney C, Israel BA. Social networks and social support. In: Glanz K, Rimer B, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. Fourth Edition. San Francisco, CA: Jossey-Bass; 2008:189-210.
4. Albrecht T, Adelman M. Communicating Social Support. Newbury Park, CA: Sage Publications; 1987.
5. Ertel KA, Glymour MM, Berkman LF. Social networks and health: a life course perspective integrating observational and experimental evidence. J Soc Pers Relatsh 2009 May 13;26(1):73-92. [doi: 10.1177/0265407509105523]
6. Everson-Rose SA, Lewis TT. Psychosocial factors and cardiovascular diseases. Annu Rev Public Health 2005;26:469-500. [doi: 10.1146/annurev.publhealth.26.021304.144542] [Medline: 15760298]
7. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]

XSL•FO
RenderX

8. Uchino BN, Cacioppo JT, Kiecolt-Glaser JK. The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. Psychol Bull 1996 May;119(3):488-531. [doi: 10.1037/0033-2909.119.3.488] [Medline: 8668748]
9. Gonzalez JS, Penedo FJ, Antoni MH, Durán RE, McPherson-Baker S, Ironson G, et al. Social support, positive states of mind, and HIV treatment adherence in men and women living with HIV/AIDS. Health Psychol 2004 Jul;23(4):413-418. [doi: 10.1037/0278-6133.23.4.413] [Medline: 15264978]
10. Catz SL, Kelly JA, Bogart LM, Benotsch EG, McAuliffe TL. Patterns, correlates, and barriers to medication adherence among persons prescribed new treatments for HIV disease. Health Psychol 2000 Mar;19(2):124-133. [Medline: 10762096]
11. Clinical Practice Guideline Treating Tobacco UseDependence 2008 Update Panel, Liaisons,Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A US public health service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
12. Wing RR, Jeffery RW. Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. J Consult Clin Psychol 1999 Feb;67(1):132-138. [doi: 10.1037//0022-006x.67.1.132] [Medline: 10028217]
13. Gorin A, Phelan S, Tate D, Sherwood N, Jeffery R, Wing R. Involving support partners in obesity treatment. J Consult Clin Psychol 2005 Apr;73(2):341-343. [doi: 10.1037/0022-006X.73.2.341] [Medline: 15796642]
14. Piferi RL, Lawler KA. Social support and ambulatory blood pressure: an examination of both receiving and giving. Int J Psychophysiol 2006 Nov;62(2):328-336. [doi: 10.1016/j.ijpsycho.2006.06.002] [Medline: 16905215]
15. Fox S, Jones S. The Social Life of Health Information. Pew Research Center. 2009 Jun 11. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2009/PIP_Health_2009.pdf [accessed 2020-05-15]
16. Fox S, Duggan M. Peer-to-Peer Health Care. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/peer-to-peer-health-care/ [accessed 2019-11-12]
17. Herbert LB, Hayes S, Bhowmick A, Zerkowski K, Green BM. Facebook...the Destination for Health Information? Understanding Patients' Use of Social Media for Health. In: American Public Health Association Annual Meeting and Exposition. 2017 Presented at: APHA'17; November 4-8, 2017; Atlanta, GA URL: https://apha.confex.com/apha/2017/meetingapp.cgi/Paper/384494 [doi: 10.4135/9781412971942.n31]
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
19. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
20. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
21. Wang X, Zhao K, Street N. Social Support and User Engagement in Online Health Communities. In: Proceedings of the International Conference on Smart Health. 2014 Presented at: ICHS'14; July 10-11, 2014; Beijing, China. [doi: 10.1007/978-3-319-08416-9_10]
22. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The effects of a digital well-being intervention on patients with chronic conditions: observational study. J Med Internet Res 2020 Jan 10;22(1):e16211 [FREE Full text] [doi: 10.2196/16211] [Medline: 31922491]
23. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. J Health Psychol 2014 Dec;19(12):1613-1625. [doi: 10.1177/1359105313496446] [Medline: 23988679]
24. Burke M, Kraut RE. The relationship between Facebook use and well-being depends on communication type and tie strength. J Comput-Mediat Comm 2016 Jul 26;21(4):265-281. [doi: 10.1111/jcc4.12162]
25. DiJulio B, Hamel L, Muñana C, Brodie M. Loneliness and Social Isolation in the United States, the United Kingdom, and Japan: An International Survey. Kaiser Family Foundation. 2018. URL: http://files.kff.org/attachment/Report-Loneliness-and-Social-Isolation-in-the-United-States-the-United-Kingdom-and-Japan-An-International-Survey [accessed 2020-05-15]
26. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
27. Coulson NS. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. Cyberpsychol Behav 2005 Dec;8(6):580-584. [doi: 10.1089/cpb.2005.8.580] [Medline: 16332169]
28. Mo PK, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. Cyberpsychol Behav 2008 Jun;11(3):371-374. [doi: 10.1089/cpb.2007.0118] [Medline: 18537512]
29. Braithwaite DO, Waldron VR, Finn J. Communication of social support in computer-mediated groups for people with disabilities. Health Commun 1999;11(2):123-151. [doi: 10.1207/s15327027hc1102_2] [Medline: 16370973]
30. Eastin MS, LaRose R. Alt.support: modeling social support online. Comput Hum Behav 2005 Nov;21(6):977-992. [doi: 10.1016/j.chb.2004.02.024]

XSL•FO
RenderX

31. Malik S, Coulson NS. 'They all supported me but I felt like I suddenly didn't belong anymore': an exploration of perceived disadvantages to online support seeking. J Psychosom Obstet Gynaecol 2010 Sep;31(3):140-149. [doi: 10.3109/0167482X.2010.504870] [Medline: 20718585]
32. Brennan PF, Moore SM, Smyth KA. Alzheimer's disease caregivers' uses of a computer network. West J Nurs Res 1992 Oct;14(5):662-673. [doi: 10.1177/019394599201400508] [Medline: 1529609]
33. Koh J, Kim Y. Knowledge sharing in virtual communities: an e-business perspective. Expert Syst Appl 2004 Feb;26(2):155-166. [doi: 10.1016/s0957-4174(03)00116-7]
34. Steadman J, Pretorius C. The impact of an online Facebook support group for people with multiple sclerosis on non-active users. Afr J Disabil 2014;3(1):132 [FREE Full text] [doi: 10.4102/ajod.v3i1.132] [Medline: 28730005]
35. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
36. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003 Feb 15;97(4):920-925 [FREE Full text] [doi: 10.1002/cncr.11145] [Medline: 12569591]
37. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
38. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
39. Partridge SR, Gallagher P, Freeman B, Gallagher R. Facebook groups for the management of chronic diseases. J Med Internet Res 2018 Jan 17;20(1):e21 [FREE Full text] [doi: 10.2196/jmir.7558] [Medline: 29343460]
40. Bhowmick A, Hribar C. Online Health Communities: A New Frontier in Health Research. Medium. 2016 Aug 25. URL: https://medium.com/@abhowmick1/online-health-communities-a-new-frontier-in-health-research-71fb73edbea2 [accessed 2020-05-15]
41. Delanty G. Community. Third Edition. London, UK: Routledge, Taylor & Francis Group; 2018.
42. van Brakel WH. Measuring health-related stigma--a literature review. Psychol Health Med 2006 Aug;11(3):307-334. [doi: 10.1080/13548500600595160] [Medline: 17130068]
43. Weiss MG, Ramakrishna J, Somma D. Health-related stigma: rethinking concepts and interventions. Psychol Health Med 2006 Aug;11(3):277-287. [doi: 10.1080/13548500600595053] [Medline: 17130065]
44. Preece J, Maloney-Krichmar D. Online communities: design, theory, and practice. J Comput-Media Commun 2005;10(4):-. [doi: 10.1111/j.1083-6101.2005.tb00264.x]
45. Barrera M, Glasgow RE, McKay HG, Boles SM, Feil EG. Do internet-based support interventions change perceptions of social support?: an experimental trial of approaches for supporting diabetes self-management. Am J Community Psychol 2002 Oct;30(5):637-654. [doi: 10.1023/A:1016369114780] [Medline: 12188054]
46. de Souza CS, Preece J. A framework for analyzing and understanding online communities. Interact Comput 2004 Jun;16(3):579-610. [doi: 10.1016/j.intcom.2003.12.006]
47. Migraine. URL: https://Migraine.com [accessed 2020-05-15]
48. MultipleSclerosis. URL: https://MultipleSclerosis.net [accessed 2020-05-15]
49. RheumatoidArthritis. URL: https://RheumatoidArthritis.net [accessed 2020-05-15]
50. IrritableBowelSyndrome. URL: https://IrritableBowelSyndrome.net [accessed 2020-05-15]
51. LungCancer. URL: https://LungCancer.net [accessed 2020-05-15]
52. ProstateCancer. URL: https://ProstateCancer.net [accessed 2020-05-15]
53. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a mixed-methods research approach to content analysis in the digital age: the combined content-analysis model and its applications to health care Twitter feeds. J Med Internet Res 2016 Mar 8;18(3):e60 [FREE Full text] [doi: 10.2196/jmir.5391] [Medline: 26957477]
54. Ringstaff A. You Don't Get It… Till You GET IT. Multiple Sclerosis. 2017 Jun 23. URL: https://multiplesclerosis.net/living-with-ms/till-you-get-it/ [accessed 2018-12-31]
55. Editorial Team. 30 Facts for World MS Day. Multiple Sclerosis. 2018 May 13. URL: https://multiplesclerosis.net/living-with-ms/30-facts-world-ms-day/ [accessed 2018-12-31]
56. Allen GM. How Social Media Has Always Helped Me Overcome MS. Multiple Sclerosis. 2018 Feb 18. URL: https://multiplesclerosis.net/living-with-ms/how-social-media-helped-overcome/ [accessed 2018-12-31]
57. Rhoades E. Parking Lot Life Lessons. Multiple Sclerosis. 2017 Apr 27. URL: https://multiplesclerosis.net/living-with-ms/speaking-of-advocacy/ [accessed 2018-12-31]
58. Teddy S. For Caregivers, a Sixth Stage of Grief. Multiple Sclerosis. 2017 May 17. URL: https://multiplesclerosis.net/topic/for-caregivers-a-sixth-stage-of-grief/ [accessed 2018-12-31]
59. Editorial Team. How Has MS Impacted Your Friendships and Relationships? Multiple Sclerosis. 2018 Jan 31. URL: https://multiplesclerosis.net/q-and-a/impacted-friendships-relationships/ [accessed 2018-12-31]

XSL•FO
RenderX

60. MoniD. I Get Tired of “Fake Fluff” Sometimes. Multiple Sclerosis. 2017 Nov 17. URL: https://multiplesclerosis.net/stories/i-get-tired-of-fake-fluff-sometimes/ [accessed 2018-12-31]
61. LizShelton. Anyone Else Turn Into an Emotional Train Wreck During/After a Migraine? Migraine. 2018 Apr 5. URL: https://migraine.com/q-and-a/emotional-wreck-during-after/ [accessed 2018-12-31]
62. Workman A. Loving Someone with Chronic Pain. Migraine. 2018 Jan 19. URL: https://migraine.com/living-migraine/loving-someone-chronic-pain/ [accessed 2018-12-31]
63. The Migraine Girl. The Messy Wake of Destruction a Migraine Can Leave Behind. Migraine. 2016 Jul 22. URL: https://migraine.com/living-migraine/the-messy-wake-of-destruction/ [accessed 2018-12-31]
64. Sanders J. Losing Time. Migraine. 2017 Oct 5. URL: https://migraine.com/living-migraine/losing-time/ [accessed 2018-12-31]
65. Hackley S. What Not to Say to Someone with Chronic Migraine: A Note on “Prevention”. Migraine. 2017 May 4. URL: https://migraine.com/living-migraine/a-note-on-prevention-and-what-not-to-say-to-someone-with-chronic-migraine/ [accessed 2018-12-31]
66. Knight K. No Nagging, Please and Thank You. Migraine. 2018 May 19. URL: https://migraine.com/living-migraine/no-nagging-please/ [accessed 2018-12-31]
67. Golden KM. Patient Perspective: Migraine Symptoms. Migraine. 2017 Jul 13. URL: https://migraine.com/living-migraine/patient-perspective-migraine-symptoms/ [accessed 2018-12-31]
68. Editorial Team. IBS Awareness Month. Irritable Bowel Syndrome. 2018 Apr 1. URL: https://irritablebowelsyndrome.net/spotlight/ibs-awareness-month-2018/ [accessed 2018-12-31]
69. Editorial Team. Beyond the Bathroom: The Impact of IBS on Quality of Life. Irritable Bowel Syndrome. 2018 Apr 12. URL: https://irritablebowelsyndrome.net/infographic/beyond-bathroom-impact-ibs-qol/ [accessed 2018-12-31]
70. HessP. What I Wish Others Knew About IBS. Irritable Bowel Syndrome. 2017 Apr 20. URL: https://irritablebowelsyndrome.net/video/wish-others-knew-about-ibs/ [accessed 2018-12-31]
71. Lumpe A. RA Impacts Much More than Bones and Joints. Rheumatoid Arthritis. 2013 Nov 29. URL: https://rheumatoidarthritis.net/living/ra-impacts-much-just-bones-joints/ [accessed 2018-12-31]
72. Wren. RA and the Zombie Apocalypse. Rheumatoid Arthritis. 2015 Jun 12. URL: https://rheumatoidarthritis.net/living/and-the-zombie-apocalypse/ [accessed 2018-12-31]
73. Haag T. Suiting Up … for a Day at the Office. Rheumatoid Arthritis. 2015 Jul 7. URL: https://rheumatoidarthritis.net/living/suiting-up-office/ [accessed 2018-12-31]
74. Mack K. What to Tell Others. Rheumatoid Arthritis. 2015 Sep 4. URL: https://rheumatoidarthritis.net/living/what-tell/ [accessed 2018-12-31]
75. Editorial Team. Community Thoughts: What Do You Wish More People Knew About RA? Rheumatoid Arthritis. 2015 May 15. URL: https://rheumatoidarthritis.net/living/community-thoughts-wish-more-knew/ [accessed 2018-12-31]
76. Wren. Pain is Like an Onion. Rheumatoid Arthritis. 2018 Jan 12. URL: https://rheumatoidarthritis.net/living/pain-is-like-an-onion/ [accessed 2018-12-31]
77. Cristola M. When You Just Don’t Want to Fight It so Hard Any Longer. Lung Cancer. 2018 Jun 1. URL: https://lungcancer.net/stories/dont-want-to-fight/ [accessed 2018-12-31]
78. Petok A. Dear Lung Cancer Patient. Lung Cancer. 2017 Nov 24. URL: https://lungcancer.net/living/dear-patient/ [accessed 2018-12-31]
79. Donaldson D. Stigma Kills. Lung Cancer. 2017 Feb 10. URL: https://lungcancer.net/living/stigma-kills/ [accessed 2018-12-31]
80. Poehlmann J. Coping With the Pressure of Looking Fine. Lung Cancer. 2017 Oct 6. URL: https://lungcancer.net/living/pressure-of-looking-fine/ [accessed 2018-12-31]
81. Elkins I. Do You Still Get Nervous Before Every Scan. Lung Cancer. 2018 Feb 8. URL: https://lungcancer.net/living/still-get-nervous-every-scan/ [accessed 2018-12-31]
82. Poehlmann J. Protecting Those We Love From Worry. Lung Cancer. 2018 May 3. URL: https://lungcancer.net/living/protecting-love-worry/ [accessed 2018-12-31]
83. Loss K. A Tweak Here, A Twinge There. Lung Cancer. 2017 Jul 27. URL: https://lungcancer.net/living/a-tweak-here-a-twinge-there/ [accessed 2018-12-31]
84. Loss K. The Road to Grateful, Not Hateful. Lung Cancer. 2017 Jul 14. URL: https://lungcancer.net/living/the-road-to-grateful-not-hateful/ [accessed 2018-12-31]
85. Editorial Team. Challenges of Living with Prostate Cancer. Prostate Cancer. 2017 Oct 20. URL: https://prostatecancer.net/living-coping/ [accessed 2018-12-31]
86. Editorial Team. Nerves in Prostate May Be New Treatment Target. Prostate Cancer. 2017 Dec 18. URL: https://prostatecancer.net/clinical/nerves-prostate-treatment-target/ [accessed 2018-12-31]
87. Myers N. The Psycho-Sexual (Side) Effect. Prostate Cancer. 2017 Dec 15. URL: https://prostatecancer.net/living/psycho-sexual-effect/ [accessed 2018-12-31]
88. Editorial Team. Is Pain a Sign of Prostate Cancer? Prostate Cancer. 2017 Nov 10. URL: https://prostatecancer.net/symptoms/pain/ [accessed 2018-12-31]
89. Smith L. The Bombs Are Dropped. Prostate Cancer. 2017 Dec 13. URL: https://prostatecancer.net/living/bombs-dropped/ [accessed 2018-12-31]

XSL•FO
RenderX

90. Editorial Team. What Is Gleason Score? Prostate Cancer. 2017 Nov 11. URL: https://prostatecancer.net/diagnosis/gleason-score/ [accessed 2018-12-31]
91. Jones W. "C-Day:" The Day I Learned I Had Prostate Cancer. Prostate Cancer. 2017 Dec 20. URL: https://prostatecancer.net/living/day-learned-prostate-cancer/ [accessed 2018-12-31]
92. Editorial Team. What Are Treatment Options for Prostate Cancer? Prostate Cancer. 2017 Oct 20. URL: https://prostatecancer.net/treatment/ [accessed 2018-12-31]
93. Dedoose. 2018. URL: https://www.dedoose.com [accessed 2020-05-15]
94. Hine C. Virtual ethnography. In: Lewis-Beck MS, Bryman A, Liao TF, editors. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: Sage Publications; 2004.
95. Cowie JM, Gurney ME. The use of Facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]
96. Community Rules. MultipleSclerosis. 2015. URL: https://multiplesclerosis.net/about-us/community-rules/ [accessed 2020-05-15]
97. Huh J, McDonald DW, Hartzler A, Pratt W. Patient moderator interaction in online health communities. AMIA Annu Symp Proc 2013;2013:627-636 [FREE Full text] [Medline: 24551364]
98. Horrell LN, Lazard AJ, Bhowmick A, Hayes S, Mees S, Valle CG. Attracting users to online health communities: analysis of lungcancer.net's Facebook advertisement campaign data. J Med Internet Res 2019 Nov 4;21(11):e14421 [FREE Full text] [doi: 10.2196/14421] [Medline: 31682589]
99. Franz D, Marsh HE, Chen JI, Teo AR. Using Facebook for qualitative research: a brief primer. J Med Internet Res 2019 Aug 13;21(8):e13544 [FREE Full text] [doi: 10.2196/13544] [Medline: 31411143]

Abbreviations

CCA: combined content analysis
IBS: irritable bowel syndrome
IRR: interrater reliability
LC: lung cancer
MS: multiple sclerosis
OHC: online health community
PC: prostate cancer
RA: rheumatoid arthritis

Edited by G Eysenbach; submitted 08.12.19; peer-reviewed by G Pravettoni, G Marton, L Vergani, K McCausland, S Oser; comments to author 06.01.20; revised version received 01.03.20; accepted 03.06.20; published 07.07.20.

Please cite as:
Green BM, Van Horn KT, Gupte K, Evans M, Hayes S, Bhowmick A
Assessment of Adaptive Engagement and Support Model for People With Chronic Health Conditions in Online Health Communities: Combined Content Analysis
J Med Internet Res 2020;22(7):e17338
URL: https://www.jmir.org/2020/7/e17338
doi:10.2196/17338
PMID:32492651

©Brian M Green, Katelyn Tente Van Horn, Ketki Gupte, Megan Evans, Sara Hayes, Amrita Bhowmick. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characteristics, Symptom Severity, and Experiences of Patients Reporting Chronic Kidney Disease in the PatientsLikeMe Online Health Community: Retrospective and Qualitative Study

Glen James[1], PhD; Elisabeth Nyman[2], MSc; Marcy Fitz-Randolph[3], DO, MPH; Anna Niklasson[2], PhD; Katarina Hedman[2], PhD; Jonatan Hedberg[2], MSc; Eric T Wittbrodt[4], PharmD, MPH; Jennie Medin[2], MPH, PhD; Carol Moreno Quinn[1], MD, PhD; Alaster M Allum[1], MBBS, BA, DipSl; Cathy Emmas[5], PhD

[1]AstraZeneca, Cambridge, United Kingdom

[2]AstraZeneca, Gothenburg, Sweden

[3]PatientsLikeMe, Inc, Cambridge, MA, United States

[4]AstraZeneca, Gaithersberg, MD, United States

[5]AstraZeneca, Luton, United Kingdom

Corresponding Author:
Glen James, PhD
AstraZeneca
136 Hills Rd
Cambridge, CB2 8PA
United Kingdom
Phone: 44 7384538548
Email: glen.james@astrazeneca.com

Abstract

Background: Chronic kidney disease (CKD) is a major global health burden, and is associated with increased adverse outcomes, poor quality of life, and substantial health care costs. While there is an increasing need to build patient-centered pathways for improving CKD management in clinical care, data in this field are scarce.

Objective: The aim of this study was to understand patient-reported experiences, symptoms, outcomes, and treatment journeys among patients with CKD through a retrospective and qualitative approach based on data available through PatientsLikeMe (PLM), an online community where patients can connect and share experiences.

Methods: Adult members (aged ≥18 years) with self-reported CKD within 30 days of enrollment, who were not on dialysis, and registered between 2011 and 2018 in the PLM community were eligible for the retrospective study. Patient demographics and disease characteristics/symptoms were collected from this retrospective data set. Qualitative data were collected prospectively through semistructured phone interviews in a subset of patients, and questions were oriented to better understand patients' experiences with CKD and its management.

Results: The retrospective data set included 1848 eligible patients with CKD, and median age was 56 years. The majority of patients were female (1217/1841, 66.11%) and most were US residents (1450/1661, 87.30%). Of the patients who reported comorbidities (n=1374), the most common were type 2 diabetes (783/1374, 56.99%), hypertension (664/1374, 48.33%), hypercholesterolemia (439/1374, 31.95%), and diabetic neuropathy (376/1374, 27.37%). The most commonly reported severe or moderate symptoms in patients reporting these symptoms were fatigue (347/484, 71.7%) and pain (278/476, 58.4%). In the qualitative study, 18 eligible patients (13 females) with a median age of 60 years and who were mainly US residents were interviewed. Three key concepts were identified by patients to be important to optimal care and management: listening to patient needs, coordinating health care across providers, and managing clinical care.

Conclusions: This study provides a unique source of real-world information on the patient experience of CKD and its management by utilizing the PLM network. The results reveal the challenges these patients face living with an array of symptoms, and report key concepts identified by patients that can be used to further improve clinical care and management and inform future CKD studies.

(J Med Internet Res 2020;22(7):e18548) doi:10.2196/18548

XSL•FO
RenderX

KEYWORDS

community networks; chronic kidney disease; real-world experience; patient experience; retrospective; observational; diabetes; interview; online; social media

Introduction

Chronic kidney disease (CKD) is a major global health concern with an estimated prevalence of 8%-16% worldwide [1,2]. It is associated with an increased risk of adverse outcomes, poor health-related quality of life (HRQoL), and substantial health care costs [1]. The patient voice has become an integral part of the management of chronic diseases and increasingly relevant to the development of new treatments. Therefore, an expanding need exists to build patient-centered pathways for improved CKD management in clinical care [3]. Patient-focused research can help identify priorities and outcomes that are important to patients and can enable the development of care components that can improve health care practice [4-6]. This information has also driven the development of tools that can promote self-management, which has been shown to slow disease progression, specifically in patients with CKD [5,7]. While qualitative studies have identified specific topics of importance to patients with respect to disease management [4,5], there is still an outstanding need for information directly from patients about their experience of symptoms and outcomes in CKD [8-10].

More recently, the collection of data on the patient perspective of new treatments during drug development has been encouraged by the US Food and Drug Administration, which has developed guidance related to the use of patient-reported outcome measures to support drug approvals and label claims [11]. Researchers in academia and industry are also increasingly seeking patients' voices and candid perspectives on a variety of clinical research-related areas through patient advisory boards. These activities have generated valuable insights, leading to improvements in study design (eg, adjustments in visit schedules), patient communication materials (eg, study pamphlets), and the development of new patient-centered endpoints [12].

Patients with serious and chronic diseases increasingly seek information and peer-to-peer support through online communities [13-16]. Online health communities and research networks, such as PatientsLikeMe (PLM), allow patients to connect with other patients, share experiences, and learn more about how to live better with their condition [17,18]. Through their engagement with such communities, patients experience improved emotional and social well-being, as well as increased confidence in their interactions with health care providers, and greater optimism and control in relation to their condition [14,15]. Ultimately, these networks provide an opportunity for clinicians, service providers, and researchers to listen to and better understand patients, and gather information outside of the clinical setting. An additional method for eliciting in-depth information is through patient interviews. Qualitative studies involving semistructured interviews have been scarcely used to understand patients' experiences of living with CKD and gain their perspective on various aspects of their care, so that any unmet needs can be identified and health care provision improved [4,10]. To date, characteristics, treatments, and self-reported outcomes in patients with CKD who are members of the PLM network have not been evaluated.

To improve patient care and to construct patient-centered pathways, it is necessary first to understand how patients experience their disease. The objective of this study was to understand patient-reported experiences, symptoms, outcomes, and treatment management among patients with CKD through a retrospective approach based on data available through PLM, and then using a qualitative approach in a subset of patients from the retrospective study cohort.

Methods

Retrospective Study Population

Adult PLM members (aged ≥18 years) were included in this study if they had registered with PLM between 2011 and 2018, if they had self-reported CKD within 30 days of registration, and if they were not receiving dialysis (N=1848).

Qualitative Study Population

This population was a subgroup of the retrospective study cohort (acknowledging patients' clinical profile, such as dialysis status, may have changed between enrollment on the retrospective study and the interview). All participants were adult PLM members (aged ≥18 years) who had self-reported CKD and were resident in the United States or other English-speaking countries (N=18).

PatientsLikeMe

PLM, described in Multimedia Appendix 1, is an online network for patients with a range of diseases that allows them to connect with other patients and share personal stories and health data. Patients are directed to the website through multiple channels, including paid advertisements, public relations, press mentions, academic collaborations, word of mouth from other patients, provider referral, and web search [18]. Members provide information on their treatment and symptoms, including whether they are experiencing any core symptoms (pain, fatigue, anxious mood, depressed mood, and insomnia) that PLM asks members about regardless of their condition [19]. Participation is voluntary; patients did not receive an honorarium for providing data on the site, and their use of the site was not dependent on the provision of data. However, patients did receive an honorarium for interviews. Confidentiality of patient data was maintained and deidentified data were analyzed.

Study Using Retrospective Data

A cross-sectional observational descriptive study using deidentified retrospective data from the PLM network database was used to characterize patients with self-reported CKD. These data were used under license for this study; restrictions applied to the availability of the data. Demographics and clinical characteristics were described for all patients who provided information. We evaluated data on CKD diagnosis, symptom

XSL•FO
RenderX

severity (both general from the PLM symptom panel and other open-field symptoms), treatments, and comorbidities reported within 30 days of PLM registration. If multiple reports of the same symptom were entered within 30 days of registration, only the first report was included in these analyses. The 30-day period was chosen to reflect the status of patients when joining the PLM network. The data were standardized in PLM; comorbidities and symptoms were standardized to corresponding International Statistical Classification of Diseases and Related Health Problems, Tenth Revision codes, and treatments were standardized to prescription codes. Data were analyzed descriptively, presenting median with interquartile range (IQR), frequency, and proportion as appropriate. The study protocol was performed in accordance with International Conference on Harmonisation Good Clinical Practice and the Declaration of Helsinki, and the applicable legislation on noninterventional studies or observational studies or both. Patients who register on PLM are made aware that patient data will be used for research purposes, and results will be shared with the PLM community.

Study Using Qualitative Data

The aim of this study was to understand the patient's experience of CKD and its management. This involved questioning patients on their experience of CKD diagnosis, treatments, and comorbidities; the impact of symptoms on HRQoL and functioning; disease management, diet, and physical exercise in relation to CKD; and challenges around their interactions with their medical team, to better understand what is missing regarding specific CKD patient-reported management.

The qualitative study was performed in a subset of patients from the retrospective study using semistructured phone interviews (60-90 minutes). PLM members who met the inclusion criteria (ie, individuals who reported CKD as a condition, were over 18 years of age, and resided in the United States or in an English-speaking country) were identified and contacted by private message or email to participate. Those who indicated they wished to participate were contacted by the research team and any questions they had about the study were discussed. Consent forms were shared by email on scheduling, and verbal consent was obtained prior to starting the interview; all received US $100 for participating. Interviews were open ended and followed a semistructured interview guide. Open-ended questions were followed by prompts to encourage more detailed conversations about issues relevant to CKD if a topic had not already been fully explored. All interviews were recorded and transcribed postinterview to ensure direct quotes were accurate; recordings were deleted after they were transcribed verbatim.

Data were descriptively evaluated, presenting median with IQR, frequency, and proportion as appropriate. The interview transcripts were reviewed consecutively in groups of 3 using a content analysis approach [20] to identify emerging themes within the interview guide categories using the ATLAS.ti software (Scientific Software Development GmbH). As additional interviews were coded and new themes arose, the coding schema was updated, refined, and further grouped to create a final list of themes. This process was continued until no further concepts were identified and content saturation had been achieved (ie, no new concepts were being mentioned). Study end was achieved when either 20 participants had been interviewed or concept saturation was reached. Qualitative analyses of individual verbatim interview transcripts included whether symptoms reflected the patient experience without any potential bias that might be introduced by the interview questions [21]. The study protocol, patient communications, and interview guide were approved by the New England Independent Review Board. Any privacy concerns were addressed by not including identifying information in the final report and by allowing only authorized research personnel to access the PLM database.

Data Availability

The data that support the findings of this study were collected from PatientsLikeMe. Restrictions apply to the availability of these data, which were used under license for this study.

Results

Retrospective Data

Demographics and Clinical History

Between 2011 and 2018, 1848 (from 2391) PLM members who self-reported CKD were not receiving dialysis and were included in this analysis (Multimedia Appendix 2). The majority of patients were female (1217/1841, 66.11%) and were US residents (1450/1661, 87.30%), with a median age of 56 years (IQR 45-64, n=1848) at the date they joined PLM (Table 1). Most of those who recorded race were white (1199/1496, 80.15%) and 79.1% (564/713) of patients who shared their education history had postsecondary education. The median age at onset of symptoms that patients associated with CKD was 43 years (IQR 28-54 years; n=378); median age of CKD diagnosis, reported by 578 members, was 47 years (IQR 33-56 years).

XSL•FO
RenderX

Table 1. Patient demographics and clinical history (retrospective study; total patients studied: 1848).

Characteristic	Total
Gender (N=1841), n (%)	
Female	1217 (66.11)
Age (years) at date joined PLM[a] (N=1848), median (IQR[b]); range	56 (45-64); 18-89
Country (N=1661), n (%)	
United States	1450 (87.30)
UK	66 (3.97)
Other	145 (8.73)
Race (N=1496), n (%)	
White	1199 (80.15)
Black or African American	137 (9.16)
Asian	67 (4.48)
Mixed race	66 (4.41)
Other[c]	27 (1.80)
Education (N=713), n (%)	
Postgraduate degree (master's, doctorate, etc.)	94 (13.18)
College bachelor's/undergraduate degree	136 (19.07)
Some college, but less than a bachelor's/undergraduate degree	334 (46.84)
High-school graduate or GED[d] (left school around age 18)	124 (17.39)
Some high school, but did not graduate (left school around age 16)	23 (3.23)
Eighth grade or less (left school around age 14)	2 (0.28)
Age (years) at first CKD[e] symptom (N=378), median (IQR)	43 (28-54)
Age (years) at first CKD diagnosis (N=578), median (IQR)	47 (33-56)
Time (months) between first symptom and diagnosis (N=371), median (IQR)	0.0 (0.0-12.0)
Time (months) between first symptom and joining PLM (N=378), median (IQR)	63.9 (28.9-9.3)
Time (months) between diagnosis and joining PLM (N=578), median (IQR)	40.9 (14.5-102.3)
Medication (N=1369), n (%)[f]	
Diabetes medication	702 (51.28)
Metformin[g]	481 (68.52)
Hypertension medication	311 (22.72)
Angiotensin-converting enzyme inhibitor[g]	212 (68.17)
Antidepressant medication	301 (21.99)
Serotonin–norepinephrine reuptake inhibitors[g]	150 (49.83)
Lipid-modifying medication	267 (19.50)
Statin[g]	247 (92.51)
Immunosuppressants	147 (10.74)
Calcineurin inhibitors[g]	30 (20.41)
Anti-inflammatory and antirheumatic medication	52 (3.80)
Nonsteroidal anti-inflammatory drugs[g]	48 (92.31)

[a]PLM: PatientsLikeMe.

[b]IQR: interquartile range.

XSL•FO
RenderX

[c]Includes American Indian or Alaska Native, or Native Hawaiian or other Pacific Islander.

[d]GED: General Educational Diploma.

[e]CKD: chronic kidney disease.

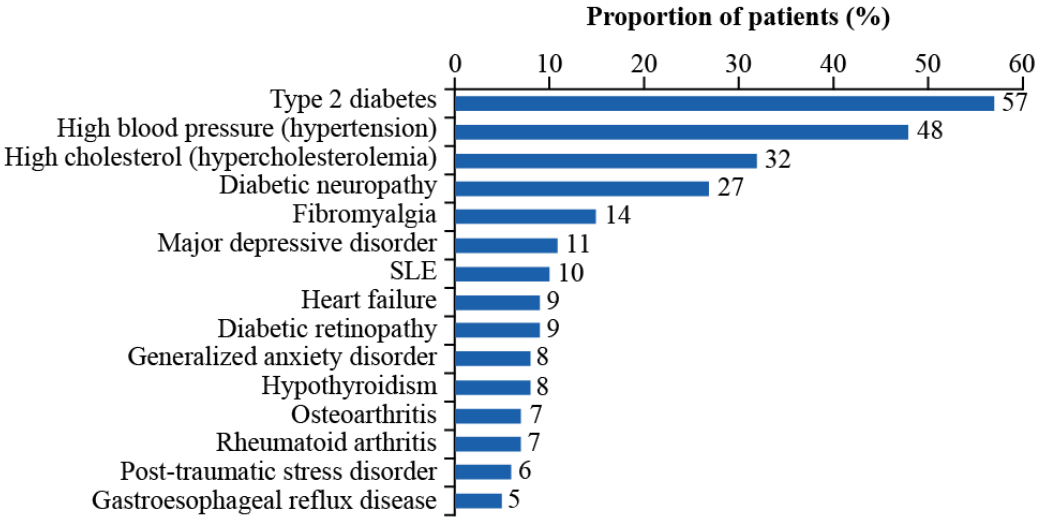
[f]Percent for medication class is percent of those reporting treatment (N=1369), whereas percent for top reported medication within treatment class is percent of treatment class.

[g]Top reported medication within the specified drug class.

Comorbidities and Treatments

Conditions in addition to CKD were reported in 1374/1848 patients (74.35%), and the median number of different conditions per patient was 3 (IQR 0-4). There were 474 patients reporting no comorbidities, 685 patients reporting 1-3 comorbidities, and 689 patients reporting more than 3 comorbidities. The most common conditions among those reporting comorbidities were type 2 diabetes (783/1374, 56.99%), hypertension (664/1374, 48.33%), hypercholesterolemia (439/1374, 31.95%), and diabetic neuropathy (376/1374, 27.37%; Figure 1). Treatments were reported by 1369 patients; the most common were for diabetes (702/1369, 51.28%) and hypertension (311/1369, 22.72%; Table 1). These comorbidities are highly prevalent in patients with CKD.

Figure 1. Patient-reported comorbidities (retrospective data). Comorbidities reported in addition to CKD (n=1374). CKD: chronic kidney disease; SLE: systemic lupus erythematosus.

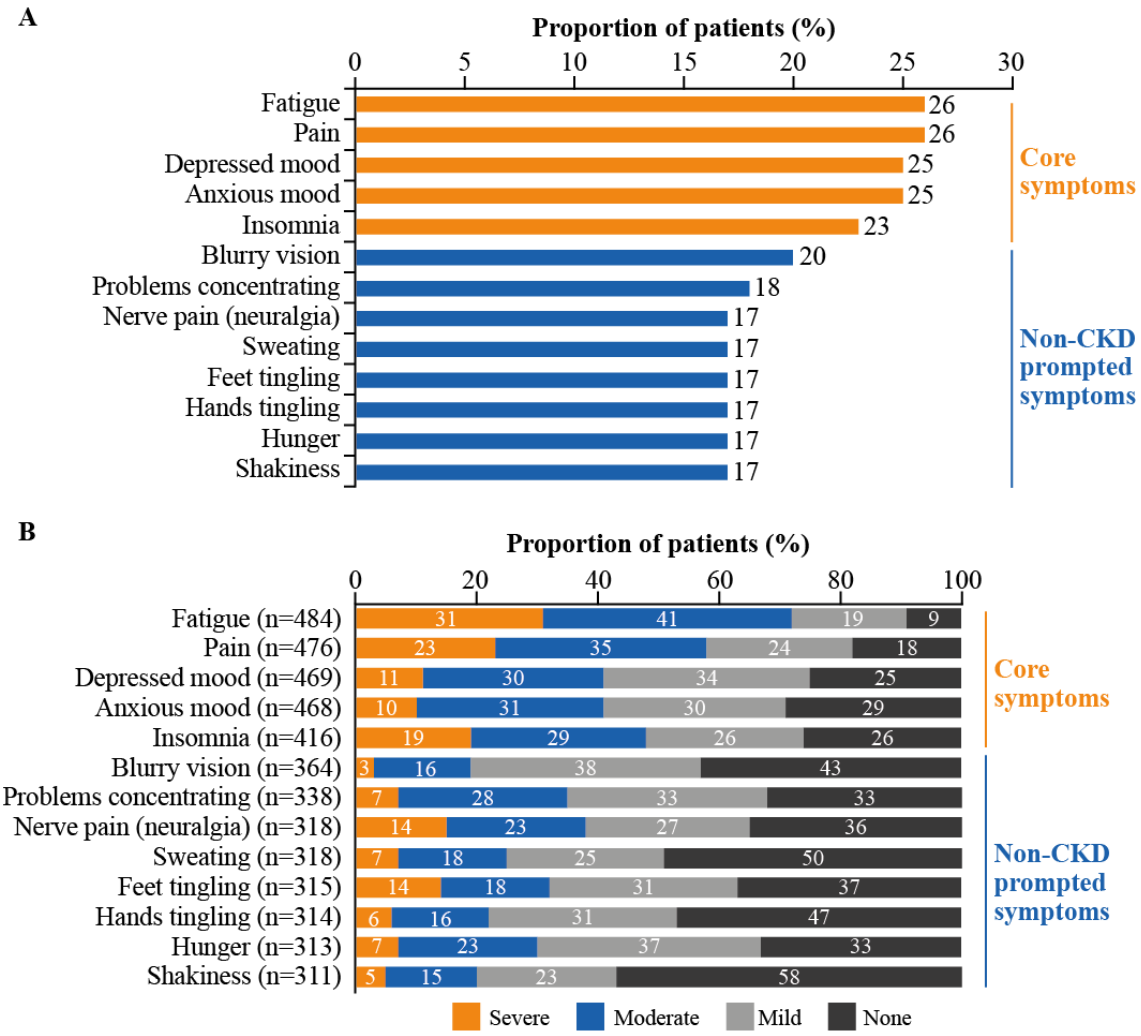


Symptoms

Less than one-half of all patients (754/1848, 40.80%) entered any symptom within 30 days of registration. Core symptoms from the PLM panel were reported by 487 patients (26.35%), and 633 patients (34.25%) reported additional symptoms (Figure 2A). Core symptoms of fatigue, pain, and insomnia were rated as moderate or severe by 71.7% (347/484), 58.4% (278/476), and 47.6% (198/416) of patients, respectively (Figure 2B). Of the additional symptoms, nerve pain, problems concentrating, and feet tingling were classified as moderate or severe by more than 30% of patients—37.1% (118/318), 34.6% (117/338), and 32.4% (102/315), respectively.

XSL•FO
RenderX

Figure 2. Patient-reported symptoms (retrospective data). (A) Core and additional patient-reported symptoms (total cohort, N=1848). (B) Severity of patient-reported symptoms. Core symptoms are symptoms that all patients registered with PatientsLikeMe are asked to track; additional symptoms are those patients themselves may choose to track, or symptoms patients may be asked about in relation to a condition other than CKD. *None* refers to those patients reporting *none* for a specific symptom; those who did not provide any response were not included. CKD: chronic kidney disease.



Qualitative Semistructured Interviews

Concept saturation (Multimedia Appendix 3) was reached with 18 patients, after which no additional patients were interviewed. In-depth patient perspectives on their disease and care were gathered during these semistructured interviews and 4 major concepts were explored: symptoms, financial burden, valued outcome, and care concepts. We explore the interview responses within these themes below.

Patient History With CKD, Including Diagnosis

This group of 18 patients had a median age of 60 (IQR 52-76) years, were mostly US residents (n=16), and female (n=13). At the time of the interview, 2 patients who had participated in the retrospective study had stage 4 CKD and were receiving or about to initiate dialysis. Most patients described their current health status as *OK* (n=7) or *Good* (n=8). The median patient age at diagnosis of CKD was 55 (IQR 47-62) years and 15 patients were able to report disease stage at diagnosis (Table 2). Most patients explained that they were diagnosed by a nephrologist following referral by their primary care physician (Textbox 1), and their self-reported diagnosis was based on an abnormal metabolic panel or urinary protein on screening tests.

All patients had comorbidities, with the most common being hypertension (12/18, 67%; Table 2). These comorbidities also had symptoms that could be associated with CKD, especially fatigue, numbness/tingling, or nausea.

XSL•FO
RenderX

Table 2. Patient demographics (qualitative study).

Characteristic	Total (N=18)
Age (years), median (interquartile range)	60 (52-76)
Female, n (%)	13 (72)
Location, n (%)	
United States	16 (89)
Ex-United States	2 (11)
Work status, n (%)	
Full-time	1 (6)
Part-time	3 (17)
Retired	4 (22)
Disabled[a]	9 (50)
Homemaker	1 (6)
Chronic kidney disease stage at diagnosis, n (%)	
Unknown	3 (17)
2	2 (11)
3a/3b	11 (61)
4	2 (11)
Most frequently reported comorbidities, n (%)	
Hypertension[b]	12 (67)
Osteoarthritis/degenerative joint disease related	9 (50)
Gastroesophageal reflux disease and related conditions	7 (39)
Anemia[b]	7 (39)
Fibromyalgia/myofascial pain	6 (33)
Peripheral neuropathy	6 (33)
Diabetes[b]	4 (22)
Hyperkalemia[b]	3 (17)

[a]Disability not due to chronic kidney disease.

[b]These comorbidities were probed for if not spontaneously described by patients.

Textbox 1. Direct patient quotes from the qualitative study. Patients provided consent to include their words in a deidentified format. CKD: chronic kidney disease.

Patient history with CKD, including how they were diagnosed

At one time I was on so many medications that it was ridiculous, and I've got off of everything except some Parkinson's medicine and blood pressure medicine and blood thinner that I have to take.

I discovered I had chronic kidney disease when I was participating in the dry eye study. During the study I was asked to provide a urine sample, which showed protein in my urine.

And she says, and another thing, she says, we're going to have to send you a specialist on is that you're showing signs of kidney disease, and she says we're looking at stage 3.

I had just had some routine lab work, and it showed that my eGFR was low consistently. He repeated it a few times. So they sent me to a nephrologist.

I never am really sure what's contributing to it, if the kidney disease is a part of that, especially my hemoglobin. I don't get anemic or anything how that fatigue really sets in. I usually attributed it to the fibromyalgia.

Patient symptom experience related to CKD

I could spit sand.

XSL•FO
RenderX

I'm doing a workup for a kidney transplant also, so that's why I had the day off today, and then just like this last month I've had like six different doctors' appointments.

I did have very severe anemia. I had many, many blood transfusions.

Financial impact of CKD

Well, I don't have the best insurance in the world, but so far I've been able to pay most of my bills.

When I was finally approved for disability, I was also approved for (inaudible) Medicare. So that helped a bunch. Because basically, I had to stop working in 2014 because of the health issues.

I'd say most of my problems are financial right now. So now it's taking the time off plus the drugs plus – it's just the whole situation and everything.

It can run up to $700 a month with MRIs, CT scans, doctor bills. It's been bad lately.

CKD patient experience with diet and physical exercise

Yeah, there are times when I can't really go out or do anything even at home, because my muscles are so cramped up.

Losing weight, trying to exercise regularly.

Not really happy [about my diet] because I grew up on a farm and meat was the main – we had meat three times a day, and I'm a real meat eater. So that has been a huge change for me.

And they keep always saying you got – you have to exercise. The extra weight is hard on the kidneys. And I don't know what my problem is. It's a lack of motivation, I think.

He always talks to me about doing what I can. Don't overdo it, obviously ... Yeah, he tries to keep me motivated and at least doing some sort of exercise daily, even if just like for 10 or 15 minutes, whatever, to keep my body moving.

Patient perception of CKD care and management

... They didn't honestly seem very concerned about my kidney health as much.

But I was always having to remind everybody, hey, I've got a kidney problem. We have to keep that in mind that it's not always the pulmonary hypertension. It's not always the small fiber neuropathy. So, I have to make sure that they understand that.

I want to have the best life that I can, and if I can't take anything for pain because of my kidneys, my life isn't going to be the best life I have. So, I have to know what things will work for me but aren't going to continue damaging my kidneys.

He's just very understanding, very supportive, and he does a good job of explaining what's going on. He gives good praise for achieving results, which helps.

I think at this point it's me. That'll change, probably. But I'm pretty independent, and I want to be in control. I think most people do.

... so, I have learned that I don't ask him my questions. I ask my nephrologist my questions. And he will answer anything. And he wants to answer. But the primary doctor just doesn't seem to care.

For the record, I would never have another transplant – and God hope my kidneys don't go south, because I would never do it again. Never in a million years.

Shifting to a new one is a little bit scary.

You really need to have a doctor that's going to really listen to you because he's going to miss something if he's not listening to you.

Well, I just hope that I can live the rest of my life and not have serious problems with my kidney.

I just make sure that I speak up for myself and explain where I'm at or ask questions. I think that that's a key component that patients don't always do. If nothing else, nursing has taught me to be an advocate for myself.

Someone's taking care of it, and I know that I will know if my kidney function starts to fall or something, they'll know, and they'll start to get it taken care of right away.

It's just that nagging thought back there about well, will I be able to continue to live the quality of life that I do, and my kidney function remain high enough that I don't need more aggressive treatment?

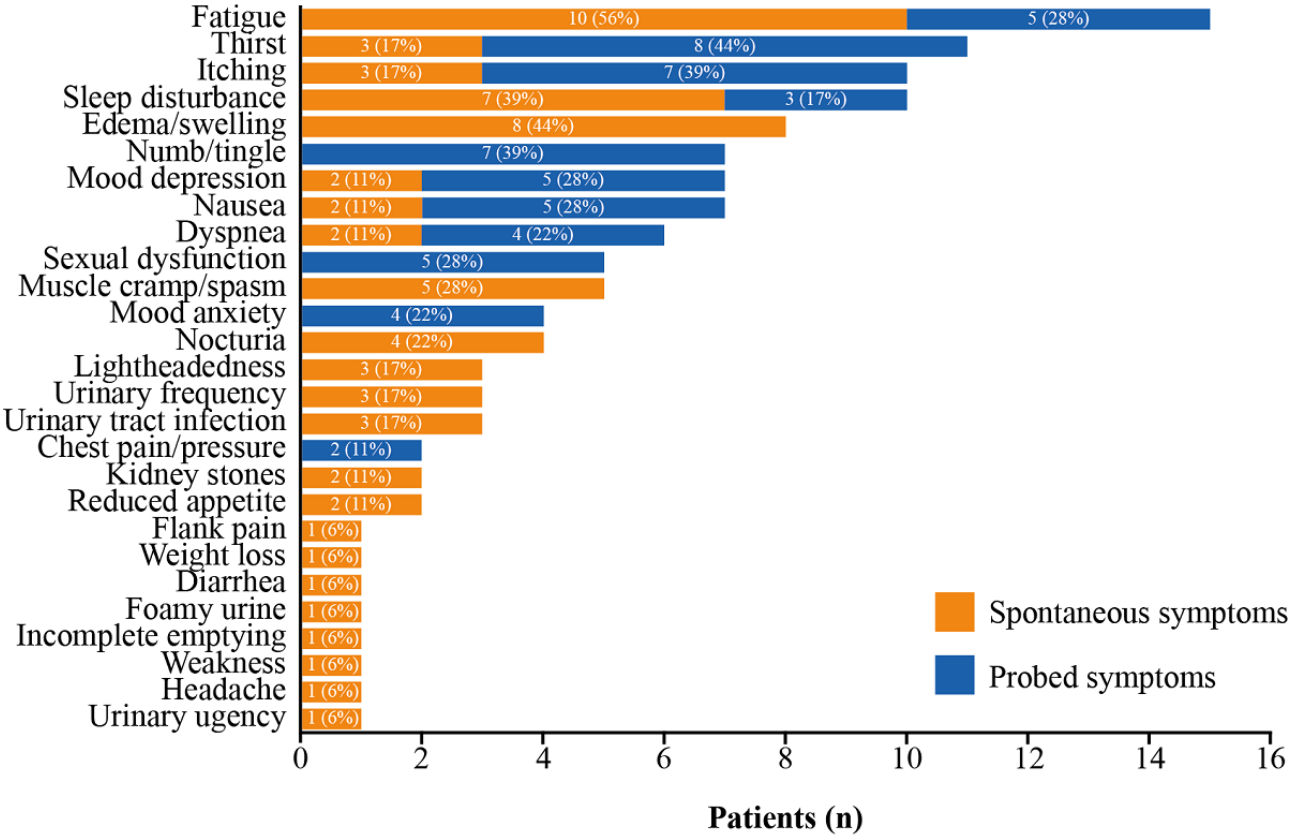
But the doctors that I have here I feel are great doctors, and they really work with me ... So, they all have to be working in concert, which I've worked really hard to get them to do.

XSL•FO
RenderX

Patient Symptom Experience and HRQoL

In total, 28 different symptoms were reported by patients as potentially related to CKD (eg, edema, urinary tract infection, and urinary frequency). Textbox 1 summarizes key quotes from patients. The most commonly reported symptoms (spontaneous and probed) included fatigue (n=15), thirst (n=11), itching (n=10), sleep disturbance (n=10), and edema/swelling (n=8) (Figure 3). Edema/swelling and muscle cramps/spasms were 2 commonly reported symptoms not probed for during interviews. These symptoms also influenced both social interactions and the activities of daily living. For example, 1 patient noted that their body would suddenly *go rubbery* on them during yard work (Textbox 1).

Figure 3. Patient-reported symptoms (qualitative study, N=18). Spontaneous symptoms are symptoms that patients themselves choose to share during the interview that they considered to be related to CKD, probed symptoms are those that interviewers probed for if patients did not mention them (fatigue, sleep disturbance, itching, numbness/tingling, chest pains, dyspnea, nausea, thirst, mood, and sexual function). If a patient had not mentioned any issues related to sleep spontaneously, they were asked about "any impacts on their sleep"; similarly, if they had not mentioned any issues related to anxiety or depression, they were asked "how does CKD affect your mood?". Uniformly, patients reported negative impacts on sleep, mood, and sexual functioning either spontaneously or prompted.



When asked which symptom had the biggest impact on their HRQoL, most patients (n=12) mentioned a symptom not related to CKD or said no symptoms had a significant impact. Of the CKD-related symptoms mentioned as having the biggest impact, fatigue was most commonly reported (n=4), followed by muscle cramping (n=2). Six patients reported experiencing fatigue *all of the time*, whereas 2 patients commented that CKD affected their sleep *all of the time* (Multimedia Appendix 4), and the impact of lack of sleep on fatigue was noted.

Only the 2 patients with end-stage renal disease (ie, stage 4 or 5, on or preparing for dialysis) stated that CKD was the single biggest health burden in their lives. One patient commented that bipolar disorder and CKD equally caused the most burden; another said diabetes and CKD. Most of the remaining patients ranked CKD third or lower in their list of comorbidities.

Financial Impact on Patients

Nine patients reported a financial impact, with 4 patients spending *some* and 3 patients spending *a lot* of their monthly budget toward medical bills; 2 patients noted that only *a little* was related to medical bills. The level of financial impact varied across patients and was influenced by insurance status, severity of CKD, and comorbidities. The financial impact of CKD was highlighted by 12 patients in total (Multimedia Appendix 4). Concepts appearing when describing financial impact of CKD included concern about loss of insurance, high deductibles, insurance dictating care, losing time at work leading to lost wages, expensive drugs, cost of specialists, and time and knowledge required to scrutinize billing (Textbox 1).

Patient Experience With Diet and Physical Exercise

Physical activity was considered important; however, nothing more than regular exercise or walking for CKD was prescribed (Textbox 1). Some patients with comorbidities did report they received physical therapy, and it was clear that physicians were actively encouraging patients to engage in physical activity.

A total of 7 patients shared that they had received recommendations for physical activity from clinicians, but all

XSL•FO
RenderX

instances were very general and not specific for CKD. Patients received more information about diet monitoring than about physical activity, because they discussed diet details when reviewing their course of treatment for CKD. Almost all patients mentioned monitoring aspects of their diet, such as salt, potassium, protein, and water intake.

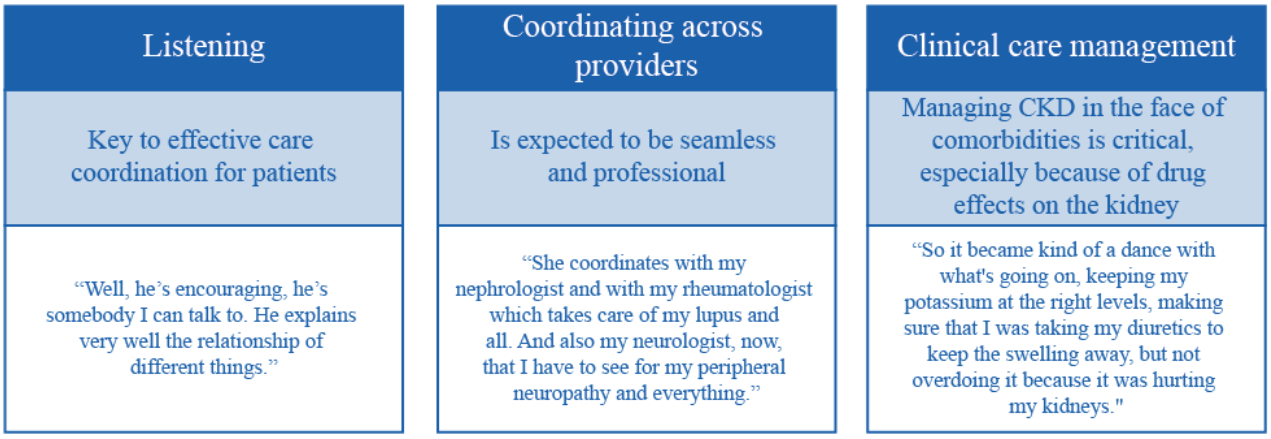
Many patients noted that they missed foods they like, such as pasta, but accepted diet changes. However, it was clear that for some patients, balancing diet changes with multiple conditions was challenging. Vitamin and mineral supplement use were reported, specifically when iron-deficiency anemia or hypokalemia were mentioned.

Disease Management

When asked what they valued in managing their CKD, most patients focused on the reduction of future harm (ie, minimizing kidney damage, preventing disease progression, and not losing kidney function; Textbox 1). For example, 1 patient recognized the importance of preserving kidney function: "I want to have the best life that I can, and if I can't take anything for pain because of my kidneys, my life isn't going to be the best life I have. So, I have to know what things will work for me but aren't going to continue damaging my kidneys." Another patient viewed their HRQoL as the most important aspect of care: "Quality of life is more important to me than length of life." Other concepts valued by patients included awareness of CKD by all their doctors, a firm diagnosis, regular monitoring, having future treatment choices (dialysis, transplant, or none) honored, preferring transplant to dialysis, and having a normal life (Textbox 1). Importantly, one patient summarized that "I want to be able to make my own choices, and if I decide that I go on dialysis, I want support for that. If I don't want to, I want support for that, too".

Patients with CKD require coordination of care with their primary care physician, renal specialists, and specialists managing their comorbidities. Most patients listed their primary care physician, psychiatrist, and nephrologist as the most helpful in managing their CKD; 4 patients said they themselves were most helpful (Textbox 1). Patients frequently commented on the aspects of their care that led to well-coordinated, effective CKD management. Concepts generated by patients fell into 3 overarching domains: listening, coordinating across providers, and clinical care management (Figure 4). Listening to patients' needs was considered central to effectively coordinating care. There was an expectation among patients that coordinated care across providers should be seamless to ensure appropriate management of their CKD in the face of comorbidities. When any of these aspects of care were absent, patients felt these were unmet needs (Textbox 1).

Figure 4. Concepts related to optimal CKD care and management (qualitative study). CKD: chronic kidney disease.



Discussion

Principal Results

The patient voice is often underrepresented in the clinical setting, yet it is critical to optimizing and improving clinical practice and research, and to better understand chronic conditions such as CKD. Our main objective in this study was to better understand the patient experience with CKD and its management through direct patient insight, unfiltered by interactions with health care professionals, from a large patient cohort within the PLM network. This study highlighted the array of symptoms that patients with CKD face, as well as key concepts that are important for optimal care and management.

To our knowledge, this is the first published analysis of real-world data gathered on CKD from an online community. Data from PLM have previously been used to explore patient perspectives on neuromyelitis optica and neuromyelitis optica spectrum disorders [19], characterize the profiles of patients with atopic dermatitis [22] and systemic lupus erythematosus [23], and understand patient preferences for type 2 diabetes self-management [24]. This is the first study to utilize the PLM network to gain insights into self-reported experiences in patients with CKD. These approaches for data collection can also translate to the clinical trial design process. Engaging patients early in the process has the potential to ultimately yield more successful studies [25]. DISCOVER CKD is an example of a study that has been influenced by this patient-centric approach (ClinicalTrials.gov identifier: NCT04034992). This PLM study was the pilot study for developing the interviews used in the DISCOVER CKD study and guided decisions about which symptoms the DISCOVER CKD study should capture [26].

XSL•FO
RenderX

Retrospective Study

The retrospective study profiled a specific cohort of PLM members with CKD who were not receiving dialysis. Their characteristics are broadly consistent with US CKD populations from United States Renal Data System (USRDS), National Health and Nutrition Examination Survey (NHANES), and the Dialysis Outcomes and Practice Patterns Study (DOPPS), although the percentage of females is slightly higher [27-29]. Members of PLM with CKD included in this cohort were predominantly white, middle-aged (median 56 years), and women, whereas the prevalence of CKD is only slightly higher for women compared with men (17% vs 13%, respectively) in the general population, with the majority aged >60 [27]. PLM members also have a higher educational level than the general population with CKD; 79.1% (564/713) had postsecondary education compared with 56%-59% in the general population [27,28]. The overrepresentation of educated men and women in the cohort and the relatively young age of PLM members compared with the general population according to the references reflect the patient population who regularly use health-based internet sites [30,31].

Comorbidities were reported by 74.35% (1374/1848) of the PLM cohort studied, the most common being type 2 diabetes (783/1374, 56.99%), hypertension (664/1374, 48.33%), hypercholesterolemia (439/1374, 31.95%), and diabetic neuropathy (376/1374, 27.37%). This is reflective of the general CKD population, as reported by the USRDS, where diabetes (36%), hypertension (31%), and self-reported cardiovascular disease (40%) are the most common comorbidities [27]. While rates of major depressive disorder (148/1374, 10.77%) were close to the frequency of depressive symptoms among predialysis patients (13%) attending a single renal center [32], they were lower than the reported prevalence of depressive symptoms among patients with CKD who are not on dialysis (21%-42%) [33]. The proportion of patients with fibromyalgia (199/1374, 14.48%) was higher than that previously reported in patients with CKD receiving hemodialysis (3.9%-12.2%) [34-37] or peritoneal dialysis (9.7%-18%) [38,39]. This may be due to the nature of the PLM database, where large communities for chronic conditions including fibromyalgia exist [17]. Therefore, these comorbidities tend to be overrepresented. Overall, the similarities between the characteristics of the PLM network and the overall US CKD population indicate that the PLM platform is a useful format for gathering insights into CKD care and management.

Qualitative Study

Information gathered from semistructured interviews with a small subset of the retrospective study population provides a unique in-depth perspective on patient experiences and captures their opinions regarding the care they receive or gaps they perceive in their care, including insights that may not be collected in a routine clinical setting or captured by other patient-reported outcome tools. Four key concepts were explored in the interviews: symptoms, financial burden, valued outcome, and care concepts.

Patients shared how their symptoms affected their HRQoL, with fatigue noted by most patients as having the biggest impact. Fatigue and insomnia were also among the top core symptoms reported in the retrospective population. Together, these findings highlight the significance of fatigue and sleep-related problems for patients, corroborating other CKD studies [40-43]. Patients with CKD are known to experience profound fatigue and it is one of the most common symptoms affecting their HRQoL [43,44]. One study reported how fatigue was associated with unemployment, comorbidities, and use of antidepressant medication; the authors suggested that the presence of fatigue may act as a clinical prognostic factor in CKD [45]. Patients experiencing sleep disorders and fatigue are also at increased risk of all-cause mortality [44]. Fatigue has also been related to poor physical functioning, although causality remains unclear [41].

The cause of fatigue experienced by this subset of PLM patients is likely to be multifactorial [41] and associated with prevalent complications such as anemia [41,45] and low albumin levels [42], psychosocial factors including depression and anxiety [41], and insomnia [46]. Improved understanding of the impact that fatigue and other symptoms such as depression and anxiety have on those living with CKD can help guide treatment and improve quality of care by prioritizing interventions that improve HRQoL, such as personalized care, precision medicine, or wearable technology and biometric devices.

Only the 2 patients with stage 4 CKD considered CKD to be their single biggest health burden. This finding is likely due to the predominance of earlier and less symptomatic stages of CKD (3a or 3b) in this study and to the nonspecific nature of CKD-related symptoms. Other patients identified other comorbidities (eg, diabetes) as more burdensome relative to CKD. These comorbidities are interrelated with CKD and one can affect another; therefore, awareness among patients regarding which symptoms are associated with specific diseases needs to be addressed in the clinical management of patients with CKD.

Financial concepts around loss of insurance and earnings, and the cost of treatment and specialist care arose from interviews. Importantly, patients' experiences of the financial impact of CKD were varied and influenced by insurance status (particularly in regions without universal health care), severity of CKD, and other comorbidities. According to the USRDS, total medical costs were US $50.4 billion for CKD (excluding end-stage renal disease) among Medicare patients in 2013 [47]. In adult patients in the United States, direct annual health care costs for individuals with CKD were US $17,472 higher compared with those without CKD from 2002 to 2011 [48]. Such financial pressures experienced by patients may impact other symptoms noted, such as depression [49], and should therefore be considered in the holistic management of patients in clinical care.

The personal experiences of some interviewees reflected but were not fully aligned with the current Kidney Disease: Improving Global Outcomes (KDIGO) guidelines [50,51]. Current guidelines recommend that those living with CKD should be encouraged to undertake physical exercise, achieve a healthy weight (BMI 20-25 kg/m^2), and stop smoking. KDIGO guidelines state that expert dietary advice and information

XSL•FO
RenderX

should be provided to patients in the context of an education program and tailored to severity of CKD. Detailed recommendations on protein intake, glycemic control, and salt and potassium intake are also provided [50,52]. However, PLM patients in this study mentioned that any lifestyle advice they received was very general and they were more likely to receive information about diet than physical activity; almost all patients noted that they monitored aspects of their diet. Importantly, some patients found balancing diet changes with multiple conditions very challenging. These insights illustrate that KDIGO guidelines on exercise may not be implemented effectively in daily practice and patients with multiple conditions need additional support from health care providers to manage the changes in their diet.

Concepts identified by patients relating to optimal CKD care and management fell into 3 domains: listening, coordinating across providers, and clinical care management. Patients valued physicians who listened to their needs, coordinated their care effectively with other specialists, and managed their CKD in conjunction with comorbidities. When any of these aspects of care were absent, patients felt these were unmet needs. These differ slightly to The Renal Association and Kidney Care UK's Kidney Patient Reported Experience Measure, a UK initiative to gain insight into patient experiences and help shape renal services, which identified 3 aspects of renal care that affect patient experience: shared decision making, transport, and discomfort associated with needle insertions [53]. Such differences may reflect differences in health care systems between countries, such as the National Health Service in the UK versus an insurance-based system in the United States. These insights provide an opportunity for health care service providers to adapt local clinical practice to improve care and better reflect the experiences and needs of patients. Patient experience has been positively associated with patient safety and health outcomes, and clinical effectiveness across a number of diseases and settings [54].

Study Limitations

A limitation of this study was that members of PLM self-reported their diagnoses and treatment—there was no independent clinical corroboration; and treatment indication, although captured by PLM, was not analyzed as few patients provide this information. In addition, the extent and nature of missing data can reduce precision and lead to a systematic bias of results; however, missing variables and missingness in variables were assessed and reported. Because PLM members were not expected to know details of their disease severity (ie, CKD stage), it is also not possible to derive information on how self-management and care needs to change with disease progression.

There is also a sampling bias as patients had an active choice to participate/enroll into the PLM network, so the cohort was self-selecting. As a result, generalizability of these results may be limited because the profiles of patients enrolling into the PLM network may differ from the profile of those not enrolling into PLM or of the general population. The observation that patients who enroll in PLM with CKD tend to be younger, more likely to have some college education, and more likely to have white ethnicity compared with the general population has also been noted in other disease groups [19,55].

The evaluation of the retrospective data in this study focused on information reported within 30 days of enrollment/registration (or as close as possible) into the PLM network. However, we cannot be sure all information reported during or outside this period relates specifically to CKD, as most patients reported multiple comorbidities. There may be a future opportunity to conduct a longitudinal study to describe patient experience over time.

Nonetheless, in the retrospective study, the use of data directly from an activated network of patients that were not filtered through an interviewer or clinician eliminated any potential interviewer bias and the use of an anonymous online survey may have improved willingness to disclose sensitive information.

Conclusions

This study provides a unique source of real-world information on the patient experience of CKD and its management by utilizing the PLM network. The data collected through PLM identified key symptoms affecting patients, including fatigue and depression/anxiety, and are broadly consistent with published data on CKD populations in the United States. In-depth semistructured interviews with members reveal the challenges patients face living with symptoms of CKD and the impacts they experience financially and on HRQoL. These interviews identified key concepts around optimal care and management of CKD such as listening to patients and coordinating health care across providers, indicating that this approach can provide insights outside of the clinical setting to inform clinical practice and drug development, and improve future studies.

Acknowledgments

The authors thank the patients, the patients' families, and the PLM interviewers involved in this study, without whom this research would not be possible. Medical writing support was provided by Georgina Collett, PhD, CMPP, and Laura Geuss, PhD, and editorial support was provided by Bethany King, BSc, all of Core Medica, London, UK, supported by AstraZeneca according to Good Publication Practice guidelines. The sponsor was involved in the study design, collection, analysis, and interpretation of data, as well as data checking of information provided in the manuscript. Ultimate responsibility for opinions, conclusions, and data interpretation lies with the authors. Development of this manuscript was supported by AstraZeneca.

XSL•FO
RenderX

Authors' Contributions

GJ, EN, KH, CE, and JH contributed to the design of the study. MF-R contributed to collection of data for the qualitative study. EN and MF-R conducted the data analysis and all authors contributed to data interpretation and manuscript preparation and review. All authors approved the final version of the manuscript and accept accountability for the overall work.

Conflicts of Interest

AMA, CMQ, EN, and JM are employees of AstraZeneca. AN, CE, ETW, GJ, JH, and KH are employees and stockholders of AstraZeneca. MF-R is an employee of PLM.

Multimedia Appendix 1
Supplementary material - Methods.
[DOCX File , 14 KB - jmir_v22i7e18548_app1.docx]

Multimedia Appendix 2
Patient disposition (retrospective data).
[DOCX File , 107 KB - jmir_v22i7e18548_app2.docx]

Multimedia Appendix 3
Concept saturation (qualitative study).
[DOCX File , 92 KB - jmir_v22i7e18548_app3.docx]

Multimedia Appendix 4
Patient-reported symptoms (qualitative study, N=18).
[DOCX File , 98 KB - jmir_v22i7e18548_app4.docx]

References

1. Bello AK, Levin A, Tonelli M, Okpechi IG, Feehally J, Harris D, et al. Assessment of Global Kidney Health Care Status. JAMA 2017 May 09;317(18):1864-1881. [doi: 10.1001/jama.2017.4046] [Medline: 28430830]
2. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. Lancet 2013 Jul 20;382(9888):260-272. [doi: 10.1016/S0140-6736(13)60687-X] [Medline: 23727169]
3. Rees S, Williams A. Promoting and supporting self-management for adults living in the community with physical chronic illness: A systematic review of the effectiveness and meaningfulness of the patient-practitioner encounter. JBI Libr Syst Rev 2009;7(13):492-582. [doi: 10.11124/01938924-200907130-00001] [Medline: 27819974]
4. Donald M, Beanlands H, Straus S, Ronksley P, Tam-Tham H, Finlay J, et al. Preferences for a self-management e-health tool for patients with chronic kidney disease: results of a patient-oriented consensus workshop. CMAJ Open 2019 Dec;7(4):E713-E720 [FREE Full text] [doi: 10.9778/cmajo.20190081] [Medline: 31822502]
5. Baay S, Hemmelgarn B, Tam-Tham H, Finlay J, Elliott MJ, Straus S, et al. Understanding Adults With Chronic Kidney Disease and Their Caregivers' Self-Management Experiences: A Qualitative Study Using the Theoretical Domains Framework. Can J Kidney Health Dis 2019;6:2054358119848126 [FREE Full text] [doi: 10.1177/2054358119848126] [Medline: 31205731]
6. Levin A, Adams E, Barrett BJ, Beanlands H, Burns KD, Chiu HH, et al. Canadians Seeking Solutions and Innovations to Overcome Chronic Kidney Disease (Can-SOLVE CKD): Form and Function. Can J Kidney Health Dis 2018;5:2054358117749530 [FREE Full text] [doi: 10.1177/2054358117749530] [Medline: 29372064]
7. Havas K, Douglas C, Bonner A. Meeting patients where they are: improving outcomes in early chronic kidney disease with tailored self-management support (the CKD-SMS study). BMC Nephrol 2018 Oct 20;19(1):279 [FREE Full text] [doi: 10.1186/s12882-018-1075-2] [Medline: 30342487]
8. Breckenridge K, Bekker HL, Gibbons E, van der Veer SN, Abbott D, Briançon S, et al. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. Nephrol Dial Transplant 2015 Oct;30(10):1605-1614 [FREE Full text] [doi: 10.1093/ndt/gfv209] [Medline: 25982327]
9. Clarke AL, Yates T, Smith AC, Chilcot J. Patient's perceptions of chronic kidney disease and their association with psychosocial and clinical outcomes: a narrative review. Clin Kidney J 2016 Jun;9(3):494-502 [FREE Full text] [doi: 10.1093/ckj/sfw014] [Medline: 27274839]
10. Teasdale EJ, Leydon G, Fraser S, Roderick P, Taal MW, Tonkin-Crine S. Patients' Experiences After CKD Diagnosis: A Meta-ethnographic Study and Systematic Review. Am J Kidney Dis 2017 Nov;70(5):656-665. [doi: 10.1053/j.ajkd.2017.05.019] [Medline: 28764920]
11. Food and Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims 2009 October 22, 2019. Food and Drug Administration. URL: https://www.fda.gov/regulatory-information/

XSL•FO
RenderX

search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims [accessed 2019-10-22]
12. Anderson A, Benger J, Getz K. Using Patient Advisory Boards to Solicit Input Into Clinical Trial Design and Execution. Clin Ther 2019 Aug;41(8):1408-1413. [doi: 10.1016/j.clinthera.2019.06.006] [Medline: 31307831]
13. Muhammad S, Allan M, Ali F, Bonacina M, Adams M. The renal patient support group: supporting patients with chronic kidney disease through social media. J Ren Care 2014 Sep;40(3):216-218. [doi: 10.1111/jorc.12076] [Medline: 25065504]
14. Hilliard ME, Sparling KM, Hitchcock J, Oser TK, Hood KK. The emerging diabetes online community. Curr Diabetes Rev 2015;11(4):261-272 [FREE Full text] [Medline: 25901500]
15. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
16. Nelakurthi AR, Pinto AM, Cook CB, Jones L, Boyle M, Ye J, et al. Should patients with diabetes be encouraged to integrate social media into their care plan? Future Sci OA 2018 Jul;4(7):FSO323 [FREE Full text] [doi: 10.4155/fsoa-2018-0021] [Medline: 30112191]
17. Wicks P, Mack Thorley E, Simacek K, Curran C, Emmas C. Scaling PatientsLikeMe via a "Generalized Platform" for Members with Chronic Illness: Web-Based Survey Study of Benefits Arising. J Med Internet Res 2018 May 07;20(5):e175 [FREE Full text] [doi: 10.2196/jmir.9909] [Medline: 29735472]
18. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
19. Eaneff S, Wang V, Hanger M, Levy M, Mealy MA, Brandt AU, et al. Patient perspectives on neuromyelitis optica spectrum disorders: Data from the PatientsLikeMe online community. Mult Scler Relat Disord 2017 Oct;17:116-122 [FREE Full text] [doi: 10.1016/j.msard.2017.07.014] [Medline: 29055439]
20. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
21. Brédart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. Health Qual Life Outcomes 2014;12(1):15. [doi: 10.1186/1477-7525-12-15]
22. Rundle C, Dellavalle R. PatientsLikeMe and atopic dermatitis: characterizing the atopic dermatitis patient profile. Dermatol Online J 2018 Aug 15;24(8):13030/qt70k8c9jn [FREE Full text] [Medline: 30677844]
23. Nyman E, Vaughan T, Desta B, Wang X, Barut V, Emmas C. Characteristics and Symptom Severity of Patients Reporting Systemic Lupus Erythematosus in the PatientsLikeMe Online Health Community: A Retrospective Observational Study. Rheumatol Ther 2020 Mar;7(1):201-213 [FREE Full text] [doi: 10.1007/s40744-020-00195-7] [Medline: 32008212]
24. Lopez JMS, Katic BJ, Fitz-Randolph M, Jackson RA, Chow W, Mullins CD. Understanding preferences for type 2 diabetes mellitus self-management support through a patient-centered approach: a 2-phase mixed-methods study. BMC Endocr Disord 2016 Jul 18;16(1):41 [FREE Full text] [doi: 10.1186/s12902-016-0122-x] [Medline: 27430259]
25. DasMahapatra P, Raja P, Gilbert J, Wicks P. Clinical trials from the patient perspective: survey in an online patient community. BMC Health Serv Res 2017 Feb 27;17(1):166 [FREE Full text] [doi: 10.1186/s12913-017-2090-x] [Medline: 28241758]
26. James G, Wei J, Wittbrodt E, Quinn C, Allum A, Hedman K. Utilising Mobile and Web-Based Technology for Capturing Patient-Specific Information: Methods from the DISCOVER CKD Program. 2019 Presented at: American Society of Nephrology; 5-10 November 2019; Washington, DC.
27. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey Data. UNITED STATES RENAL DATA SYSTEM. 2018. URL: https://www.usrds.org/2018/view/Default.aspx [accessed 2019-10-03]
28. Myers OB, Pankratz VS, Norris KC, Vassalotti JA, Unruh ML, Argyropoulos C. Surveillance of CKD epidemiology in the US - a joint analysis of NHANES and KEEP. Sci Rep 2018 Oct 26;8(1):15900 [FREE Full text] [doi: 10.1038/s41598-018-34233-w] [Medline: 30367154]
29. DPM United States. DOPPS Practice Monitor United States. URL: https://www.dopps.org/dpm/DPMSlideBrowser.aspx?type=Topic&id=33
30. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
31. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
32. Amira O. Prevalence of symptoms of depression among patients with chronic kidney disease. Niger J Clin Pract 2011;14(4):460-463 [FREE Full text] [doi: 10.4103/1119-3077.91756] [Medline: 22248950]
33. Shirazian S, Grant CD, Aina O, Mattana J, Khorassani F, Ricardo AC. Depression in Chronic Kidney Disease and End-Stage Renal Disease: Similarities and Differences in Diagnosis, Epidemiology, and Management. Kidney Int Rep 2017 Jan;2(1):94-107 [FREE Full text] [doi: 10.1016/j.ekir.2016.09.005] [Medline: 29318209]

XSL•FO
RenderX

34. Yuceturk TE, Yucel AE, Yuceturk H, Kart-Koseoglu H, Unuvar R, Ozdemir FN, et al. Fibromyalgia: its prevalence in haemodialysis patients and its relationships with clinical and laboratory parameters. Nephrol Dial Transplant 2005 Nov;20(11):2485-2488. [doi: 10.1093/ndt/gfi028] [Medline: 16046505]
35. Couto CI, Natour J, Carvalho AB. Fibromyalgia: its prevalence and impact on the quality of life on a hemodialyzed population. Hemodial Int 2008 Jan;12(1):66-72. [doi: 10.1111/j.1542-4758.2008.00243.x] [Medline: 18271844]
36. Leblebici B, Özelsancak R, Yılmaz EE, Doruk P. Fibromyalgia syndrome in Turkish hemodialysis patients. Hemodial Int 2016 Jan;20(1):106-110. [doi: 10.1111/hdi.12332] [Medline: 26198740]
37. Samimagham H, Haghighi A, Tayebi M, Jenabi A, Arabi M, Kianmehr N. Prevalence of fibromyalgia in hemodialysis patients. Iran J Kidney Dis 2014 May;8(3):236-239 [FREE Full text] [Medline: 24878948]
38. Okumus M, Parpucu H, Kocaoglu S, Ceceli E, Duranay M, Borman P. The Frequency of Fibromyalgia Syndrome and the Quality of Life in Patients with Peritoneal Dialysis. OJRA 2012;02(04):88-93. [doi: 10.4236/ojra.2012.24017]
39. Ozelsancak R, Doruk Analan P, Leblebici B. Factors Associated with Fibromyalgia Syndrome in Peritoneal Dialysis Patients. Physiother Can 2019;71(1):77-81 [FREE Full text] [doi: 10.3138/ptc.2017-45.ep] [Medline: 30787502]
40. Almutary H, Bonner A, Douglas C. Symptom burden in chronic kidney disease: a review of recent literature. J Ren Care 2013 Sep;39(3):140-150. [doi: 10.1111/j.1755-6686.2013.12022.x] [Medline: 23826803]
41. Artom M, Moss-Morris R, Caskey F, Chilcot J. Fatigue in advanced kidney disease. Kidney Int 2014 Sep;86(3):497-505 [FREE Full text] [doi: 10.1038/ki.2014.86] [Medline: 24694985]
42. Jhamb M, Liang K, Yabes J, Steel JL, Dew MA, Shah N, et al. Prevalence and correlates of fatigue in chronic kidney disease and end-stage renal disease: are sleep disorders a key to understanding fatigue? Am J Nephrol 2013;38(6):489-495 [FREE Full text] [doi: 10.1159/000356939] [Medline: 24335380]
43. Joshwa B, Campbell ML. Fatigue in Patients with Chronic Kidney Disease: Evidence and Measures. Nephrol Nurs J 2017;44(4):337-343. [Medline: 29160968]
44. Yang XH, Zhang BL, Gu YH, Zhan XL, Guo LL, Jin HM. Association of sleep disorders, chronic pain, and fatigue with survival in patients with chronic kidney disease: a meta-analysis of clinical trials. Sleep Med 2018 Nov;51:59-65. [doi: 10.1016/j.sleep.2018.06.020] [Medline: 30099353]
45. Gregg LP, Jain N, Carmody T, Minhajuddin AT, Rush AJ, Trivedi MH, et al. Fatigue in Nondialysis Chronic Kidney Disease: Correlates and Association with Kidney Outcomes. Am J Nephrol 2019;50(1):37-47. [doi: 10.1159/000500668] [Medline: 31167183]
46. Lindner AV, Novak M, Bohra M, Mucsi I. Insomnia in Patients With Chronic Kidney Disease. Semin Nephrol 2015 Jul;35(4):359-372. [doi: 10.1016/j.semnephrol.2015.06.007] [Medline: 26355254]
47. Golestaneh L, Alvarez PJ, Reaven NL, Funk SE, McGaughey KJ, Romero A, et al. All-cause costs increase exponentially with increased chronic kidney disease stage. Am J Manag Care 2017 Jun;23(10 Suppl):S163-S172 [FREE Full text] [Medline: 28978205]
48. Ozieh MN, Bishu KG, Dismuke CE, Egede LE. Trends in healthcare expenditure in United States adults with chronic kidney disease: 2002-2011. BMC Health Serv Res 2017 May 22;17(1):368 [FREE Full text] [doi: 10.1186/s12913-017-2303-3] [Medline: 28532412]
49. Song M, Ward SE, Hladik GA, Bridgman JC, Gilet CA. Depressive symptom severity, contributing factors, and self-management among chronic dialysis patients. Hemodial Int 2016 Apr;20(2):286-292 [FREE Full text] [doi: 10.1111/hdi.12317] [Medline: 25998623]
50. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update Work Group. KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl (2011) 2017 Jul;7(1):1-59 [FREE Full text] [doi: 10.1016/j.kisu.2017.04.001] [Medline: 30675420]
51. KDIGO. Chapter 2: Definition, identification, and prediction of CKD progression. Kidney Int Suppl (2011) 2013 Jan;3(1):63-72 [FREE Full text] [doi: 10.1038/kisup.2012.65] [Medline: 25018976]
52. KDIGO. Summary of Recommendation Statements. Kidney Int Suppl (2011) 2013 Jan;3(1):5-14 [FREE Full text] [doi: 10.1038/kisup.2012.77] [Medline: 25598998]
53. Stannard C, Gair R, Hawkins J, Wellsted D. The experience of chronic kidney disease patients: The results of the 2017 and 2018 renal association / kidney care UK kidney patient reported experience measure. Nephrol Dial Transplant 2019;34(Supplement 1):i632-i637. [doi: 10.1093/ndt/gfy104.fp339]
54. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 03;3(1):e001570 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
55. Fox S, Duggan M. Health Online. Pew Research Center Internet & Technology. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-05-18]

Abbreviations

CKD: chronic kidney disease

XSL•FO
RenderX

HRQoL: health-related quality of life
IQR: interquartile range
PLM: PatientsLikeMe

Edited by G Eysenbach; submitted 13.03.20; peer-reviewed by P Ronksley, A Nikolaev, D Carvalho; comments to author 06.04.20; revised version received 01.06.20; accepted 04.06.20; published 15.07.20.

Please cite as:
James G, Nyman E, Fitz-Randolph M, Niklasson A, Hedman K, Hedberg J, Wittbrodt ET, Medin J, Moreno Quinn C, Allum AM, Emmas C
Characteristics, Symptom Severity, and Experiences of Patients Reporting Chronic Kidney Disease in the PatientsLikeMe Online Health Community: Retrospective and Qualitative Study
J Med Internet Res 2020;22(7):e18548
URL: http://www.jmir.org/2020/7/e18548/
doi:10.2196/18548
PMID:32673242

©Glen James, Elisabeth Nyman, Marcy Fitz-Randolph, Anna Niklasson, Katarina Hedman, Jonatan Hedberg, Eric T Wittbrodt, Jennie Medin, Carol Moreno Quinn, Alaster M Allum, Cathy Emmas. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Impact of the Grading and Assessment of Predictive Tools Framework on Clinicians and Health Care Professionals' Decisions in Selecting Clinical Predictive Tools: Randomized Controlled Trial

Mohamed Khalifa[1], MD, MSc; Farah Magrabi[1], BSc, MSc, PhD; Blanca Gallego Luxan[2], BSc, MSc, PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia

[2]Centre for Big Data Research in Health, Faculty of Medicine, University of New South Wales, Sydney, Australia

Corresponding Author:
Mohamed Khalifa, MD, MSc
Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine and Health Sciences
Macquarie University
75 Talavera Rd, North Ryde, NSW 2113
Sydney, 2113
Australia
Phone: 61 438632060
Email: mohamed.khalifa@mq.edu.au

Abstract

Background: While selecting predictive tools for implementation in clinical practice or for recommendation in clinical guidelines, clinicians and health care professionals are challenged with an overwhelming number of tools. Many of these tools have never been implemented or evaluated for comparative effectiveness. To overcome this challenge, the authors developed and validated an evidence-based framework for grading and assessment of predictive tools (the GRASP framework). This framework was based on the critical appraisal of the published evidence on such tools.

Objective: The aim of the study was to examine the impact of using the GRASP framework on clinicians' and health care professionals' decisions in selecting clinical predictive tools.

Methods: A controlled experiment was conducted through a web-based survey. Participants were randomized to either review the derivation publications, such as studies describing the development of the predictive tools, on common traumatic brain injury predictive tools (control group) or to review an evidence-based summary, where each tool had been graded and assessed using the GRASP framework (intervention group). Participants in both groups were asked to select the best tool based on the greatest validation or implementation. A wide group of international clinicians and health care professionals were invited to participate in the survey. Task completion time, rate of correct decisions, rate of objective versus subjective decisions, and level of decisional conflict were measured.

Results: We received a total of 194 valid responses. In comparison with not using GRASP, using the framework significantly increased correct decisions by 64%, from 53.7% to 88.1% (88.1/53.7=1.64; t_{193}=8.53; P<.001); increased objective decision making by 32%, from 62% (3.11/5) to 82% (4.10/5; t_{189}=9.24; P<.001); decreased subjective decision making based on guessing by 20%, from 49% (2.48/5) to 39% (1.98/5; t_{188}=−5.47; P<.001); and decreased prior knowledge or experience by 8%, from 71% (3.55/5) to 65% (3.27/5; t_{187}=−2.99; P=.003). Using GRASP significantly decreased decisional conflict and increased the confidence and satisfaction of participants with their decisions by 11%, from 71% (3.55/5) to 79% (3.96/5; t_{188}=4.27; P<.001), and by 13%, from 70% (3.54/5) to 79% (3.99/5; t_{188}=4.89; P<.001), respectively. Using GRASP decreased the task completion time, on the 90th percentile, by 52%, from 12.4 to 6.4 min (t_{193}=−0.87; P=.38). The average System Usability Scale of the GRASP framework was very good: 72.5% and 88% (108/122) of the participants found the GRASP useful.

Conclusions: Using GRASP has positively supported and significantly improved evidence-based decision making. It has increased the accuracy and efficiency of selecting predictive tools. GRASP is not meant to be prescriptive; it represents a high-level

XSL•FO
RenderX

approach and an effective, evidence-based, and comprehensive yet simple and feasible method to evaluate, compare, and select clinical predictive tools.

(J Med Internet Res 2020;22(7):e15770) doi:10.2196/15770

KEYWORDS

clinical prediction rule; clinical decision rules; evidence-based medicine; evaluation study

Introduction

Background

Clinical decision support (CDS) systems have been discussed to enhance evidence-based practice and support cost-effectiveness [1-10]. On the basis of the three-level classification by Shortliffe, clinical predictive tools, referred to as predictive tools in this paper, belong to the highest CDS level, providing patient-specific recommendations based on clinical scenarios, which usually follow clinical rules and algorithms, a cost-benefit analysis, or clinical pathways [11,12]. Such tools include various applications, ranging from the simplest manual clinical prediction rules to the most sophisticated machine learning algorithms [13,14]. These research-based applications provide diagnostic, prognostic, or therapeutic decision support. They quantify the contributions of relevant patient characteristics to derive the likelihood of diseases, predict their courses and possible outcomes, or support decision making on their management [15,16].

When selecting predictive tools for implementation in clinical practice or for recommendation in clinical guidelines, clinicians and health care professionals, referred to as *professionals* in this paper, involved in decision making are challenged with an overwhelming and ever-growing number of tools. Many of these tools have never been implemented or evaluated for comparative effectiveness [17-19]. By definition, health care professionals include all clinicians who provide direct care to patients, in addition to professionals who work in laboratories, researchers, and public health experts [20]. Professionals usually rely on previous experience, subjective evaluation, or recent exposure to predictive tools in making selection decisions. Objective methods and evidence-based approaches are rarely used in such decisions [21,22]. When developing clinical guidelines, some professionals search the literature for studies that describe the development, implementation, or evaluation of predictive tools. Others look for systematic reviews comparing the tools' performance or development methods. However, there are no available approaches to objectively summarize or interpret such evidence [23,24]. In addition, predictive tool selection decisions are time-consuming as they seek a consensus of subjective expert views [25]. Furthermore, when experts make their decisions subjectively, they face much decisional conflict; they are less confident in the decisions they make and sometimes less satisfied with them [26].

To overcome this major challenge, the authors developed and published a new evidence-based framework for grading and assessment of predictive tools (the GRASP framework) [27]. The authors have also validated and updated the GRASP framework through the feedback of a wide group of international experts [28]. Furthermore, the authors applied the GRASP framework to evaluate and compare 14 pediatric head injury clinical predictive tools. This study is now published [29]. The GRASP framework aims to provide standardized objective information on predictive tools to support the search for and selection of effective tools. On the basis of the critical appraisal of published evidence, GRASP uses 3 dimensions to grade clinical predictive tools: (1) phase of evaluation, (2) level of evidence, and (3) direction of evidence.

Phase of Evaluation

Predictive tools are assigned the letters A, B, or C based on the highest phase of evaluation: before implementation, during planning for implementation, or after implementation respectively. If a tool's predictive performance, as reported in the literature, has been tested retrospectively for validity using observational data, it is assigned phase C. If a tool's usability or potential effect have been tested prospectively using small pilots, which may or may not reflect routine practice, it is assigned phase B. Potential effect of a tool is the expected, estimated, or calculated impact of using the tool, assuming it has been successfully implemented and used in clinical practice. Finally, if a tool has been implemented in clinical practice and there is published evidence evaluating its achieved postimplementation impact prospectively, it is assigned phase A.

Level of Evidence

A numerical score within each phase is assigned based on the level of evidence associated with each tool. A tool is assigned grade C1 if it has been tested for external validity multiple times, grade C2 if it has been tested for external validity only once, and grade C3 if it has been tested only for internal validity. Grade C0 means that the tool did not show sufficient internal validity to be used in clinical practice. Grade B1 is assigned to a predictive tool that has been evaluated during the planning for implementation, for both of its potential effect, on clinical effectiveness, patient safety, or health care efficiency, and for its usability. Grade B2 is assigned to a predictive tool that has been evaluated only for its potential effect, while if it has been studied only for its usability, it is assigned grade B3. Finally, if a predictive tool had been implemented and evaluated for its postimplementation impact on clinical effectiveness, patient safety, or health care efficiency, then it is assigned grade A1 if there is at least one experimental study of good quality evaluating its postimplementation impact, grade A2 if there are observational studies evaluating its impact, and grade A3 if the postimplementation impact has been evaluated only through subjective studies, such as expert panel reports.

Direction of Evidence

For each phase and level of evidence, a direction of evidence is assigned based on the collective conclusions reported in the

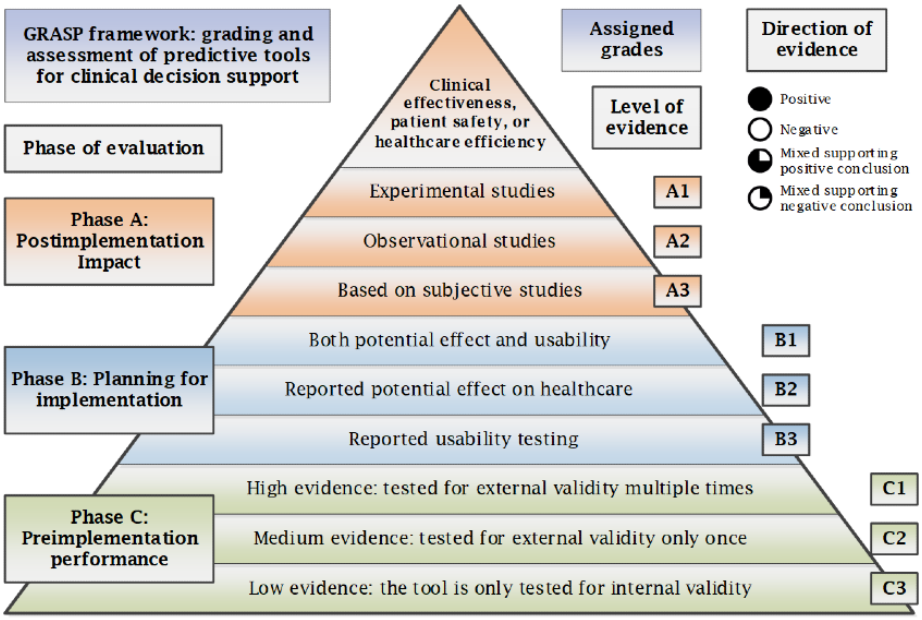
XSL•FO
RenderX

studies. The evidence is considered positive if all studies about a predictive tool reported positive conclusions and negative if all studies reported negative or equivocal conclusions. The evidence is considered mixed if some studies reported positive results, while others reported either negative or equivocal conclusions. To determine the overall direction of evidence, a protocol is used to sort the mixed evidence to support an overall positive or negative conclusion. The protocol is based on 2 main criteria: (1) degree of matching between the evaluation study conditions and the original tool specifications and (2) quality of the evaluation study. Studies evaluating tools in closely matching conditions to the tools' specifications and providing high-quality evidence are considered first for their conclusions in deciding the overall direction of evidence.

The final grade assigned to a predictive tool is based on the highest phase of evaluation, supported by the highest level of positive evidence, or mixed evidence that supports a positive conclusion. More details on the process of critical appraisal of published evidence, summarizing the evidence, and assigning grades to predictive tools are discussed in a published study that describes the development of the GRASP framework [27]. The GRASP framework concept is shown in Figure 1, and a detailed report of the GRASP framework is presented in Multimedia Appendix 1.

The aim of this study was to evaluate the impact of using GRASP on the decisions made by professionals in selecting predictive tools for CDS. The objective was to explore whether the GRASP framework positively supports professionals' evidence-based decision making and improves their accuracy and efficiency in selecting clinical predictive tools. To explore this impact, a group of hypotheses have been proposed including that using the GRASP framework by professionals is going to (1) make their decisions more accurate, that is, selecting the best predictive tools; (2) make their decisions more objective, informed, and evidence-based, that is, decisions are based on the information provided by the framework; (3) make their decisions less subjective, that is, decisions are less based on guessing, prior knowledge, or experience; (4) make their decisions more efficient, that is, decisions are made in less time; and (5) make them face less decisional conflict, that is, become more confident in their decisions and more satisfied with them. We also proposed that using GRASP can move professionals who have less knowledge, less experience, and are less familiar with predictive tools to an equal or even higher accuracy of decision making than professionals who have more knowledge, have more experience, and are more familiar with tools when they do not use GRASP.

Figure 1. The grading and assessment of predictive tools framework concept.



Methods

The Study Design

This study was based on experimental methods. It aimed to examine the performance and outcomes of professionals' decisions in selecting predictive tools with and without using the GRASP framework. Through a web-based survey, the experiment involved asking participants to select the best predictive tool for implementation in clinical practice or for recommendation in clinical practice guidelines from a group of 5 similar tools doing the same predictive task, one time with and another time without using the GRASP framework. In addition, participants were asked a few questions regarding the process of making their decisions through the 2 scenarios. Participants were also requested to provide their feedback on the perceived usability and usefulness of the evidence-based summary of the GRASP framework. This experiment does not include registration of the randomized controlled trial as it does not involve any patients, medications, or treatments.

The emergency department (ED) is among the top health care specialties that are increasingly utilizing predictive tools, especially in the area of managing traumatic brain injury (TBI), which is the leading cause of death and disability among trauma patients [30-33]. Two groups of predictive tools designed to

XSL•FO
RenderX

exclude TBI in the ED were prepared. The first group included 5 tools for predicting TBI in pediatrics: Paediatric Emergency Care Applied Research Network (PECARN) head injury rule, Children's Head injury ALgorithm for the prediction of Important Clinical Events, Canadian Assessment of Tomography for Childhood Head injury rule, Palchak head injury rule, and Atabaki head injury rule [34-38]. The PECARN, being the most validated and the only tool that has been implemented in clinical practice and discussed to have a positive postimplementation impact, is the best tool among the 5 [39,40]. The second group includes 5 tools for predicting TBI in adults: the Canadian CT Head Rule (CCHR), New Orleans Criteria (NOC), Miller criteria for head computed tomography, Kimberley Hospital Rule, and Ibanez model for head computed tomography [41-45]. The CCHR and NOC, being the only tools that have been implemented in clinical practice and are the most validated, showing high predictive performance, are the best tools among the 5 [46-48]. Two scenarios were prepared for this experiment. The first is the control scenario, which includes providing participants with basic information about each tool, the full text of the original studies describing the tools, and allowing them to search the internet for information. The second is the experiment scenario, including providing participants with the main component of the GRASP framework, which is an evidence-based summary of the predictive tools and the full GRASP report on each tool, in addition to allowing them to search the internet for information. To minimize bias, eliminate the pre-exposure effect, and improve the robustness, the experiment includes randomizing the 2 groups of predictive tools and the 2 scenarios. Accordingly, the participants go randomly through 1 of 4 scenarios: (1) pediatric tools without GRASP and then adult tools with GRASP, (2) pediatric tools with GRASP and then adult tools without GRASP, (3) adult tools without GRASP and then pediatric tools with GRASP, and (4) adult tools with GRASP and then pediatric tools without GRASP. Figure 2 shows the survey workflow and the randomization of participants.

The authors recruited a wide group of international professionals to participate in this experiment through a web-based survey. To identify potential participants who work at the ED and those who have knowledge or experience about CDS tools, published studies were used to retrieve the authors' emails and invite them. To retrieve studies on CDS systems, tools, models, algorithms, and pathways or rules used in the ED, emergency service, or emergency medicine published over the last 5 years by professionals who work in the EDs or services of their health care organizations or those who conducted emergency medicine, EDs, or emergency services research, 4 databases were used: MEDLINE, EMBASE, CINAHL, and Google Scholar. The authors expected a response rate of approximately 10%. Before the deployment of the survey, a pilot test was conducted by 10 expert professionals. The feedback of the pilot test was used to improve the survey. Professionals who participated in the pilot test were excluded from participation in the final survey. An invitation email, introducing details about the study objectives, the GRASP framework, the experiment task, the survey completion time, which was estimated at 20 min, and a participation consent was submitted to the identified potential participants with the link to the web-based survey. A reminder email, in 2 weeks, was sent to the potential participants who did not respond or complete the survey. Figure 3 shows the CONSORT 2010 flow diagram of the progress of the randomized trial of the 2 groups: intervention group (GRASP) and control group (No GRASP), showing the enrollment, intervention allocation, follow-up, and data analysis [49,50].

XSL•FO
RenderX

Figure 2. Survey workflow and randomization of the 4 scenarios.

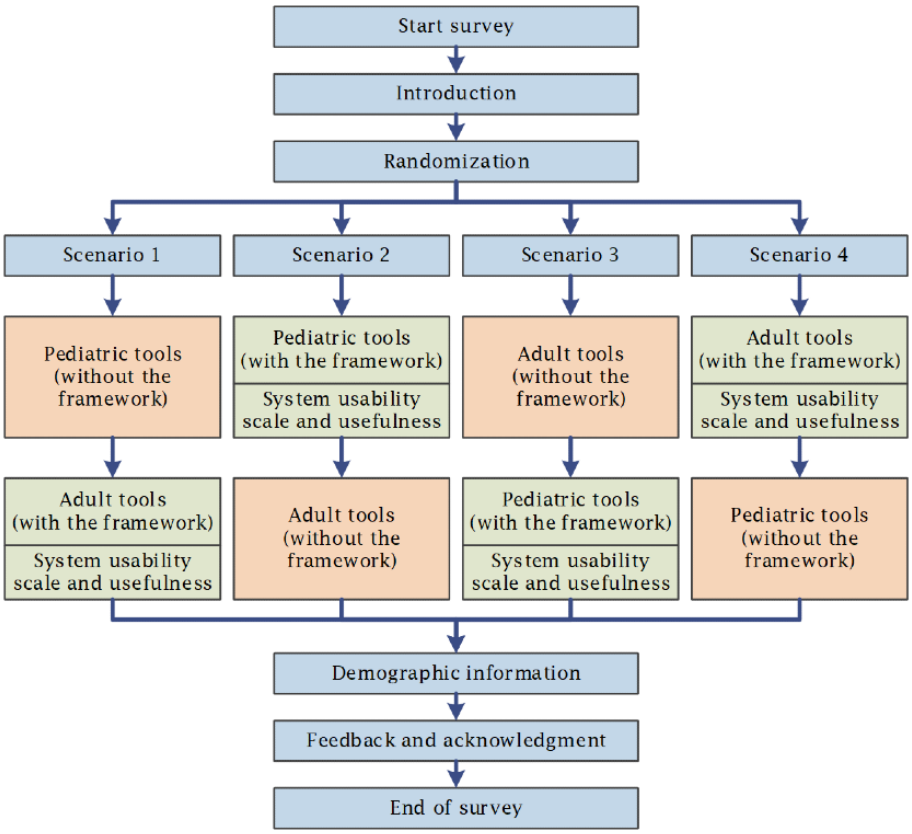
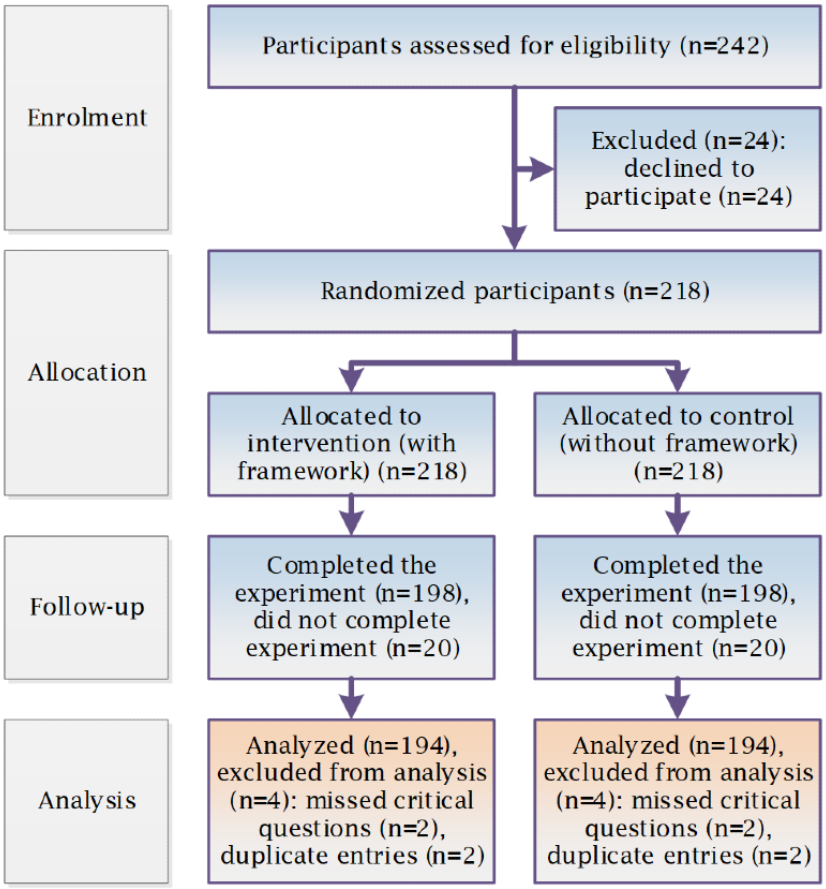


Figure 3. Consolidated Standards of Reporting Trials 2010 flow diagram of the experiment. GRASP: grading and assessment of predictive tools.



XSL•FO
RenderX

The Study Survey

The web-based survey was developed using the Qualtrics Experience Management Solutions Platform [51]. The survey, illustrated through screenshots in the Multimedia Appendix 1, includes 5 sections. The first section includes an introduction to the study objectives, the GRASP framework, and the experiment task. In addition, participants are provided with contacts to request further information or submit complaints. The second section includes randomizing the 2 scenarios and the 2 groups of predictive tools to create the 4 scenarios described above.

In this section, participants are asked to assume that they are the heads of busy EDs and are responsible for selecting the best tool, the most validated in the literature or implemented in clinical practice, out of the 5 diagnostic head injury predictive tools. The PECARN is the correct answer among the 5 pediatric tools, and both the CCHR and the NOC are correct answers among the 5 adult tools. On a 5-point Likert scale, participants were asked to show how much they agreed to the following: (1) they made their decisions based on guessing, (2) they made their decisions based on prior knowledge or experience, (3) they made their decisions based on the information provided in the survey, (4) they were confident in their decisions, and (5) they were satisfied with their decisions. The third section includes asking participants to provide their feedback on the usability of the evidence-based summary of the GRASP framework through a standard set of System Usability Scale (SUS) questions. The SUS is a simple, 10-item attitude Likert scale that provides a global view of subjective assessments of usability. It was developed by John Brooke in 1986 as a tool to be used in the engineering of electronic systems. It is now widely accepted as a tool for the evaluation of system usability [52,53]. Participants were also asked to provide free-text feedback on whether they think the framework is useful or not and why they think so. The fourth section includes participants' demographics, such as their clinical or health care role, specialty, gender, age group, years of experience, and how much they are familiar with head injury predictive tools.

The Study Sample Size

As an initial estimate of the required sample size for this study, and based on similar studies, evaluating the impact of using information systems on professionals' objective, informed, and evidence-based decisions, the authors aimed to recruit a sample of 40 to 60 participants [54-56]. More specifically, it was estimated that a sample size of 46 participants would be sufficient to test for at least a 10% difference, between the 2 arms of the experiment, in the measured outcomes, when using a paired two-tailed *t* test with a significance level of 0.05 and power of 0.95. Calculations were conducted using G*Power software [57].

Analysis and Outcomes

To test the 5 proposed hypotheses, the study was designed to compare the 2 scenarios, making decisions with and without using the GRASP framework, based on a group of 7 measures: (1) time needed for tool selection decision making; (2) accuracy of tool selection decisions; (3) making decisions subjectively based on guessing; (4) making decisions subjectively based on prior knowledge or experience; (5) making decisions objectively based on the information and evidence provided; (6) levels of participants' confidence in their decisions; and (7) levels of participants' satisfaction with their decisions. The accuracy of making decisions, with and without GRASP, was also compared along with knowledge, experience, and familiarity with predictive tools. Table 1 shows the 5 proposed hypotheses and their related 7 outcome measures. To avoid an inflated Type I error and account for the 5 tested hypotheses and the 7 compared measures, the Bonferroni correction was used, by setting the alpha value of the paired samples *t* test to .007 instead of .05. The sample size was re-estimated to 96 participants. In addition, the SUS was calculated for the average rate and distribution of scores. The perceived usefulness and free-text feedback were analyzed. The demographic variables were analyzed for possible correlations or differences.

As this randomized controlled experiment was conducted via a web-based survey of clinicians and health care professions, Table 2 shows the checklist for reporting results from the internet surveys (checklist for reporting results from the internet surveys).

Table 1. Proposed hypotheses and related outcome measures.

Proposed hypotheses	Related outcome measures
Using GRASP[a] will make predictive tools' selection decisions more accurate, that is, selecting the best predictive tools	Accuracy of tools' selection decisions
Using GRASP will make decisions more objective, informed, and evidence-based, that is, decisions are based on the information provided by the framework	Making decisions objectively based on the information and evidence provided in the experiment
Using GRASP will make decisions less subjective, that is, less based on guessing, prior knowledge, or experience	Making decisions subjectively based on guessing and making decisions subjectively based on prior knowledge or experience
Using GRASP will make decisions more efficient, that is, decisions are made in less time	The time needed for tools' selection decision making
Using GRASP will make participants face less decisional conflict, that is, be more confident and satisfied with decisions	Levels of participants' confidence in their decisions and levels of participants' satisfaction with their decisions

[a]GRASP: grading and assessment of predictive tools.

XSL•FO
RenderX

Table 2. The checklist for reporting results from the internet surveys checklist.

Item category and checklist item	Explanation
Design	
Describe survey design	A randomized controlled trial experiment testing the impact of using the GRASP framework on clinicians and health care professionals' decisions in selecting predictive tools for CDS[a], using a convenience invited sample to participate in the experiment
IRB[b] approval and informed consent process	
IRB approval	The experiment was approved by the Human Research Ethics Committee, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia
Informed consent	Informed consent was introduced at the beginning of the survey for participants to agree before they take the survey, including the length of time of the survey, types of data collected and its storage, investigators, and the purpose of the study
Data protection	Collected personal information was protected through Macquarie University account on Qualtrics survey system
Development and pretesting	
Development and testing	The first author (MK) developed the survey and pilot tested the questions and its usability before deploying the survey to the participants
Recruitment process and description of the sample having access to the questionnaire	
Open survey versus closed survey	This was a closed survey; only invited participants had access to complete the survey
Contact mode	An initial contact, via email, was sent to all invited participants. Only those who agreed to participate completed the web-based survey
Advertising the survey	The survey was not advertised. Only invited participants were informed of the study and completed the survey
Survey administration	
Web or email	The survey was developed using the Qualtrics survey platform, and the link to the web-based survey was sent to invited participants via email. Responses were automatically collected through the Qualtrics survey platform then retrieved by the investigators for analysis
Context	Only invited participants were informed of the study via email
Mandatory/voluntary	The survey was not mandatory for invited participants
Incentives	The only incentive was that participants could request to be acknowledged in the published study. Participants were also informed of the results of the survey after the analysis is complete
Time/date	Data were collected over 6 weeks, from March 11 to April 21, 2019
Randomization of items or questionnaires	To prevent biases, items were randomized. Figure 2 shows the survey workflow and randomization of 4 scenarios
Adaptive questioning	Four scenarios were used and randomized, but they were not conditionally displayed
Number of items	From 5 to 8 items per page
Number of screens (pages)	The questionnaire was distributed over 5 pages
Completeness check	Completeness checks were used after the questionnaire was submitted, and mandatory items were highlighted. Items provided a nonresponse option "not applicable" or "don't know"
Review step	Respondents were able to review and change their answers before submitting their answers
Response rates	
Unique site visitor	We used the IP[c] addresses to check for unique survey visitors

XSL•FO
RenderX

Item category and checklist item	Explanation
View rate (ratio of unique survey visitors/unique site visitors)	Only invited participants had access to the survey. Survey visitors included those who completed the survey and those who started the survey but did not complete it or gave incomplete answers
Participation rate (ratio of unique visitors who agreed to participate/unique first survey page visitors)	The recruitment rate was 90% (218 participants agreed to participate out of 242 invited participants who visited the first page)
Completion rate (ratio of users who finished the survey/users who agreed to participate)	The completion rate was 91% (198 participants completed the survey out of 218 participants who agreed to participate)
Preventing multiple entries from the same individual	
Cookies used	Cookies were not used to assign a unique user identifier; instead, we used users' computer IP to identify unique users
IP address check	The IP addresses of participants' computers were used to identify potential duplicate entries from the same user. Only 2 duplicate entries were captured and were eliminated before analysis
Log file analysis	We also checked the provided demographic information, of all participants, to make sure the 2 identified duplicates were the only incidents
Registration	Data were collected and the user IP and other demographic data were used later on to eliminate duplicate entries before analysis. Most recent entries were used in the analysis
Analysis	
Handling of incomplete questionnaires	Only completed surveys were used in the analysis
Questionnaires submitted with an atypical timestamp	The task completion time was captured. However, no specific timeframe was used. In the analysis, we excluded statistical outliers, since the survey allowed users to re-enter after a while, for example, the next day. This is discussed in the paper
Statistical correction	No statistical correction was required

[a]CDS: clinical decision support.

[b]IRB: institutional review board.

[c]IP: internet protocol.

Results

Descriptive Analysis

Out of 5857 relevant publications retrieved, 3282 professionals were identified and invited to participate in the survey. Over the survey duration of 6 weeks, from March 11 to April 21, 2019, we received a total of 194 valid responses, with a response rate of 5.9%. Valid responses were identified as those who completed the survey until the end and answered questions in all the survey sections, with no missing sections. Six participants missed answering one or more questions in one or more of the survey sections, 5 participants did not provide their demographics, and 57 participants did not wish to be acknowledged in the study. The detailed distributions of participants based on gender, age group, years of experience, clinical and health care role, clinical specialty, familiarity with head injury predictive tools, and their countries are illustrated in Multimedia Appendix 1, Figures 4-10.

The GRASP Impact on Participants' Decisions

Using the GRASP framework, an evidence-based summary of predictive tools and a detailed report on each predictive tool, along with allowing participants to search the internet for further information, made them select the correct tools 88.1% of the time. Without GRASP, that is, using the basic information about the predictive tools, the full text of the studies describing each tool, along with allowing participants to search the internet for further information, they selected the correct tools 53.7% of the time. This shows a statistically significant improvement of 64% (88.1/53.7=1.64; $P<.001$). On a 5-point Likert scale, where strongly agree is considered equal to 5 and strongly disagree is considered equal to 1, the participants reported that they made their tools' selection decisions based on guessing with an average of 1.98 (SD 1.22), when they used GRASP, compared with an average of 2.48 (SD 1.37), when they did not use GRASP. This shows a statistically significant reduction of 20% ($P<.001$). Participants reported that they made their tools' selection decisions based on their prior knowledge or experience with an average of 3.27 (SD 1.44) when they used GRASP, compared with an average of 3.55 (SD 1.31), when they did not use GRASP. This shows a statistically significant reduction of 8% ($P=.004$).

Participants reported that they made their tools' selection decisions based on the information provided in the survey with an average of 4.10 (SD 1.10) when they used GRASP, compared with an average of 3.11 (SD 1.42), when they did not use GRASP. This shows a statistically significant increase of 32% ($P<.001$). Participants reported that they were confident in their decisions, with an average of 3.96 (SD 0.87), when they used GRASP, compared with an average of 3.55 (SD 1.15), when they did not use GRASP. This shows a statistically significant increase of 11% ($P<.001$). Participants reported that they were

XSL•FO
RenderX

satisfied with their decisions with an average of 3.99 (SD 0.89), when they used GRASP, compared with an average of 3.54 (SD 1.20), when they did not use GRASP. This shows a statistically significant increase of 13% ($P<.001$). The duration of completing the task of selecting predictive tools showed high variability, with many statistical outliers. In addition to the average, the authors used the percentiles to avoid the effect of extreme outliers. The average duration of making the selection decisions showed a statistically insignificant reduction of 52% from 14.5 7 min ($P=.39$). There is also a reduction of 18.9% from 2.2 to 1.8 min on the 50th percentile, 37.3% from 5.3 to 3.3 min on the 75th percentile, 48% from 12.4 to 6.4 min on the 90th percentile, and 30.6% from 19.2 to 13.3 min on the 95th percentile. Table 3 shows the impact of using GRASP on the 7 measures: decision accuracy, guessing, subjective decisions, objective decisions, confidence in decisions, satisfaction with decisions, and task completion duration 90th percentile in minutes.

Using a paired samples *t* test, Table 4 shows the estimation for the paired difference of the 7 measures and the effect size, calculating and interpreting the eta-squared statistic, based on the guidelines proposed by Cohen [58].

Table 5 compares physicians to nonphysicians, emergency medicine to other specialties, familiar with tools to nonfamiliar, male to female, younger to older, and less experienced to more experienced participants. The GRASP detailed report is shown in Multimedia Appendix 1. The GRASP evidence-based summaries of the 2 groups of pediatric and adult predictive tools are shown in Multimedia Appendix 1.

Table 3. The impact of using grading and assessment of predictive tools on participants' decisions (n=194).

Criteria	No GRASP[a]	GRASP	Change (%)	*P* value
Score (0 to 100%)	53.7	88.1	64	<.001
Guessing (1 to 5)	2.48	1.98	−20	<.001
Subjective (1 to 5)	3.55	3.27	−8	.003
Objective (1 to 5)	3.11	4.10	32	<.001
Confidence (1 to 5)	3.55	3.96	11	<.001
Satisfaction (1 to 5)	3.54	3.99	13	<.001
Time in min (90th percentile)	12.4	6.4	−48	.38

[a]GRASP: grading and assessment of predictive tools.

Table 4. Estimation for paired difference and effect size.

Measure	Mean (SD)	SE	99.3% CI[a]	*t* test (*df*)	*P* value	Effect size[b]	
						Value	Actual size
Score	0.340 (0.555)	0.040	0.231 to 0.449	8.53 (193)	<.001	0.274	Large
Guessing	−0.519 (1.303)	0.095	−0.777 to −0.260	−5.47 (188)	<.001	0.134	Moderate
Subjective	−0.319 (1.464)	0.107	−0.613 to −0.028	−2.99 (187)	.003	0.044	Small
Objective	1.005 (1.496)	0.109	0.709 to 1.302	9.24 (189)	<.001	0.307	Large
Confidence	0.392 (1.261)	0.092	0.141 to 0.642	4.27 (188)	<.001	0.086	Moderate
Satisfaction	0.439 (1.235)	0.090	0.194 to 0.684	4.89 (188)	<.001	0.110	Moderate
Duration[c]	−447 (7152)	514	−1847 to 952	−0.87 (193)	.39	N/A[d]	N/A

[a]Bonferroni correction conducted.

[b]Effect size calculated using the eta-square statistic (0.01=small effect, 0.06=moderate effect, and 0.14=large effect [58]).

[c]Task completion duration is reported in seconds.

[d]N/A: not applicable.

XSL•FO
RenderX

Table 5. Comparing the impact of grading and assessment of predictive tools on participant groups.

Health care professional group	Criteria						
	Score (0 to 100%)	Guessing (1 to 5)	Subjective (1 to 5)	Objective (1 to 5)	Confidence (1 to 5)	Satisfaction (1 to 5)	Time in min (90th percentile)
Role							
Physicians (n=130)							
No GRASP[a]	61.4	2.4	3.7	3.0	3.6	3.6	10.9
GRASP	89.0	2.0	3.5	4.0	4.0	4.0	6.1
Change (%)	45	−18	−5	33	10	12	−44
P value	<.001	<.001	.080	<.001	<.001	<.001	.62
Nonphysicians (n=59)							
No GRASP	37	2.7	3.3	3.5	3.5	3.5	15.3
GRASP	85	2.0	2.8	4.4	3.8	3.9	6.6
Change (%)	127	−25	−16	28	10	14	−57
P value	<.001	<.001	.007	<.001	.047	.008	.26
Specialty							
Emergency (n=94)							
No GRASP	73	2.4	4.1	2.8	3.8	3.8	11.0
GRASP	93	1.9	3.7	3.8	4.1	4.1	6.5
Change (%)	29	−19	−10	36	6	7	−41
P value	<.001	<.001	.009	<.001	.07	.04	.51
Nonemergency (n=95)							
No GRASP	36	2.6	3.0	3.4	3.3	3.2	15.0
GRASP	83	2.0	2.9	4.4	3.8	3.8	6.5
Change (%)	129	−21	−6	28	15	19	−57
P value	<.001	<.001	.096	<.001	.001	<.001	.11
Familiarity with tools							
Familiar (n=108)							
No GRASP	67.0	2.3	4.1	2.8	3.8	3.8	8.1
GRASP	89.6	1.8	3.7	3.8	4.1	4.1	5.3
Change (%)	34.0	−22.0	−10.0	39.0	8.0	8.0	−34.0
P value	<.001	<.001	.007	<.001	.016	.013	.51
Not familiar (n=81)							
No GRASP	36	2.7	2.8	3.6	3.3	3.2	18.2
GRASP	85	2.2	2.7	4.5	3.7	3.8	7.9

XSL•FO
RenderX

Health care professional group	Criteria						
	Score (0 to 100%)	Guessing (1 to 5)	Subjective (1 to 5)	Objective (1 to 5)	Confidence (1 to 5)	Satisfaction (1 to 5)	Time in min (90th percentile)
Change (%)	134	−18	−5	23	14	19	−57
P value	<.001	.002	0.16	<.001	.003	<.001	.24
Gender							
Males (n=120)							
No GRASP	54.2	2.3	3.5	3.1	3.7	3.6	13.5
GRASP	82.2	2.0	3.3	4.1	3.9	4.0	7.4
Change (%)	52	−14	−7	33	8	10	−45
P value	<.001	.005	.08	<.001	.009	.002	.41
Females (n=67)							
No GRASP	55	2.9	3.5	3.3	3.3	3.4	12.2
GRASP	97	2.0	3.1	4.3	3.9	4.0	5.3
Change (%)	78	−30	−12	29	17	18	−56
P value	<.001	<.001	.004	<.001	.004	.001	.54
Age (years)							
Younger (<45 years, n=112)							
No GRASP	59	2.6	3.6	3.1	3.5	3.5	9.1
GRASP	87	2.0	3.3	4.1	4.0	4.0	6.0
Change (%)	48	−25	−7	34	13	14	−34
P value	<.001	<.001	.06	<.001	.001	.001	.45
Older (>45 years, n=77)							
No GRASP	47	2.3	3.5	3.2	3.6	3.6	15.9
GRASP	88	2.0	3.2	4.1	3.9	4.0	7.7
Change (%)	89	−13	−10	28	7	10	−52
P value	<.001	.03	.009	<.001	.08	.004	.19
Experience							
Less experience (<15 years, n=94)							
No GRASP	59	2.6	3.6	3.0	3.5	3.4	8.1
GRASP	87	2.0	3.3	4.0	3.9	4.0	6.5
Change (%)	48	−24	−7	36	12	16	−20
P value	<.001	<.001	.09	<.001	.009	.001	.46
More experience (>15 years, n=95)							

XSL•FO
RenderX

Health care professional group	Criteria						
	Score (0 to 100%)	Guessing (1 to 5)	Subjective (1 to 5)	Objective (1 to 5)	Confidence (1 to 5)	Satisfaction (1 to 5)	Time in min (90th percentile)
No GRASP	49	2.4	3.5	3.3	3.6	3.6	15.0
GRASP	88	2.0	3.2	4.2	3.9	4.0	6.8
Change (%)	80	−16	−10	28	9	9	−54
P value	<.001	.004	.006	<.001	.004	.004	.11

[a]GRASP: grading and assessment of predictive tools.

The GRASP Usability and Usefulness

The overall SUS rate of the GRASP framework and evidence-based summary, considering the responses of all 194 participants, was 72.5%, which represents a very good level of usability [59,60]. Examining the influence of demographics on the SUS rates, only 2 factors showed significant influence: the gender of participants and their familiarity with predictive tools. The female participants reported a statistically significant higher SUS rate (76.2%) in comparison with the male participants (70.8%), showing that female participants, more than male participants, thought GRASP is easy to use. Using the statistical Spearman correlation test, the degree of familiarity with head injury predictive tools showed a weak negative statistically significant correlation with the GRASP SUS score (P=.03). This indicates that participants who were less familiar with predictive tools thought that the GRASP framework was easy to use more than participants who were more familiar with the tools.

Among the 194 valid responses of participants, almost two-third (122) provided free-text feedback on the GRASP evidence-based summary usefulness and explained their feedback. The qualitative analysis of the open-ended question was conducted using the NVivo Version 12.3 software package [61]. Most respondents (88%, 108/122) reported that they found the GRASP evidence-based summary useful. They explained their responses with various reasons, mainly that the evidence-based summary was simple, clear, and logical. Some reported that the visual presentation was attractive, intuitive, and self-explanatory. Others reported that it concisely and comprehensively provided a summary of extensive information, and some reported that the presented information was consistent, easily comparable, making it easy to make informed decisions. A smaller group of 12% of participants reported that they found the GRASP evidence-based summary useless. They reported that it did not provide enough information to make informed decisions. Some reported that it was not clear enough, or simple enough, to understand and use to select predictive tools. One health care professional reported that “it is too complicated and needs to be simplified further,” while another reported that “it is oversimplified and missing some important parameters.” One health care professional reported “it might be more helpful when the decision is less clear” and added, “I would like to see more info on the strengths/weaknesses of each tool.”

Discussion

Brief Summary

The use of GRASP has positively supported, and significantly improved, evidence-based decision making and increased the accuracy and efficiency of selecting predictive tools. Using the GRASP framework has significantly increased correct decisions and objective decision making, and significantly decreased subjective decision making based on guessing, prior knowledge, or experience. Moreover, using the GRASP framework significantly decreased decisional conflict, increasing the confidence and satisfaction of participants with their decisions. Furthermore, using the GRASP framework decreased the task completion time for selecting predictive tools. In addition, the average SUS of the GRASP framework was very good, and most participants found the GRASP framework useful.

It is a challenging task for most health care professionals to critically evaluate a growing number of predictive tools, proposed in the literature, to select effective tools for implementation in clinical practice or for recommendation in clinical guidelines, to be used by other professionals. Although most of these predictive tools have been assessed for predictive performance, only a few have been implemented and evaluated for comparative effectiveness or postimplementation impact. Professionals need an evidence-based approach to provide them with standardized objective information on predictive tools to support their search for and selection of effective tools for clinical tasks. On the basis of the critical appraisal of the published evidence, the GRASP framework uses 3 dimensions to grade predictive tools: (1) phase of evaluation, (2) level of evidence, and (3) direction of evidence. The final grade assigned to a tool is based on the highest phase of evaluation, supported by the highest level of positive evidence, or mixed evidence that supports a positive conclusion. In this study, we present an evaluation of the impact of the GRASP framework on professionals’ decisions in selecting predictive tools for CDS.

The Impact of GRASP on Participants’ Decisions

The GRASP framework provides a systematic and transparent approach for professionals to make objective, well-informed,

XSL•FO
RenderX

and evidence-based decisions regarding the selection of predictive tools. This is very similar to the findings of using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework in evaluating the quality of evidence and strength of recommendations regarding treatment methods and decisions endorsed in clinical guidelines [62,63]. The quality of decision making, while developing clinical guidelines, depends significantly on the quality of the evidence-informed analysis and advice provided [64]. Similarly, supporting professionals with evidence improves their accuracy and helps them make better clinical decisions and better organizational decisions [65,66]. Similarly, using GRASP and providing professionals with evidence-based information on predictive tools significantly improved professionals' accuracy of decisions in selecting the best predictive tools.

Providing professionals with GRASP evidence-based information also enabled them to minimize subjective decision making, such as guessing, prior knowledge, or previous experience. This has been discussed in other studies investigating the role of utilizing evidence-based resources in decreasing subjective bias in making clinical, population-related, and health policy decisions [67,68]. Evidence-based information on GRASP was associated with a decrease in professionals' decisional conflict by increasing their confidence in their decisions and their satisfaction with them. This has been discussed in similar studies reporting the impact of evidence-based information on decreasing decisional conflicts faced by both professionals and patients when they make clinical decisions [69-71]. When time is a sensitive factor for critical clinical and population decisions, efficient decision making becomes important [72]. Here comes the role of evidence-based decision making, which is discussed to be not only more accurate, objective, and of higher quality but also much more efficient [73,74]. Similarly, providing professionals with GRASP evidence-based information improved their efficiency in making predictive tools' selection decisions.

Using GRASP made nurses and other professionals make more accurate decisions than physicians when they are not using GRASP. Using GRASP, clinicians of specialties other than emergency medicine make better decisions than emergency medicine clinicians without GRASP. Furthermore, using GRASP, professionals who were not familiar with head injury predictive tools made better decisions than professionals who were familiar with the tools without GRASP. Furthermore, the use of GRASP made decisions more efficient. Accordingly, using GRASP has moved professionals with less knowledge, less experience, and less familiarity with predictive tools to higher accuracy, higher efficiency, and better decision-making levels than professionals who had more knowledge, had more experience, and were more familiar with tools, but did not use GRASP.

The Usability and Usefulness of GRASP

The usability of systems is an important foundation for successful implementation and utilization [75]. Usability can be evaluated by measuring the effectiveness of task management with accuracy and completeness, measuring the efficiency of utilizing resources in completing tasks and measuring users' satisfaction, comfort with, and positive attitudes toward, the use of the tools [76,77]. One of the validated and simply applicable methods of measuring usability is the SUS [52,53]. When users have more experience with a system, they tend to provide higher, more favorable SUS scores for the system usability over users with either no or limited experience [78]. On the other hand, when users have less experience with a system, they tend to see new tools illustrating the system, or new approaches to understanding it, more usable than users who have extensive experience with the system itself [79]. This explains why the degree of familiarity with the tools was negatively correlated with the GRASP SUS score, where participants less familiar with tools provided higher SUS scores for GRASP than participants who were more familiar. It is reported in the literature that gender does not influence the perceived usability or usefulness of systems [80,81]. This was not the case with GRASP, where female participants provided higher SUS scores than males. Furthermore, female participants also thought that GRASP is more useful than males. Both findings could be explained by the greater improvement in female participants' confidence and satisfaction with their decisions when they used GRASP compared with male participants. Some participants' suggestions, reported in the free-text feedback, can be used in the future to add more information to the GRASP detailed report on each tool.

Study Conclusions

Through this study, the GRASP framework is presented as an effective evidence-based approach to support professionals' decisions when selecting predictive tools for implementation in clinical practice or for recommendation in clinical practice guidelines. Using the GRASP framework and the evidence-based summary improved the accuracy of selecting the best predictive tools, with an increased objective, informed, and evidence-based decision making and decreased subjective decision making based on guessing, prior knowledge, or experience. Using GRASP also decreased the decisional conflict faced by professionals by improving their confidence and satisfaction with their decisions. Using GRASP has also improved the efficiency of professionals in making their selection decisions by decreasing the time needed to complete the decision-making task.

The GRASP framework represents a high-level approach to provide professionals with an evidence-based and comprehensive, yet simple and feasible method to evaluate and select predictive tools. However, when professionals need further information, the detailed framework report provides them with the required details to support their decision making. The GRASP framework is designed for 2 levels of users:

(1) Expert users, such as health care researchers, experienced in evidence-based evaluation methods. They will use the framework to critically evaluate published evidence, assign grades to predictive tools, and report their details.

(2) End users, such as clinicians and health care professionals, responsible for selecting tools for implementation in clinical practice or for recommendation in clinical guidelines. They will use the GRASP framework detailed reports on tools and their

XSL•FO
RenderX

assigned grades, produced by expert users, to compare existing predictive tools and select the most suitable tools [27].

The GRASP framework is not meant to be absolutely prescriptive. A lower grade tool could be preferred by a health care professional to improve clinical outcomes that are not supported by a higher grade one. For example, a practicing clinician may prefer an A2 tool showing improved patient safety in 2 observational studies rather than an A1 tool showing reduced health care costs in three experimental studies because they are now trying to improve patient safety to avoid reducing health care costs. It all depends on the objectives and priorities that the clinicians and health care professionals are trying to achieve. In addition, sometimes, more than one predictive tool should be endorsed in clinical practice guidelines, each supported by its requirements for application, conditions of use, and recommended for its most prominent outcomes of predictive performance or postimplementation impact on health care and clinical outcomes. Furthermore, even when GRASP assigns high grades to predictive tools, some of these tools may not be simply recommended for use in a different country or population than the ones that were used to develop and validate the tools in the first place. This might happen because of the population-related differences in the risks associated with the incidence of certain medical conditions, outcomes, or prognoses. This necessitates adjustment of the tools to the local context, thereby producing new versions of the tools, which requires re-evaluation by GRASP.

Although the GRASP framework has been developed to assess and grade predictive tools and other similar CDS systems, the application of the framework concept of grading tools and systems based on the published evidence is not limited to predictive tools or CDS systems. The GRASP framework concept can be applied to assess and grade many other types of clinical tools, systems, and methods.

Study Limitations and Future Work

Although we received a large and sufficient number of 194 valid responses, the very low response rate of 5.9% could have been improved if potential participants were motivated by some incentives. They could have also been motivated if more support was provided through their organizations, which need more resources to synchronize such efforts. For the sake of keeping the survey feasible, for most busy professionals, the number of questions was kept limited and the time required to complete the survey was kept in the range of 20 min. However, some of the participants showed their willingness to provide more detailed feedback, which could have been done through interviews, for example, but this was out of the scope of the study and was not initially possible with the huge number of invited participants. The reduction in the decision-making duration of selecting predictive tools, while using GRASP, was statistically insignificant, because of the high variability and extreme statistical outliers, with and without GRASP. This could be explained by the fact that the Qualtrics platform of the survey measures the task completion duration by subtracting the time of loading the page from the time of pushing the Next button after completing the task and not the actual time the participants spent active on the page, which is currently under development [82].

To enable a wider global audience of clinicians, health care professionals, and clinical guideline developers to access detailed information, reported evidence, and assigned grades of different predictive tools, it is essential to implement the GRASP framework into a web-based platform. However, maintaining such a grading system up to date is a challenging task, as this requires continuous updating of the predictive tools grading and assessments when newly published evidence becomes available. In addition, the entire process is currently conducted manually, which represents a large burden on assessing and grading the huge number of existing predictive tools and those continuously emerging. Accordingly, it is essential to use automated or semiautomated methods for searching and processing new information to keep the GRASP framework information, grades, and assessments updated. Finally, we recommend that the GRASP framework be utilized by working groups of professional organizations to grade predictive tools to provide consistent results and increase reliability and credibility for end users. These professional organizations should also support disseminating such evidence-based information on predictive tools, similar to announcing and disseminating new updates of clinical practice guidelines.

Acknowledgments

The authors would like to thank Peter Petocz, Professor of Statistics at Macquarie University, for his guidance and support in the statistical analysis. They would also like to thank all the professors, doctors, and researchers who participated in evaluating the impact of the GRASP framework, including Aaron Brody, Adam Rose, Adel Hamed Elbaih, Aleix Martínez-Pérez, Alireza Razzaghi, Amanda Montalbano, Andreas Huefner, Andreas Sundermeyer, Andres M Rubiano, Andrew MacCormick, Andrew Miller, Antoinette Conca, Ashraf Bedaewi, Asit Misra, Ausra Snipaitiene, Azadeh Tafakori, Başak Bayram, Bedia Gulen, Bhagavatula Indira Devi, Brent Becker, Brian Rowe, Chad Cannon, Chen-June Seak, Chimwemwe Mula, Christian Waydhas, Christophe Marti, Christopher R Carpenter, Claire Vajdic, Claudine Blum, Clermont E Dionne, Corrado Tagliati, David R Vinson, Dawn Moeller, Deepak Batura, Deepika Mohan, DJ Choi, Duane Steward, Dustin Ballard, Edward Baker, Elise Gane, Elizabeth Manias, Elizabeth Powell, Ellen Weber, Emanuele Gilardi, Eveline Hitti, Ewout W Steyerberg, Eva Bitzer, Fawaz Abdullah Alharbi, Fayza Haider, Fernanda Bellolio, Fleur Lorton, Francesco Fleres, Gabriel Rodrigues, Gerry Lee, Giacomo Veronese, Gianfranco Sinagra, Giovanni de Simone, Guillaume Foldes-Busque, Gulten Sucu, Guo XH, Gwendolyn Vuurberg, Herbert J Fernandes, Hon Lon Tam, Hsu Teh-Fu, Ibrahim Jatau Abubakar, Ilias Galanopoulos, Iqra Manzoor, Jennifer Hoffmann, JM Ferreras, Joel Noutakdie Tochie, John Kellett, Jonathan Elmer, Jorge Acosta-Reyes, Juan Del Castillo-Calcáneo, Julia Lappin,

XSL•FO
RenderX

Kai Eggers, Karin Pukk Härenstam, Katarina Lockman Frostred, Ken Hillman, Kevin Heard, Kimberly D Johnson, Kristen Miller, Kristin Baltrusaitis, L Beenen, Larry Figgs, Lauren Southerland, Lina Jankauskaitė, Luca Molinari, Majed Althagafi, Makini Chisolm-Straker, Marc Probst, Marcello Covino, Marco Daverio, Marie Méan, Mariza Elsi, Mark Ebell, Martin Nordberg, Matt Reed, Matthew J Douma, Mauro Podda, Melanie Dechamps, Michael Joseph Barrett, Mieke Deschodt, Mojtaba Mojtahedzadeh, Molly Moore Jeffery, Muhammad Waseem, Nathan Hoot, Niccolò Parri, Nicola Ramacciati, Nurhafiza Yezid, Olga H Torres, Özlem Köksal, Paola Fugazzola, Paolo Navalesi, Paul L Aronson, Paul Monagle, Pedro J Marcos, Peter Dayan, Peter Nugus, Pinchas Halpern, Prabath Lodewijks, Prosen Gregor, Rachel Seymour, Rachid Mahmoudi, Rafael Oliveira Ximenes, Rafal Karwowski, Rahul Kashyap, Ricardo Fernandes, Rodolfo J Oviedo, Ross I Donaldson, Sabrina De Winter, Sandeep Sahu, Sangil Lee, Sebastián Camino, Sheri Carson, Sivera Berben, Stephane Tshitenge, Sukhyang Lee, Suzanne Tamang, Thomas Hartka, Thys Frédéric, Tim Söderlund, Tiziana Ciarambino, Toby Keene, Tomas Vedin, Vincenzo G Menditto, Wei-Chieh Lee, and William Mower.

This work was financially supported by the Commonwealth Government Funded Research Training Program, Australia, to help the authors carry out the study. The funding body played no role in the design of the study and collection, analysis, or interpretation of the data or in writing the manuscript. This study was approved by the Human Research Ethics Committee, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia, on October 4, 2018. Reference no: 5201834324569. Project ID: 3432.

Authors' Contributions

MK mainly contributed to the conception, detailed design, and conduction of the study. BG and FM supervised the study from a scientific perspective. BG was responsible for the overall supervision of the work done, while FM was responsible for providing advice on the enhancement of the methodology used. All the authors have been involved in drafting the manuscript and revising it. Finally, all the authors approved the manuscript to be published and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The GRASP Framework detailed report.
[PDF File (Adobe PDF File), 2439 KB - jmir_v22i7e15770_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 16168 KB - jmir_v22i7e15770_app2.pdf]

References

1. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
2. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. J Am Med Assoc 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
3. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Br Med J 2005 Apr 2;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
4. Oman KS, Houser J. Evidence-Based Practice: An Implementation Guide for Healthcare Organizations. Burlington, Massachusetts: Jones & Bartlett Publishers; 2010.
5. Osheroff J. Improving Outcomes with Clinical Decision Support: An Implementer's Guide. Boca Raton, Florida: Productivity Press; 2012.
6. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-145 [FREE Full text] [doi: 10.1197/jamia.M2334] [Medline: 17213487]
7. Légat L, van Laere S, Nyssen M, Steurbaut S, Dupont AG, Cornu P. Clinical decision support systems for drug allergy checking: systematic review. J Med Internet Res 2018 Sep 7;20(9):e258 [FREE Full text] [doi: 10.2196/jmir.8206] [Medline: 30194058]
8. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of decision support in computerized provider order entry: systematic literature review. JMIR Med Inform 2018 Jan 24;6(1):e3 [FREE Full text] [doi: 10.2196/medinform.7170] [Medline: 29367187]

XSL•FO
RenderX

9. Monteiro L, Maricoto T, Solha I, Ribeiro-Vaz I, Martins C, Monteiro-Soares M. Reducing potentially inappropriate prescriptions for older patients using computerized decision support tools: systematic review. J Med Internet Res 2019 Nov 14;21(11):e15385 [FREE Full text] [doi: 10.2196/15385] [Medline: 31724956]
10. Bezemer T, de Groot MC, Blasse E, ten Berg MJ, Kappen TH, Bredenoord AL, et al. A human(e) factor in clinical decision support systems. J Med Internet Res 2019 Mar 19;21(3):e11732 [FREE Full text] [doi: 10.2196/11732] [Medline: 30888324]
11. Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. New York, USA: Springer; 2014:643-674.
12. Shortliffe E, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. New York, USA: Springer; 2013.
13. Adams ST, Leveson SH. Clinical prediction rules. Br Med J 2012 Jan 16;344:d8312. [doi: 10.1136/bmj.d8312] [Medline: 22250218]
14. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med 1985 Sep 26;313(13):793-799. [doi: 10.1056/NEJM198509263131306] [Medline: 3897864]
15. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Aust J Physiother 2006;52(3):157-163. [doi: 10.1016/s0004-9514(06)70024-1]
16. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, USA: Springer; 2008.
17. Ebell MH. Evidence-Based Diagnosis: A Handbook of Clinical Prediction Rules. Volume 1. New York, USA: Springer; 2001.
18. Kappen T, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KG. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res 2018;2:11 [FREE Full text] [doi: 10.1186/s41512-018-0033-6] [Medline: 31093561]
19. Taljaard M, Tuna M, Bennett C, Perez R, Rosella L, Tu JV, et al. Cardiovascular disease population risk tool (CVDPoRT): predictive algorithm for assessing CVD risk in the community setting. A study protocol. BMJ Open 2014 Oct 23;4(10):e006701 [FREE Full text] [doi: 10.1136/bmjopen-2014-006701] [Medline: 25341454]
20. Clinician. Dictionary by Merriam-Webster. 2019 Jan 01. URL: https://www.merriam-webster.com/dictionary/clinician [accessed 2019-01-01]
21. Ansari S, Rashidian A. Guidelines for guidelines: are they up to the task? A comparative assessment of clinical practice guideline development handbooks. PLoS One 2012;7(11):e49864 [FREE Full text] [doi: 10.1371/journal.pone.0049864] [Medline: 23189167]
22. Kish MA, Infectious Diseases Society of America. Guide to development of practice guidelines. Clin Infect Dis 2001 Mar 15;32(6):851-854. [doi: 10.1086/319366] [Medline: 11247707]
23. Shekelle P, Woolf SH, Eccles M, Grimshaw J. Developing clinical guidelines. West J Med 1999 Jun;170(6):348-351 [FREE Full text] [Medline: 18751155]
24. Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 2003 Sep;14(3):207-222. [doi: 10.1111/1467-8551.00375]
25. Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. Br Med J 2005 Sep 17;331(7517):631-633 [FREE Full text] [doi: 10.1136/bmj.331.7517.631] [Medline: 16166137]
26. O'Connor A. User Manual-Decisional Conflict Scale. Patient Decision Aids - Ottawa Hospital Research. 2010. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf [accessed 2019-01-01]
27. Khalifa M, Magrabi F, Gallego B. Developing a framework for evidence-based grading and assessment of predictive tools for clinical decision support. BMC Med Inform Decis Mak 2019 Oct 29;19(1):207 [FREE Full text] [doi: 10.1186/s12911-019-0940-7] [Medline: 31664998]
28. Khalifa M, Magrabi F, Gallego B. Validating and updating GRASP: a new evidence-based framework for grading and assessment of clinical predictive tools. arxiv 2019:- epub ahead of print(1907.11524) [FREE Full text] [doi: 10.21203/rs.3.rs-15929/v1]
29. Khalifa M, Gallego B. Grading and assessment of clinical predictive tools for paediatric head injury: a new evidence-based approach. BMC Emerg Med 2019 Jun 14;19(1):35 [FREE Full text] [doi: 10.1186/s12873-019-0249-y] [Medline: 31200643]
30. Azim AB, Joseph B. Traumatic brain injury. In: Salim A, Brown CV, Inaba K, Martin M, editors. Surgical Critical Care Therapy: A Clinically Oriented Practical Approach. New York, USA: Springer; 2018:1-10.
31. Greve MW, Zink BJ. Pathophysiology of traumatic brain injury. Mt Sinai J Med 2009 Apr;76(2):97-104. [doi: 10.1002/msj.20104] [Medline: 19306379]
32. Maguire JL, Boutis K, Uleryk EM, Laupacis A, Parkin PC. Should a head-injured child receive a head CT scan? A systematic review of clinical prediction rules. Pediatrics 2009 Jul;124(1):e145-e154. [doi: 10.1542/peds.2009-0075] [Medline: 19564261]
33. Maguire JL, Kulik DM, Laupacis A, Kuppermann N, Uleryk EM, Parkin PC. Clinical prediction rules for children: a systematic review. Pediatrics 2011 Sep;128(3):e666-e677. [doi: 10.1542/peds.2011-0043] [Medline: 21859912]
34. Atabaki SM, Stiell IG, Bazarian JJ, Sadow KE, Vu TT, Camarca MA, et al. A clinical decision rule for cranial computed tomography in minor pediatric head trauma. Arch Pediatr Adolesc Med 2008 May;162(5):439-445. [doi: 10.1001/archpedi.162.5.439] [Medline: 18458190]

XSL•FO
RenderX

35. Osmond MH, Klassen TP, Wells GA, Correll R, Jarvis A, Joubert G, Pediatric Emergency Research Canada (PERC) Head Injury Study Group. CATCH: a clinical decision rule for the use of computed tomography in children with minor head injury. Can Med Assoc J 2010 Mar 9;182(4):341-348 [FREE Full text] [doi: 10.1503/cmaj.091421] [Medline: 20142371]
36. Dunning J, Daly JP, Lomas J, Lecky F, Batchelor J, Mackway-Jones K, Children's Head Injury Algorithm for the Prediction of Important Clinical Events Study Group. Derivation of the children's head injury algorithm for the prediction of important clinical events decision rule for head injury in children. Arch Dis Child 2006 Nov;91(11):885-891 [FREE Full text] [doi: 10.1136/adc.2005.083980] [Medline: 17056862]
37. Kuppermann N, Holmes JF, Dayan PS, Hoyle JD, Atabaki SM, Holubkov R, Pediatric Emergency Care Applied Research Network (PECARN). Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. Lancet 2009 Oct 3;374(9696):1160-1170. [doi: 10.1016/S0140-6736(09)61558-0] [Medline: 19758692]
38. Palchak MJ, Holmes JF, Vance CW, Gelber RE, Schauer BA, Harrison MJ, et al. A decision rule for identifying children at low risk for brain injuries after blunt head trauma. Ann Emerg Med 2003 Oct;42(4):492-506. [doi: 10.1067/s0196-0644(03)00425-6] [Medline: 14520320]
39. Bressan S, Romanato S, Mion T, Zanconato S, da Dalt L. Implementation of adapted PECARN decision rule for children with minor head injury in the pediatric emergency department. Acad Emerg Med 2012 Jul;19(7):801-807 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01384.x] [Medline: 22724450]
40. Atabaki SM, Jacobs BR, Brown KM, Shahzeidi S, Heard-Garris NJ, Chamberlain MB, et al. Quality improvement in pediatric head trauma with PECARN rules implementation as computerized decision support. Pediatr Qual Saf 2017;2(3):e019 [FREE Full text] [doi: 10.1097/pq9.0000000000000019] [Medline: 30229157]
41. Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT head rule for patients with minor head injury. Lancet 2001 May 5;357(9266):1391-1396. [doi: 10.1016/s0140-6736(00)04561-x] [Medline: 11356436]
42. Bezuidenhout AF, Hurter D, Maydell AT, van Niekerk F, de Figueiredo SA, Harvey J, et al. The Kimberley hospital rule (KHR) for urgent computed tomography of the brain in a resource-limited environment. S Afr Med J 2013 Jul 29;103(9):646-651. [doi: 10.7196/samj.6876] [Medline: 24300685]
43. Miller EC, Holmes JF, Derlet RW. Utilizing clinical factors to reduce head CT scan ordering for minor head trauma patients. J Emerg Med 1997;15(4):453-457. [doi: 10.1016/s0736-4679(97)00071-1] [Medline: 9279694]
44. Ibañez J, Arikan F, Pedraza A, Sánchez E, Poca MA, Rodriguez D, et al. Reliability of clinical guidelines in the detection of patients at risk following mild head injury: results of a prospective study. J Neurosurg 2004 May;100(5):825-834. [doi: 10.3171/jns.2004.100.5.0825] [Medline: 15137601]
45. Haydel MJ, Preston CA, Mills TJ, Luber S, Blaudeau E, DeBlieux PM. Indications for computed tomography in patients with minor head injury. N Engl J Med 2000 Jul 13;343(2):100-105. [doi: 10.1056/NEJM200007133430204] [Medline: 10891517]
46. Bouida W, Marghli S, Souissi S, Ksibi H, Methammem M, Haguiga H, et al. Prediction value of the Canadian CT head rule and the New Orleans criteria for positive head CT scan and acute neurosurgical procedures in minor head trauma: a multicenter external validation study. Ann Emerg Med 2013 May;61(5):521-527. [doi: 10.1016/j.annemergmed.2012.07.016] [Medline: 22921164]
47. Papa L, Stiell IG, Clement CM, Pawlowicz A, Wolfram A, Braga C, et al. Performance of the Canadian CT head rule and the New Orleans Criteria for predicting any traumatic intracranial injury on computed tomography in a United States Level I trauma center. Acad Emerg Med 2012 Jan;19(1):2-10 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01247.x] [Medline: 22251188]
48. Smits M, Dippel DW, de Haan GG, Dekker HM, Vos PE, Kool DR, et al. External validation of the Canadian CT head rule and the New Orleans criteria for CT scanning in patients with minor head injury. J Am Med Assoc 2005 Sep 28;294(12):1519-1525. [doi: 10.1001/jama.294.12.1519] [Medline: 16189365]
49. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
50. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11:32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
51. Qualtrics XM: The Leading Experience Management Software. 2018. URL: https://www.qualtrics.com/ [accessed 2020-01-01]
52. Brooke J. SUS-a quick and dirty usability scale. Usab Eval Indus 1996;189(194):4-7. [doi: 10.1201/9781498710411]
53. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
54. del Fiol G, Haug PJ, Cimino JJ, Narus SP, Norlin C, Mitchell JA. Effectiveness of topic-specific infobuttons: a randomized controlled trial. J Am Med Inform Assoc 2008;15(6):752-759 [FREE Full text] [doi: 10.1197/jamia.M2725] [Medline: 18755999]
55. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]

XSL•FO
RenderX

56. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321 [FREE Full text] [doi: 10.1197/jamia.M1717] [Medline: 15684126]
57. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
58. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Abingdon, UK: Routledge; 2013.
59. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
60. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. Int J Hum-Comput Interact 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
61. Bazeley P, Jackson K. Qualitative Data Analysis with NVivo. Thousand Oaks, CA: Sage Publications; 2013.
62. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, GRADE Working Group. GRADE evidence to decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. Br Med J 2016 Jun 28;353:i2016. [doi: 10.1136/bmj.i2016] [Medline: 27353417]
63. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
64. Head BW. Toward more 'evidence-informed' policy making? Public Admin Rev 2015 Nov 20;76(3):472-484. [doi: 10.1111/puar.12475]
65. Barends E, Rousseau DM. Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions. London, UK: Kogan Page Publishers; 2018.
66. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma 2015;5:4 [FREE Full text] [doi: 10.1186/s13336-015-0019-3] [Medline: 25834725]
67. Hunink M, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge, UK: Cambridge University Press; 2014.
68. McCaughey D, Bruning NS. Rationality versus reality: the challenges of evidence-based decision making for health policy makers. Implement Sci 2010 May 26;5:39 [FREE Full text] [doi: 10.1186/1748-5908-5-39] [Medline: 20504357]
69. Bate L, Hutchinson A, Underhill J, Maskrey N. How clinical decisions are made. Br J Clin Pharmacol 2012 Oct;74(4):614-620 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04366.x] [Medline: 22738381]
70. McAlpine K, Lewis KB, Trevena LJ, Stacey D. What is the effectiveness of patient decision aids for cancer-related decisions? A systematic review subanalysis. JCO Clin Cancer Inform 2018 Dec;2:1-13 [FREE Full text] [doi: 10.1200/CCI.17.00148] [Medline: 30652610]
71. Vlemmix F, Warendorf J, Rosman A, Kok M, Mol B, Morris J, et al. Decision aids to improve informed decision-making in pregnancy care: a systematic review. BJOG 2013 Feb;120(3):257-266 [FREE Full text] [doi: 10.1111/1471-0528.12060] [Medline: 23145991]
72. Brownson R, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health 2009;30:175-201. [doi: 10.1146/annurev.publhealth.031308.100134] [Medline: 19296775]
73. Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013 Oct 28;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]
74. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
75. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
76. Frøkjær E, Hertzum M, Hornbæk K. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 2000 Presented at: CHI'00; April 1-6, 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332455]
77. Khajouei R, Wierenga P, Hasman A, Jaspers M. Clinicians satisfaction with CPOE ease of use and effect on clinicians' workflow, efficiency and medication safety. Int J Med Inform 2011 May;80(5):297-309. [doi: 10.1016/j.ijmedinf.2011.02.009] [Medline: 21419695]
78. McLellan S, Muddimer A, Peres SC. The effect of experience on system usability scale ratings. J Usability Stud 2012;7(2):56-67 [FREE Full text]
79. Albert W, Tullis T. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Burlington, Massachusetts: Morgan Kaufmann; 2013.
80. Harrati N, Bouchrika I, Tari A, Ladjailia A. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. Comput Hum Behav 2016 Aug;61:463-471. [doi: 10.1016/j.chb.2016.03.051]
81. Orfanou K, Tselios N, Katsanos C. Perceived usability evaluation of learning management systems: empirical evaluation of the system usability scale. Int Rev Res Open Dist Learn 2015 Apr 15;16(2):-. [doi: 10.19173/irrodl.v16i2.1955]

XSL•FO
RenderX

82. Barnhoorn JS, Haasnoot E, Bocanegra BR, van Steenbergen H. QRTEngine: an easy solution for running online reaction time experiments using qualtrics. Behav Res Methods 2015 Dec;47(4):918-929 [FREE Full text] [doi: 10.3758/s13428-014-0530-7] [Medline: 25407763]

Abbreviations

CCHR: Canadian CT Head Rule
CDS: clinical decision support
ED: emergency department
GRASP: grading and assessment of predictive tools
NOC: New Orleans Criteria
PECARN: Pediatric Emergency Care Applied Research Network
SUS: System Usability Scale
TBI: traumatic brain injury

Edited by G Eysenbach, S Nelson; submitted 04.08.19; peer-reviewed by M Baez, M Afzal, T Tillmann; comments to author 21.01.20; revised version received 05.03.20; accepted 14.05.20; published 09.07.20.

Please cite as:
Khalifa M, Magrabi F, Gallego Luxan B
Evaluating the Impact of the Grading and Assessment of Predictive Tools Framework on Clinicians and Health Care Professionals' Decisions in Selecting Clinical Predictive Tools: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e15770
URL: https://www.jmir.org/2020/7/e15770
doi:10.2196/15770
PMID:32673228

©Mohamed Khalifa, Farah Magrabi, Blanca Gallego Luxan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Paving the Way for the Implementation of a Decision Support System for Antibiotic Prescribing in Primary Care in West Africa: Preimplementation and Co-Design Workshop With Physicians

Nathan Peiffer-Smadja[1,2,3,4], MSc, MD; Armel Poda[5,6], MD, PhD; Abdoul-Salam Ouedraogo[6,7], PharmD, PhD; Jean-Baptiste Guiard-Schmid[8], MD; Tristan Delory[9,10,11], MD, MPH; Josselin Le Bel[9,12], MD, PhD; Elisabeth Bouvet[3,9], MD, PhD; Sylvie Lariven[3,9], MD; Pauline Jeanmougin[9], MD, MSc; Raheelah Ahmad[2,13], PhD; François-Xavier Lescure[1,3,9], MD, PhD

[1]Infection Antimicrobials Modelling Evolution (IAME), UMR 1137, University of Paris, French Institute for Medical Research (INSERM), Paris, France

[2]National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, London, United Kingdom

[3]Infectious Diseases Department, Bichat-Claude Bernard Hospital, Assistance-Publique Hôpitaux de Paris, Paris, France

[4]Sorbonne Université, Paris, France

[5]Department of Infectious Diseases, University Hospital Souro Sanou, Bobo-Dioulasso, Burkina Faso

[6]Institut Supérieur des Sciences de la Santé, Université Nazi Boni, Bobo-Dioulasso, Burkina Faso

[7]Service de Bactériologie Virologie, University Hospital Souro Sanou, Bobo-Dioulasso, Burkina Faso

[8]Initiatives Conseil International - Santé, Ouagadougou, Burkina Faso

[9]Antibioclic, Paris, France

[10]Institut Pierre Louis d'Épidémiologie et de Santé Publique, Sorbonne Université, French Institute for Medical Research (INSERM), Paris, France

[11]Innovation and Clinical Research Unit, Annecy-Genevois Hospital, Épagny Metz-Tessy, France

[12]Department of General Practice, Université Paris Diderot, Université de Paris, Paris, France

[13]School of Health Sciences, City, University of London, London, United Kingdom

Corresponding Author:
Nathan Peiffer-Smadja, MSc, MD
Infection Antimicrobials Modelling Evolution (IAME), UMR 1137
University of Paris
French Institute for Medical Research (INSERM)
UMR 1137
16 rue Henri Huchard
Paris, 75870
France
Phone: 33 618662638
Email: nathan.psmadja@gmail.com

Abstract

Background: Suboptimal use of antibiotics is a driver of antimicrobial resistance (AMR). Clinical decision support systems (CDSS) can assist prescribers with rapid access to up-to-date information. In low- and middle-income countries (LMIC), the introduction of CDSS for antibiotic prescribing could have a measurable impact. However, interventions to implement them are challenging because of cultural and structural constraints, and their adoption and sustainability in routine clinical care are often limited. Preimplementation research is needed to ensure relevant adaptation and fit within the context of primary care in West Africa.

Objective: This study examined the requirements for a CDSS adapted to the context of primary care in West Africa, to analyze the barriers and facilitators of its implementation and adaptation, and to ensure co-designed solutions for its adaptation and sustainable use.

Methods: We organized a workshop in Burkina Faso in June 2019 with 47 health care professionals representing 9 West African countries and 6 medical specialties. The workshop began with a presentation of Antibioclic, a publicly funded CDSS for antibiotic

XSL•FO
RenderX

prescribing in primary care that provides personalized antibiotic recommendations for 37 infectious diseases. Antibioclic is freely available on the web and as a smartphone app (iOS, Android). The presentation was followed by a roundtable discussion and completion of a questionnaire with open-ended questions by participants. Qualitative data were analyzed using thematic analysis.

Results: Most of the participants had access to a smartphone during their clinical consultations (35/47, 74%), but only 49% (23/47) had access to a computer and none used CDSS for antibiotic prescribing. The participants considered that CDSS could have a number of benefits including updating the knowledge of practitioners on antibiotic prescribing, improving clinical care and reducing AMR, encouraging the establishment of national guidelines, and developing surveillance capabilities in primary care. The most frequently mentioned contextual barrier to implementing a CDSS was the potential risk of increasing self-medication in West Africa, where antibiotics can be bought without a prescription. The need for the CDSS to be tailored to the local epidemiology of infectious diseases and AMR was highlighted along with the availability of diagnostic tests and antibiotics using national guidelines where available. Participants endorsed co-design involving all stakeholders, including nurses, midwives, and pharmacists, as central to any introduction of CDSS. A phased approach was suggested by initiating and evaluating CDSS at a pilot site, followed by dissemination using professional networks and social media. The lack of widespread internet access and computers could be circumvented by a mobile app with an offline mode.

Conclusions: Our study provides valuable information for the development and implementation of a CDSS for antibiotic prescribing among primary care prescribers in LMICs and may, in turn, contribute to improving antibiotic use, clinical outcomes and decreasing AMR.

(J Med Internet Res 2020;22(7):e17940) doi:10.2196/17940

KEYWORDS

decision support systems, clinical; antibiotic resistance, microbial; drug resistance, microbial; antibiotic stewardship; implementation science; Africa, Western; diffusion of innovation; medical informatics applications

Introduction

Antimicrobial Resistance

Antimicrobial resistance (AMR) is a global threat that affects both high-income countries (HIC) and low- and middle-income countries (LMIC). The drivers of AMR, including antibiotic misuse, lack of infection prevention and control, and poor sanitation, particularly affect LMICs [1,2]. Data suggest that 80% to 90% of all antibiotics used in humans are prescribed in primary care [3]. Although antibiotic use has been increasing over the last decade in LMICs [2,4], antibiotic prescribing indicators in primary care for the African region deviate significantly from the World Health Organization's (WHO) reference targets [5]. Deviations from existing guidelines on antibiotic use, including delays and overuse, are linked both to patients' behavior and that of health care professionals [6,7]. There is thus a need to find interventions that could improve antibiotic prescribing in primary care to respond to the global threat of AMR.

Electronic Clinical Decision Support Systems

Electronic clinical decision support systems (CDSS) have been devised to provide prescribers with rapid access to updated information, which is required to make appropriate therapeutic decisions [8]. CDSS can be split into CDSS providing unsolicited information (eg, alerts for drug interactions) and CDSS providing solicited information (eg, diagnostic support systems). Computer representations of patient care guidelines are among commonly used CDSS [9,10]. They allow the prescriber to enter basic clinical information of a patient and access the diagnostic or therapeutic guidelines adapted to the patient's unique situation. Machine learning CDSS that find their own decision rules from the data are increasingly being developed and may replace knowledge-based CDSS in which human experts define the decision rules in the future [11]. CDSS adapted in primary care have the potential to reduce overall antibiotic use and improve the appropriateness of antibiotic prescribing [12,13]. In LMICs, the introduction of such aids is in the early stages of development [14], but it could have a measurable impact within the context of sparse infectious diseases specialists. Primary care prescribers in LMICs often lack clinical practice guidelines adapted to their setting and do not have access to antimicrobial stewardship programs [15]. Although they may not have access to health care information systems, CDSS providing diagnostic guidance or easy access to therapeutic guidelines may help them make the right decision at the point-of-care [16]. Improving the adequacy of antibiotic prescribing might, in turn, lead to better clinical outcomes and decreased AMR. Moreover, CDSS have the potential to be implemented in multiple health facilities without recruiting new prescribers [17] and could optimize the utilization of resources that are scarce in LMICs.

Implementation of Clinical Decisions Support System in West Africa

Consequently, the development and implementation of CDSS that fit into clinical work might be part of interventions to mitigate against drivers of AMR in LMICs. However, the adaptation of CDSS for antibiotic prescribing in routine clinical care and their sustainability are often limited [10,18]. As is the case with many innovations in health care, multiple CDSS have been abandoned after their development and initial evaluation because of multiple factors, including the lack of preimplementation work [19]. Indeed, at the global level, the situation analysis for AMR in human health often lack the assessment of technological innovations [20]. Moreover, many CDSS failed to demonstrate a clinical impact because of the low uptake of the CDSS and poor adherence to the generated advice [21,22]. According to the WHO, the adaptation of

XSL•FO
RenderX

electronic CDSS is particularly low in LMICs [23]. Interventions to implement CDSS are challenging because of cultural and structural constraints, and these challenges are even more pronounced in primary care where prescribing can be done by a variety of health care professionals or when resources are limited [14,24]. West Africa hosts about 381 million people and includes 16 countries, among them some of the lowest income countries in the world [25,26]. French-speaking West African countries, including Benin, Burkina Faso, Cameroon, Chad, Ivory Coast, Republic of Congo, Guinea, Mali, Niger, Senegal, and Togo, were among the LMICs with the highest increase in global antibiotic consumption between 2000 and 2015 [2]. As highlighted by the WHO, countries in the African region are the least advanced regarding electronic and mobile health [23].

Objectives

Before developing or implementing a CDSS for antibiotic prescribing in West Africa, we need to understand the contextual, structural, and behavioral determinants that will facilitate or prevent the use of CDSS for antibiotic prescribing. Our study aimed to examine the requirements for a CDSS adapted to the context of primary care in West Africa and to analyze the barriers and facilitators of its implementation and adaptation by prescribers using qualitative methods during a preimplementation workshop.

Methods

Study Design

Our aims were (1) to understand the potential benefits and risks of a CDSS for antibiotic prescribing in primary care in West Africa, (2) to analyze the barriers and facilitators of its implementation, and (3) to ensure co-designed solutions for its adaptation and sustainable use. To this end, we organized a workshop with West African antibiotic prescribers that began with a presentation of Antibioclic, a CDSS for antibiotic prescribing currently widely used in France and freely available on the web and as a smartphone app (further details below). We chose to begin the workshop with an example of an existing CDSS rather than with a theoretical description of a CDSS for antibiotic prescribing. This strategy was considered by the organizing team as the most promising to elicit reactions among the participants. The main characteristics of a CDSS were presented, and the organizers went through a number of clinical scenarios on infectious diseases using Antibioclic. Then, the organizers moderated a roundtable discussion with all the participants. During this first-of-its-kind workshop, the participants generated barriers and facilitators to the use of a CDSS such as Antibioclic in West Africa and were asked to hypothesize if its use could lead to benefits or risks, including unexpected events. Participants were specifically asked to debate the challenges in the implementation and adaptation of a CDSS for antibiotic prescribing in West Africa and to co-design solutions.

Antibioclic—An Example of a Clinical Decision Support System for Antibiotic Prescribing

Antibioclic is a French CDSS for antibiotic prescribing in primary care, [27] targeting 37 common infectious diseases, freely available on the web and as a smartphone app on iOS and Android. It was codeveloped by general practitioners, specialists of infectious diseases, and engineers in 2011. Clinicians can enter the diagnosis of a patient on the website or app, they are then asked a few targeted questions (age group, comorbidities, renal function, breastfeeding, and pregnancy). Finally, they receive a tailored recommendation of antibiotic regimen, dose, and duration according to French national guidelines. Multimedia Appendices 1 and 2 describe how to use Antibioclic respectively with slides or video in English. Antibioclic was developed using a systematic method to transform clinical practice guidelines from the French National Authority for Health and the French Infectious Diseases Society (SPILF) into decision trees in Antibioclic. Antibioclic is updated frequently and modified as soon as a new guideline is published. The number of Antibioclic users in France has steadily increased over the past years from a median (IQR) of 414 (245-494) a day in 2012 to 5365 (2891-5769) a day in 2018, without any saturation to date (Multimedia Appendix 3). The smartphone app has been downloaded 22,970 times on Android and 15,200 times on iOS. More details are described in a recent study [27].

Workshop Location and Participants

The 3-hour workshop to discuss the implementation and adaptation of a CDSS for antibiotic prescribing in West Africa was organized in Burkina Faso by 2 professors of infectious diseases, 1 from Burkina Faso (AP) and 1 from France (FL). To maximize participation and cost efficiency, the workshop took place during the French-speaking university course *Antibiologie et Antibiothérapie en Afrique Subsaharienne* (*Antibiotic therapy in Sub-Saharan Africa*) organized in Bobo-Dioulasso, Burkina Faso, on June 2019. This course lasts 5 weeks and is organized every year by the University Nazi Boni of Bobo-Dioulasso and the University of Montpellier. The course is open to West African health professionals working on AMR or antibiotic therapy, whether in primary, secondary, or tertiary care, with no selection among participants. Attendance to the workshop was voluntary for the participants of the course.

Electronic Questionnaire

After the roundtable discussion, an electronic questionnaire with closed- and open-ended questions was given to each participant (Multimedia Appendix 4). The questionnaire allowed the collection of in-depth individual data on the implementation and adaptation of a CDSS for antibiotic prescribing. The questionnaire was codeveloped by a multidisciplinary team including 6 members of the Antibioclic study team: infectious diseases clinicians (AP and FL), a microbiologist (AO), public health specialists (JB and GS) with experience in the West African context, and an implementation science and knowledge mobilization lecturer (RA). The first part collected demographic characteristics and information regarding the use of electronic tools for antibiotic prescribing. The second part displayed open-ended questions about potential outcomes and users of a

XSL•FO
RenderX

CDSS for antibiotic prescribing in West Africa and the challenges of its implementation and adaptation.

Data Analysis

Data from the closed-ended questions of the questionnaire were imported into R software version 3.2.4 (R Foundation for Statistical Computing). Numerical data are presented as absolute numbers, proportions, and median (IQR). The entire roundtable discussion was audio recorded and transcribed verbatim by the AMK France professional company. Data from the roundtable discussion and open-ended questions were analyzed using thematic analysis and were coded to identify key categories, which were developed into themes using NVivo 12 software (QSR International). The analysis of questionnaires and the roundtable discussion supported cross-validation and triangulation of the findings.

Ethical Considerations

Written consent was obtained from each participant. No distinguishable personal information was recorded, and all the data were analyzed anonymously. This survey was approved by the research ethics committee of Centre Muraz (Health Research Institute, Bobo-Dioulasso, Burkina Faso) and was compliant with the European General Data Protection Regulation.

Results

Participants

All 47 participants of the course, 19 women and 38 men with a median (IQR) age of 31 (30-38) years, participated in the workshop and completed the questionnaire (Table 1). Most of the participants (35/47, 74%) were from Burkina Faso, but 8 other West African countries were represented. Approximately half of the participants (24/47, 51%) worked in a university hospital, whereas the others worked in general hospitals (14/47, 30%), public health institutes (4/47, 9%), dispensary (1/47, 2%), private hospital (1/47, 2%), or pharmacies (1/47, 2%). The low number of participants working as primary care prescribers may be related to a more difficult access to university degrees. The specialty of the participants was general practice (21/47, 45%), microbiology (13/47, 28%), pharmacy (4/47, 9%), anesthesiology and intensive care (3/47, 6%), infectious diseases (2/47, 4%), and neurosurgery (2/47, 4%). Only 26% (12/47) of participants, all in Burkina Faso, stated that they used national guidelines to prescribe antibiotics. The others used French guidelines (20/47, 43%), WHO guidelines (20/47, 43%), or local hospital guidelines (4/47, 9%).

XSL•FO
RenderX

Table 1. Demographic characteristics of participants (N=47).

Characteristics	Values
Age (years), median (IQR)	31 (30-38)
Gender, n (%)	
Men	28 (60)
Women	19 (40)
Country of practice, n (%)	
Burkina Faso	35 (74)
Togo	3 (6)
Senegal	2 (4)
Mali	2 (4)
Gabon[a]	1 (2)
Guinea	1 (2)
Guinea-Bissau	1 (2)
Ivory Coast	1 (2)
Niger	1 (2)
Working structure, n (%)	
University hospital	24 (51)
General hospital	14 (30)
Public health institute	4 (9)
Dispensary	1 (2)
Private hospital	1 (2)
Pharmacy	1 (2)
Medical specialty, n (%)	
General practice	21 (45)
Microbiology	13 (28)
Pharmacist	4 (9)
Anesthesiology and intensive care	3 (6)
Infectious diseases	2 (4)
Neurosurgery	2 (4)
Guidelines used for clinical practice, n (%)	
French	20 (43)
World Health Organization	20 (43)
National[b]	12 (26)
Hospital	4 (9)
American	2 (4)
European	2 (4)
Portuguese	1 (2)

[a]Central Africa.

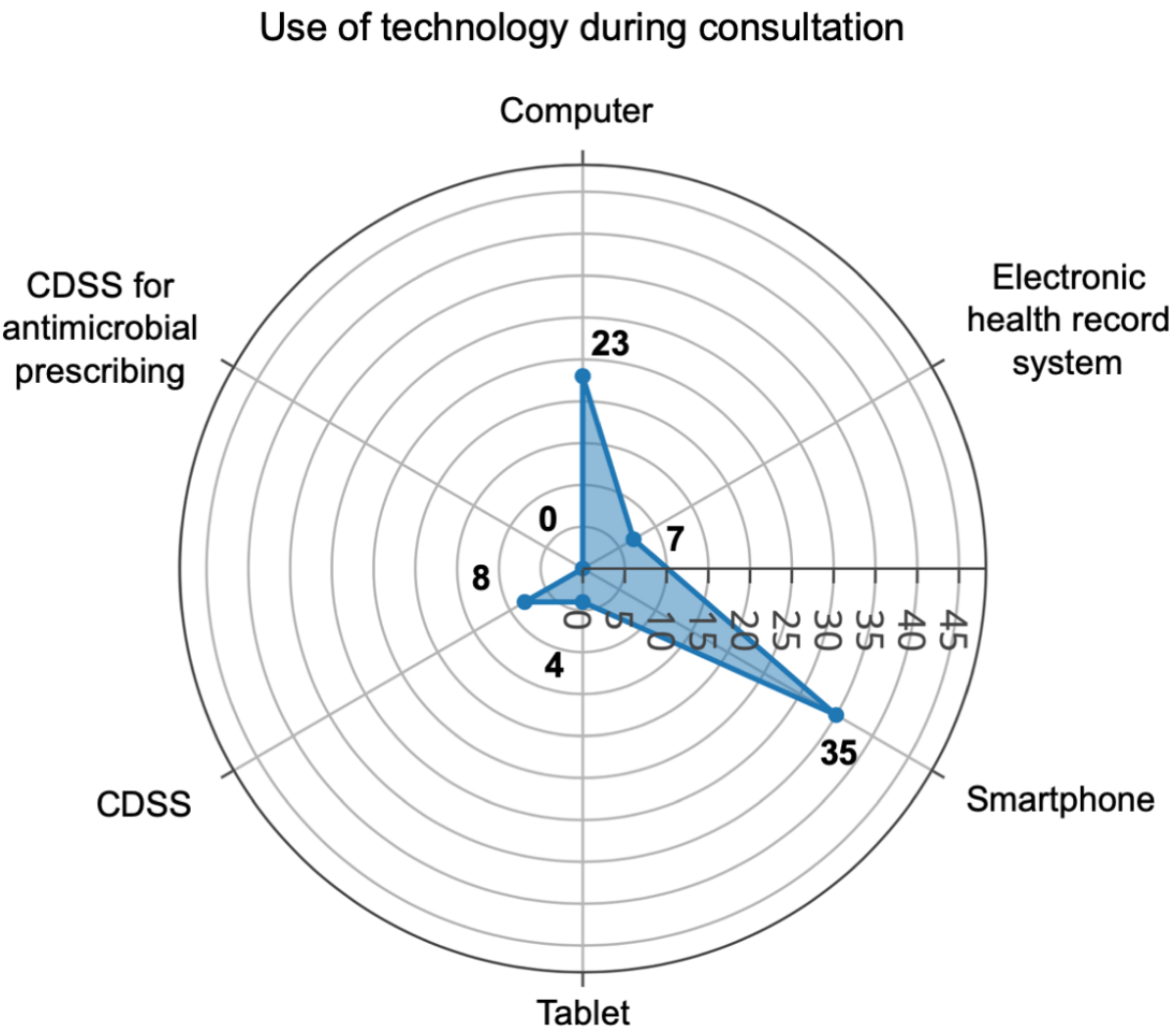
[b]The 12 participants were from Burkina Faso.

XSL•FO
RenderX

Current Use of Technology and Clinical Decision Support Systems

Most of the participants (35/47, 74%) had a smartphone with internet access during their clinical consultations, but less than half (23/47, 49%) of the participants had access to a computer, and 15% (7/47) of the participants had no access to any electronic tools during their consultations. Most of the clinicians did not use an electronic health record system (39/47, 83%) or CDSS (36/47, 77%) for any clinical decision making during their clinical consultations. The only CDSS mentioned by the participants was a tool for basic medical calculations and a tool that displays medical cards. No CDSS for antibiotic prescribing (nor for any antimicrobial drug) was used by the participants (Figure 1).

Figure 1. Current use of technology during consultation among participants. CDSS: clinical decision support systems.

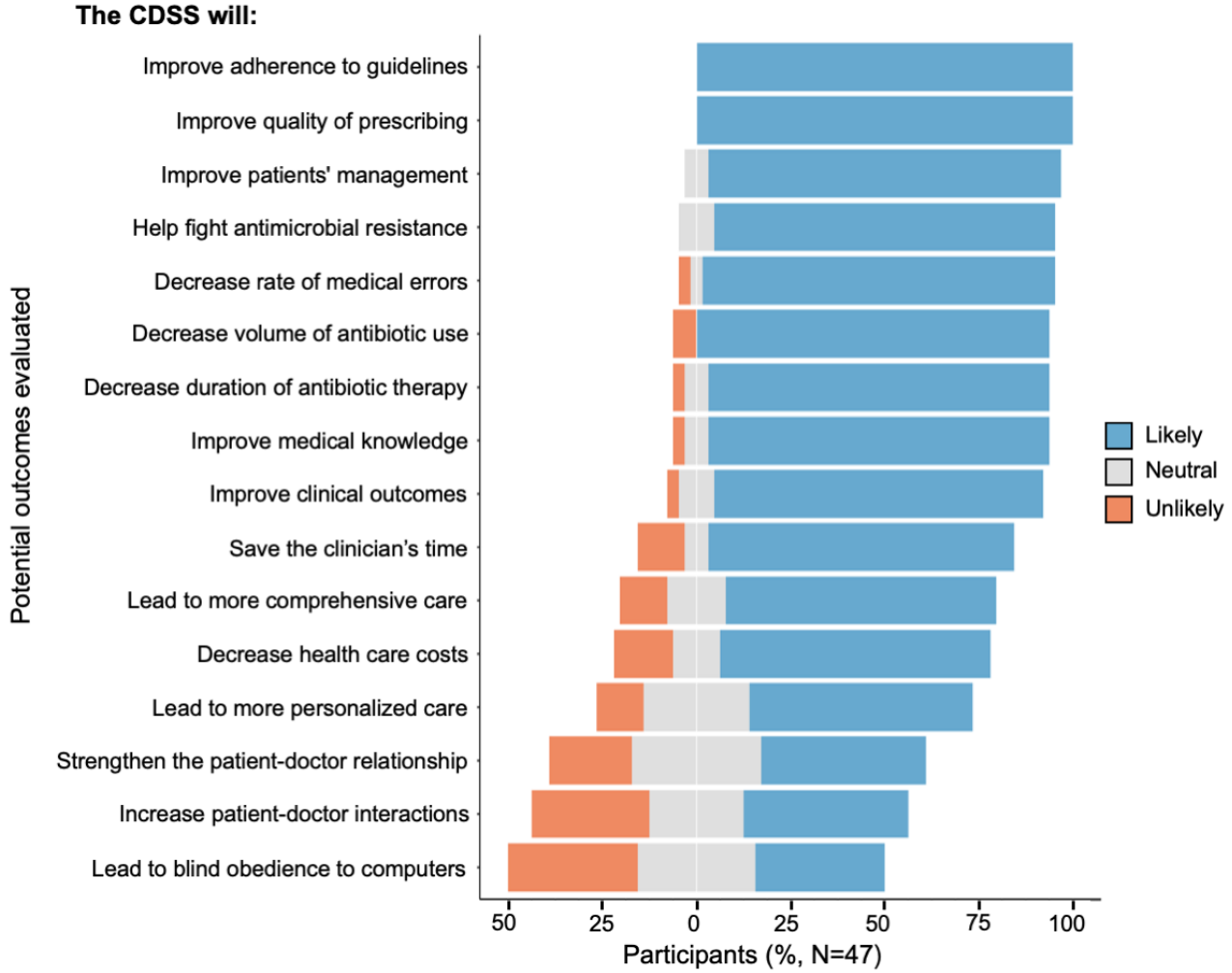


Potential of a Clinical Decision Support Systems for Antibiotic Prescribing

Most of the participants expected favorable outcomes from the implementation of a CDSS for antibiotic prescribing in primary care in their countries (Figure 2). They thought a CDSS was likely to improve the quality of antibiotic prescribing (47/47, 100%), improve adherence to guidelines (47/47, 100%), improve the management and care of patients (44/47, 94%), decrease the number of medical errors (44/47, 94%), improve the volume of antibiotic use (44/47, 94%), improve the duration of antibiotic therapy (43/47, 91%), fight against AMR (43/47, 91%), increase the medical knowledge of users (43/47, 91%), and save time during consultation (38/47, 81%). Participants disagreed on the fact that such a tool could strengthen the physician-patient relationship (20/47, 43% judged it likely; 11/47, 23% unlikely; and 16/47, 34% were neutral) or could lead to blind obedience to electronic tools from prescribers (16/47, 34% judged it likely; 16/47, 34% unlikely; and 15/47, 32% were neutral).

XSL•FO
RenderX

Figure 2. Expected outcomes of a clinical decision support system for antibiotic prescribing. CDSS: clinical decision support systems.



During the roundtable discussion and through the open-ended questions, the participants identified a number of consequences of a CDSS on a wider scale. Participants underlined the lack of antimicrobial stewardship teams in West Africa. They hypothesized that a CDSS could help many isolated structures to have access to antibiotic prescribing recommendations and could relieve existing antimicrobial stewardship teams of simple antibiotic prescribing advice and let them focus on complex cases. One participant noted that a CDSS could be used to increase surveillance capability in primary care. Indeed, its use could be monitored and lead to reports on the incidence of infectious diseases and the rate of antibiotic prescriptions. The participants agreed that this is particularly important in West African countries where epidemiological data are scarce. The use of French guidelines was not seen as a barrier to a CDSS. Indeed, most of the participants declared that they did not have national guidelines to prescribe antibiotics in their country. However, the participants pointed out that a successful CDSS could have a positive impact in encouraging national societies to develop guidelines that could then be integrated into the CDSS. More broadly, they agreed that this CDSS could help to increase the awareness of health professionals on the importance of appropriate prescribing and the risks of AMR. The participants also agreed that a CDSS could be used in the initial training and continuous medical education of health professionals by updating antibiotic prescribing knowledge:

> *If you use it continually, you certainly update your knowledge.* [General practitioner in Burkina Faso]

Regarding the scope of a CDSS for antibiotic prescribing, the participants stated that it could have a positive impact on the management of urinary tract infections (84%); skin and soft tissue infections (84%); gastrointestinal tract infections (83%); genital infections (80%); pneumonia (77%); meningitis (70%); ear, nose, and throat infections (70%); dental infections (67%); and surgical site infections (61%). The participants identified a wide diversity of potential users of a CDSS for antibiotic prescribing in West Africa, including general practitioners, infectious diseases specialists, physicians in other specialties, pharmacists, nurses, clinical microbiologists, and midwives.

Challenges

The participants identified a number of challenges and proposed a number of solutions for the adaptation of a CDSS for antibiotic prescribing in a West African setting (Table 2), and these are grouped by system level.

XSL•FO
RenderX

Table 2. Challenges and potential solutions for the development and implementation of a clinical decision support system in the West African context.

Level	Challenges to CDSS[a] development	Challenges to CDSS implementation	Potential solutions
Country level	• Scarce epidemiological data on the prevalence and incidence of infectious diseases and the level of antimicrobial resistance	• N/A[b]	• Encouraging studies to better analyze the local and regional epidemiology • Developing and updating the CDSS according to local and regional epidemiology regarding infectious diseases, microbiology, and antimicrobial resistance • Including tuberculosis and common parasitic diseases
	• Lack of national guidelines	• N/A	• The CDSS should follow local, regional, and national guidelines where they exist. If they do not, the CDSS could follow French or WHO[c] guidelines as they are used by most participants • The CDSS should be developed for the subcontinent of West Africa and then could be further adapted to each country • To easily adapt the CDSS to local and national guidelines, the programming and code of the CDSS should be in open access
	• Limited availability of diagnostic tests and antibiotics	• N/A	• Adapting the suggestions to locally available diagnostic tests and antibiotics by working with national scientific societies
Health care structure level	• N/A	• Lack of internet access and information technology infrastructure	• Development of an offline mode of the CDSS • Development of a mobile version on iOS and Android • Increasing the availability of computers and internet access in West Africa
	• N/A	• Independently operating and geographically isolated health structures such as dispensaries	• Pilot testing of the CDSS in a primary care structure linked to an academic hospital before disseminating the tool to other health structures • Field communication with primary care prescribers • Using the network of the Ministry of Health
Individual level: clinicians and patients	• Diversity of training needs for primary care prescribers	• N/A	• Co-designing the CDSS with general practitioners, nurses, midwives, microbiologists, dentists, and pharmacists • Allowing for different modules of access for health professionals to meet the different information needs
	• N/A	• Lack of awareness and training about CDSS	• Dedicated training for primary care prescribers Communicating through scientific and professional societies, using traditional and social media • Involving all the stakeholders, including health authorities, the Ministry of Health, and the media
	• N/A	• Risk of increasing self-treatment with antibiotics as they are available without prescription in most West African countries	• Limiting access to registered health professionals (disagreement between participants) • Regulating access to antibiotics without prescription
	• N/A	• Risk of deskilling and dependency of prescribers	• Improving the training of prescribers about antibiotic prescribing

XSL•FO
RenderX

Level	Challenges to CDSS[a] development	Challenges to CDSS implementation	Potential solutions
	• N/A	• Risk to lose patients' confidence by following the advice of an electronic tool	• Education of patients about the need to use reference books or electronic sources to provide the best care • Ensuring the independence of the tool from pharmaceutical companies

[a]CDSS: clinical decisions support system.

[b]N/A: not applicable.

[c]WHO: World Health Organization.

Country Level

At the country level, the participants pointed out the scarcity of epidemiological data in West African countries, the difference between epidemiology of diseases in HICs and LMICs, the lack of national guidelines, and the limited availability of diagnostic tests and antibiotics. To produce computer-interpretable clinical guidelines, CDSS developers need to have access to clinical practice guidelines. For countries where national guidelines for infectious diseases management do not exist, the participants first suggested the development of a CDSS based on French or WHO guidelines to encourage the establishment of national guidelines. To address the specific needs of primary care prescribers in West Africa, the CDSS should not only target the management of common bacterial diseases but also the management of parasitic diseases and tuberculosis. Participants stressed the importance of tailoring the CDSS to local capacity and infrastructure, such as resource availability and antibiotics, with the help of national scientific societies.

Health Care Structure Level

The first challenge at the organization level was the lack of internet access and information technology (IT) infrastructure. Taking into account the wide availability of smartphones with internet access, the participants highlighted major opportunities through the development of mobile phone versions of CDSS (on iOS and Android) with an offline mode. The second challenge was the important number of independently operating and geographically isolated health structures, such as dispensaries, in the West African health system. To reach these structures, participants suggested the involvement of relevant stakeholders, including health authorities, the Ministry of Health, scientific societies, the media, and opinion leaders, using traditional and social media:

> *It is important to use the traditional networks of the Ministry of Health where information goes down hierarchically but also social media that are more and more used.* [Physician in a university hospital, Burkina Faso]

The organization of a phased approach was suggested by initiating and evaluating CDSS in a pilot primary care structure linked to an academic hospital followed by dissemination to other health structures:

> *You have to show people that it's used by people like them starting with a pilot test site.* [Physician in a general hospital, Burkina Faso]

Individual Level: Clinicians and Patients

The diversity of primary care prescribers who have different levels of decision support needs was identified as a challenge to developing effective CDSS. A physician working in a general hospital stated that in West Africa "nurses see a lot of patients, certainly more than physicians" highlighting the need to codevelop the CDSS with primary care health care professionals who frequently prescribe antibiotics, such as general practitioners, nurses, midwives, microbiologists, dentists, and pharmacists. The implementation of a CDSS in West Africa should take into account the lack of widespread knowledge on the use of CDSS by organizing training sessions with primary care prescribers. One participant said that some health care professionals or patients could be refractory to new technologies, but this was not considered a major barrier by the other participants. The participants identified a number of facilitators for the adaptation of a CDSS by primary care prescribers and patients, including the independence of the tool from pharmaceutical companies, codevelopment with users, free access to the CDSS, and good usability.

The most discussed unexpected consequence was that an open access CDSS such as Antibioclic could amplify self-medication of antibiotics due to availability without prescription:

> *My point of view is really mixed. This represents an undeniable progress for antibiotic prescribing but in the context of our country, there is a non-negligible risk of self-medication by non-healthcare professionals.* [General practitioner, Burkina Faso]

> *A very good tool for the treatment of infections but which could be dangerous in the southern countries because of self-medication and the accessibility of antibiotics without prescription.* [Physician in a university hospital, Burkina Faso]

In this context, the CDSS may be used as a substitute for expert health care opinion. However, some participants highlighted the fact that much information is already available on the web and that Antibioclic is already available, including in Burkina Faso, with no reported use by patients. In fact, there was uncertainty regarding if implementing a CDSS in West Africa would amplify the phenomenon of self-medication or would help by providing a reliable source to patients who would want to self-medicate anyway:

> *Some patients will buy antibiotics without prescription, with or without the tool.* [General practitioner, Burkina Faso]

XSL•FO
RenderX

> *On the contrary, if patients have access to national recommendations, this reinforces and this even facilitates your message.* [Professor in a university hospital, Burkina Faso]

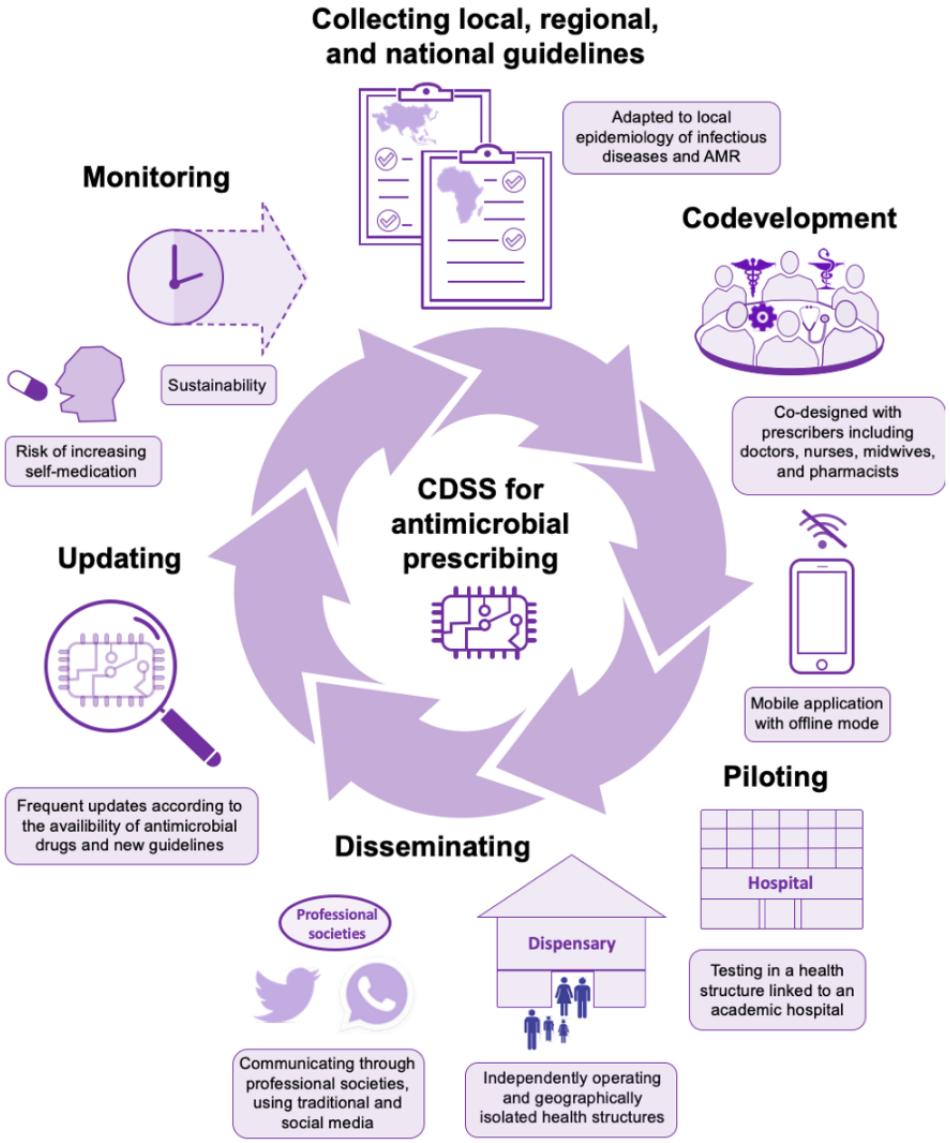
A consensus emerged that the best solution is to increase efforts to limit access to antibiotics without prescription.

Discussion

Principal Findings

We conducted a preimplementation workshop with West African health professionals to examine the requirements for a CDSS adapted to the context of primary care in their countries. The use of CDSS for antibiotic prescribing (and other clinical decisions, in fact) is currently missing, but the importance and need for such tools to support antibiotic prescribing in primary care were stressed by all participants. Regarding the implementation of a CDSS in primary care, the participants encouraged a procedure of co-designing the tool with health professionals and stakeholders involved in antibiotic prescribing and initiation via a primary care pilot site, which is linked to an academic hospital (Figure 3). The diffusion of the CDSS would be more successful if supported by networks of medical specialists and the Ministry of Health but should also take advantage of social media because of its high use among health care professionals. The lack of widespread internet access and computers during consultation was a barrier to the use of the CDSS, but it could be circumvented by the development of a mobile app with an offline mode. The CDSS should be tailored to the local epidemiology of infectious diseases and AMR, and the high prevalence of malaria, tuberculosis, and HIV in West Africa should be taken into account when developing the CDSS [28]. The availability of diagnostic tests and antibiotics may differ between West African countries and may vary over time, requiring frequent updates of the CDSS.

Figure 3. Key steps for the development and implementation of a clinical decision support systems for antibiotic prescription in low- and middle-income countries. AMR: antimicrobial resistance; CDSS: clinical decision support systems.



The participants pointed out the potential risk of self-medication in countries in which antibiotics can be bought with no prescription. Such a risk should certainly be anticipated and monitored, but most of the participants did not think that the

XSL•FO
RenderX

CDSS should have restricted access. On the contrary, some participants thought that a CDSS could provide an opportunity to strengthen the education of the general public about antibiotic prescribing and AMR. Development and implementation of a CDSS was seen to be associated with a number of benefits related to improving clinical care and reducing AMR and also to strengthening and updating practitioners' knowledge of antibiotic prescribing recommendations, encouraging the development of national guidelines, and developing surveillance capabilities in primary care. Indeed, the implementation of a CDSS for antibiotic prescribing could be coupled with measures that allow an increase in the capability of surveillance of antibiotic use. Examples include real-time monitoring of requests expressed on the platform and extracting reports about antibiotic use [27,29]. Moreover, the implementation of a CDSS in the community could have implications beyond antibiotic prescribing in primary care. Indeed, primary care is the gateway to secondary and tertiary care, and AMR in primary care circulates in the whole health system [30].

Comparison With Prior Work

The lack of preimplementation and co-design work concerning the assessment of CDSS for antibiotic prescribing has been reported in recent reviews [10,31]. More broadly, research on the implementation strategies for CDSS in primary care is lacking [32]. A qualitative design involving 33 face-to-face interviews with general practitioners has been used to codevelop computer-delivered interventions to promote guidelines for antibiotic prescribing in the United Kingdom. The authors focused on the technical development of prompts and the determinants of their acceptability but did not study the implementation of the CDSS in clinical practice [33]. Postimplementation research found that the key factor for the use of this CDSS by general practitioners in primary care was their awareness of the implementation of the system within their practice [22]. The lack of end-user engagement with implementation and the rigidity of the guidelines incorporated in the prompts were barriers to the effective adaptation of the system, stressing the importance of a systematic and theoretically informed process of implementation for CDSS [34]. The nonadaptation, abandonment, scale-up, spread, and sustainability framework has been developed to support the implementation of technological innovations in health and social care [19]. It lists challenges in 7 domains: the condition or illness, the technology, the value proposition, the adopter system (comprising professional staff, patient, and lay caregivers), the organization(s), the wider (institutional and societal) context, and the interaction and mutual adaptation between all these domains over time.

Few studies have focused on CDSS for the management of infectious diseases in LMICs and in particular in West Africa. A recent review identified 6 CDSS for the management of febrile children in primary care in LMICs [35]. In a pilot cluster-randomized controlled study, the clinical algorithm for management of childhood illness available on smartphones (e-ALMANACH) increased the number of children screened for red flags and decreased antibiotic prescription in Tanzania [36]. This result was confirmed in another controlled study in which the use of e-ALMANACH to manage children improved clinical outcomes and reduced antibiotic prescription by 80% [37]. Electronic point-of-care tool, an electronic algorithm using point-of-care testing results including C-reactive protein and procalcitonin, was shown to improve the clinical outcomes of children with febrile illnesses as compared with the e-ALMANACH CDSS [38,39]. Another study described a protocol for the implementation on a large scale in Burkina Faso of an electronic tool for integrated management of childhood illnesses [40]. The results of this trial have not been published yet. In general, few data are available on the adaptation and sustainability of such interventions. One team analyzed the implementation cost of a CDSS for antenatal and delivery care and found that most of the cost resulted from the recruitment and training of nurses and midwives using the CDSS and from buying laptops [41]. A CDSS that is easily used by health care professionals and is available as a free mobile app could lead to a decreased implementation cost. A study in Burkina Faso confirmed a positive attitude toward electronic decision support systems and the need for simplicity and good usability of the tool. They found similar barriers, such as the lack of infrastructure and IT systems and the limited availability of drugs and diagnostic tests [42]. The development of CDSS in LMICs could be facilitated by the development of mobile health [43] and by sharing open-source codes [44]; however, other challenges remain: countries often have to buy a back-office or another hosting solution and have to pay for the maintenance of the server [45,46]. These costs, if not anticipated, can limit the sustainability of CDSS.

Limitations

The participants stressed the diversity of prescribers and health structures in West Africa that must be taken into account when codeveloping the CDSS. However, most of the participants in our workshop were physicians specialized in general practice working in Burkina Faso university hospitals. This bias was related to our recruitment of participants via a university degree organized by French and Burkina Faso universities. We, thus, need to involve more primary care prescribers such as nurse prescribers, midwives, and physicians working in dispensaries to allow for a broader view of primary care antibiotic prescribing. Indeed, prescribing habits may vary between professional groups [47,48], and we would like to capture potential differences among prescribers concerning the use, needs, and expectations regarding CDSS for antibiotic prescribing.

To open the discussion, we decided to provide an example of a currently used CDSS. Indeed, as we anticipated that most of the participants had no experience using electronic decision support systems for antibiotic prescribing, we wanted to show them an existing CDSS and collect opinions. This approach was likely to influence the answers of participants and the discussion during the roundtable discussions. However, we think that the barriers and facilitators identified in the study are not limited to Antibioclic. Indeed, the contextual barriers that have been found have been described in other studies. Nonprescription dispensing of antibiotics in the community has been widely described both in LMICs [49,50] and in HICs [51]. In a recent review, the authors did not find any clinical practice guideline applicable to prehospital care in LMICs among 276 guidelines

XSL•FO
RenderX

[15]. For countries where national guidelines for the management of infectious diseases do not exist, the participants suggested the development of a CDSS first based on French or WHO guidelines to encourage the establishment of national guidelines.

The governance system and public health priorities could be important additional factors that need to be studied in West Africa before the implementation of a CDSS. Work has been done in European countries about governance to address AMR [52], and relevant frameworks for assessing this important factor exist [20,53]. However, the analysis and governance frameworks for antibiotic prescribing often lack the assessment of technological innovation, such as electronic decision support. Including CDSS in the national action plan as in the United Kingdom's 2019-2024 national plan that targets to "be able to report on the percentage of prescriptions supported by a diagnostic test or decision support tool by 2024" is probably an important step toward the development and implementation of support systems [54]. We are not aware of Western African national action plans mentioning the development and use of CDSS.

Conclusions

Preimplementation research is needed to ensure that CDSS are adapted to the context in which they are deployed. Our study provided valuable information to develop and implement a CDSS for antibiotic prescribing among primary care prescribers in LMICs. We plan to organize workshops to co-design a CDSS tailored to the context of primary care in West Africa. Optimizing the use of antibiotics in primary care may have beneficial consequences for the entire health system and can contribute to limiting the spread of AMR. Most of the barriers and facilitators that we identified may easily relate to a broad spectrum of CDSS, including systems for clinical decisions other than antibiotic prescribing.

Acknowledgments

The authors thank all the participants in the workshop and the team of the university course *Antibiologie et Antibiothérapie en Afrique Subsaharienne* organized by the University Nazi Boni of Bobo-Dioulasso and the University of Montpellier and supported by the WHO, West African Health Organization, and SPILF. RA acknowledges the support of the National Institute for Health Research Fellowship in knowledge mobilization at the National Institute of Health Research Health Protection Research Unit in Healthcare-Associated Infection and Antimicrobial Resistance at Imperial College and of the Economic and Social Research Council as part of the Antimicrobial Cross Council initiative supported by the 7 UK research councils, and the authors also acknowledge the support of the Global Challenges Research Fund. This study was funded by a PhD studentship from I^le-de-France Regional Health Agency, Paris, France, granted to NP and by a grant from the French Association de Chimiothérapie Anti-Infectieuse via SPILF awarded to NP.

Conflicts of Interest

None declared.

Multimedia Appendix 1
How to use Antibioclic (slides).
[PPTX File , 31628 KB - jmir_v22i7e17940_app1.pptx]

Multimedia Appendix 2
How to use Antibioclic (video).
[MOV File , 201584 KB - jmir_v22i7e17940_app2.mov]

Multimedia Appendix 3
Antibioclic users.
[PPTX File , 282 KB - jmir_v22i7e17940_app3.pptx]

Multimedia Appendix 4
Electronic questionnaire.
[DOCX File , 75 KB - jmir_v22i7e17940_app4.docx]

References

1. World Health Organization. Antimicrobial Resistance: Global Report on Surveillance. Geneva, Switzerland: World Health Organization; 2014.
2. Klein EY, van Boeckel TP, Martinez EM, Pant S, Gandra S, Levin SA, et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. Proc Natl Acad Sci U S A 2018 Apr 10;115(15):E3463-E3470 [FREE Full text] [doi: 10.1073/pnas.1717295115] [Medline: 29581252]

XSL•FO
RenderX

3. Goossens H, Ferech M, Stichele RV, Elseviers M, ESAC Project Group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet 2005;365(9459):579-587. [doi: 10.1016/S0140-6736(05)17907-0] [Medline: 15708101]
4. Farooqui HH, Selvaraj S, Mehta A, Heymann DL. Community level antibiotic utilization in India and its comparison vis-à-vis European countries: evidence from pharmaceutical sales data. PLoS One 2018;13(10):e0204805 [FREE Full text] [doi: 10.1371/journal.pone.0204805] [Medline: 30332450]
5. Ofori-Asenso R, Brhlikova P, Pollock AM. Prescribing indicators at primary health care centers within the WHO African region: a systematic analysis (1995-2015). BMC Public Health 2016 Aug 22;16(1):724 [FREE Full text] [doi: 10.1186/s12889-016-3428-8] [Medline: 27545670]
6. Carrara E, Pfeffer I, Zusman O, Leibovici L, Paul M. Determinants of inappropriate empirical antibiotic treatment: systematic review and meta-analysis. Int J Antimicrob Agents 2018 Apr;51(4):548-553. [doi: 10.1016/j.ijantimicag.2017.12.013] [Medline: 29277528]
7. Limmathurotsakul D, Sandoe JA, Barrett DC, Corley M, Hsu LY, Mendelson M, et al. 'Antibiotic footprint' as a communication tool to aid reduction of antibiotic consumption. J Antimicrob Chemother 2019 Aug 1;74(8):2122-2127 [FREE Full text] [doi: 10.1093/jac/dkz185] [Medline: 31074489]
8. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. J Am Med Assoc 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
9. Durieux P, Nizard R, Ravaud P, Mounier N, Lepage E. A clinical decision support system for prevention of venous thromboembolism: effect on physician behavior. J Am Med Assoc 2000 Jun 7;283(21):2816-2821. [doi: 10.1001/jama.283.21.2816] [Medline: 10838650]
10. Rawson T, Moore L, Hernandez B, Charani E, Castro-Sanchez E, Herrero P, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? Clin Microbiol Infect 2017 Aug;23(8):524-532 [FREE Full text] [doi: 10.1016/j.cmi.2017.02.028] [Medline: 28268133]
11. Peiffer-Smadja N, Rawson T, Ahmad R, Buchard A, Pantelis G, Lescure F, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2019 Sep 17:- epub ahead of print. [doi: 10.1016/j.cmi.2019.09.009] [Medline: 31539636]
12. Samore MH, Bateman K, Alder SC, Hannah E, Donnelly S, Stoddard GJ, et al. Clinical decision support and appropriateness of antimicrobial prescribing: a randomized trial. J Am Med Assoc 2005 Nov 9;294(18):2305-2314. [doi: 10.1001/jama.294.18.2305] [Medline: 16278358]
13. Holstiege J, Mathes T, Pieper D. Effects of computer-aided clinical decision support systems in improving antibiotic prescribing by primary care providers: a systematic review. J Am Med Inform Assoc 2015 Jan;22(1):236-242 [FREE Full text] [doi: 10.1136/amiajnl-2014-002886] [Medline: 25125688]
14. Cox J, Vlieghe E, Mendelson M, Wertheim H, Ndegwa L, Villegas M, et al. Antibiotic stewardship in low- and middle-income countries: the same but different? Clin Microbiol Infect 2017 Nov;23(11):812-818 [FREE Full text] [doi: 10.1016/j.cmi.2017.07.010] [Medline: 28712667]
15. McCaul M, Clarke M, Bruijns SR, Hodkinson PW, de Waal B, Pigoga J, et al. Global emergency care clinical practice guidelines: a landscape analysis. Afr J Emerg Med 2018 Dec;8(4):158-163 [FREE Full text] [doi: 10.1016/j.afjem.2018.09.002] [Medline: 30534521]
16. Laka M, Milazzo A, Merlin T. Can evidence-based decision support tools transform antibiotic management? A systematic review and meta-analyses. J Antimicrob Chemother 2020 Jan 20:- epub ahead of print. [doi: 10.1093/jac/dkz543] [Medline: 31960021]
17. Mainous AG, Lambourne CA, Nietert PJ. Impact of a clinical decision support system on antibiotic prescribing for acute respiratory infections in primary care: quasi-experimental trial. J Am Med Inform Assoc 2013 Mar 1;20(2):317-324 [FREE Full text] [doi: 10.1136/amiajnl-2011-000701] [Medline: 22759620]
18. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S. Computerized clinical decision support for prescribing: provision does not guarantee uptake. J Am Med Inform Assoc 2010;17(1):25-33 [FREE Full text] [doi: 10.1197/jamia.M3170] [Medline: 20064798]
19. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
20. Ahmad R, Zhu NJ, Leather AJ, Holmes A, Ferlie E, ASPIRES Study Co-Investigators. Strengthening strategic management approaches to address antimicrobial resistance in global human health: a scoping review. BMJ Glob Health 2019;4(5):e001730 [FREE Full text] [doi: 10.1136/bmjgh-2019-001730] [Medline: 31565417]
21. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care. Int J Med Inform 2012 Aug;81(8):521-526. [doi: 10.1016/j.ijmedinf.2012.03.002] [Medline: 22483528]

XSL•FO
RenderX

22. McDermott L, Yardley L, Little P, van Staa T, Dregan A, McCann G, eCRT Research Team. Process evaluation of a point-of-care cluster randomised trial using a computer-delivered intervention to reduce antibiotic prescribing in primary care. BMC Health Serv Res 2014 Dec 3;14:594 [FREE Full text] [doi: 10.1186/s12913-014-0594-1] [Medline: 25700144]
23. World Health Organization. MHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
24. Nuttall D. Nurse prescribing in primary care: a metasynthesis of the literature. Prim Health Care Res Dev 2018 Jan;19(1):7-22 [FREE Full text] [doi: 10.1017/S1463423617000500] [Medline: 28786368]
25. United Nations iLibrary. 2019. World Population Prospects, The 2019 Revision - Volume I: Comprehensive Tables URL: https://www.un-ilibrary.org/population-and-demography/world-population-prospects-the-2019-revision-volume-i-comprehensive-tables_15994a82-en [accessed 2020-04-28]
26. The World Bank. 2019. GNI Per Capita, PPP (Current International $) URL: https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD [accessed 2020-04-28]
27. Delory T, Jeanmougin P, Lariven S. A computerized decision support system (CDSS) for antibiotic prescription in primary care–antibioclic: implementation, adoption and sustainable use in the era of extended antimicrobial resistance. J Antimicrob Chemother 2020:75. [doi: 10.1093/jac/dkaa167]
28. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the global burden of disease study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
29. Thursky K. Use of computerized decision support systems to improve antibiotic prescribing. Expert Rev Anti Infect Ther 2006 Jun;4(3):491-507. [doi: 10.1586/14787210.4.3.491] [Medline: 16771625]
30. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. Br Med J 2010 May 18;340(2):c2096. [doi: 10.1136/bmj.c2096] [Medline: 20483949]
31. Peiffer-Smadja N, Lescure F, Maatoug R, Rohaut B. Re: 'determinants of in-hospital antibiotic prescription behaviour' by Lambregts et al. Clin Microbiol Infect 2019 May;25(5):635-637. [doi: 10.1016/j.cmi.2018.12.010] [Medline: 30594653]
32. Heselmans A, van de Velde S, Donceel P, Aertgeerts B, Ramaekers D. Effectiveness of electronic guideline-based implementation systems in ambulatory care settings-a systematic review. Implement Sci 2009 Dec 30;4:82 [FREE Full text] [doi: 10.1186/1748-5908-4-82] [Medline: 20042070]
33. McDermott L, Yardley L, Little P, Ashworth M, Gulliford M, eCRT Research Team. Developing a computer delivered, theory based intervention for guideline implementation in general practice. BMC Fam Pract 2010 Nov 18;11:90 [FREE Full text] [doi: 10.1186/1471-2296-11-90] [Medline: 21087469]
34. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
35. Keitel K, D'Acremont V. Electronic clinical decision algorithms for the integrated primary care management of febrile children in low-resource settings: review of existing tools. Clin Microbiol Infect 2018 Aug;24(8):845-855 [FREE Full text] [doi: 10.1016/j.cmi.2018.04.014] [Medline: 29684634]
36. Rambaud-Althaus C, Shao A, Samaka J, Swai N, Perri S, Kahama-Maro J, et al. Performance of health workers using an electronic algorithm for the management of childhood illness in Tanzania: a pilot implementation study. Am J Trop Med Hyg 2017 Jan 11;96(1):249-257 [FREE Full text] [doi: 10.4269/ajtmh.15-0395] [Medline: 28077751]
37. Shao AF, Rambaud-Althaus C, Samaka J, Faustine AF, Perri-Moore S, Swai N, et al. New algorithm for managing childhood illness using mobile technology (ALMANACH): a controlled non-inferiority study on clinical outcome and antibiotic use in tanzania. PLoS One 2015;10(7):e0132316 [FREE Full text] [doi: 10.1371/journal.pone.0132316] [Medline: 26161535]
38. Keitel K, Kagoro F, Samaka J, Masimba J, Said Z, Temba H, et al. A novel electronic algorithm using host biomarker point-of-care tests for the management of febrile illnesses in Tanzanian children (e-POCT): a randomized, controlled non-inferiority trial. PLoS Med 2017 Oct;14(10):e1002411 [FREE Full text] [doi: 10.1371/journal.pmed.1002411] [Medline: 29059253]
39. Keitel K, Samaka J, Masimba J, Temba H, Said Z, Kagoro F, et al. Safety and efficacy of C-reactive protein-guided antibiotic use to treat acute respiratory infections in tanzanian children: a planned subgroup analysis of a randomized controlled noninferiority trial evaluating a novel electronic clinical decision algorithm (ePOCT). Clin Infect Dis 2019 Nov 13;69(11):1926-1934. [doi: 10.1093/cid/ciz080] [Medline: 30715250]
40. Blanchet K, Lewis JJ, Pozo-Martin F, Satouro A, Somda S, Ilboudo P, et al. A mixed methods protocol to evaluate the effect and cost-effectiveness of an Integrated electronic Diagnosis Approach (IeDA) for the management of childhood illnesses at primary health facilities in Burkina Faso. Implement Sci 2016 Aug 4;11(1):111 [FREE Full text] [doi: 10.1186/s13012-016-0476-5] [Medline: 27488566]
41. Dalaba MA, Akweongo P, Williams J, Saronga HP, Tonchev P, Sauerborn R, et al. Costs associated with implementation of computer-assisted clinical decision support system for antenatal and delivery care: case study of Kassena-Nankana district of northern Ghana. PLoS One 2014 Sep 2;9(9):e106416 [FREE Full text] [doi: 10.1371/journal.pone.0106416] [Medline: 25180831]

XSL•FO
RenderX

42. Zakane SA, Gustafsson LL, Tomson G, Loukanova S, Sié A, Nasiell J, et al. Guidelines for maternal and neonatal 'point of care': needs of and attitudes towards a computerized clinical decision support system in rural Burkina Faso. Int J Med Inform 2014 Jun;83(6):459-469. [doi: 10.1016/j.ijmedinf.2014.01.013] [Medline: 24613689]
43. Adepoju IO, Albersen BJA, de Brouwere V, van Roosmalen J, Zweekhorst M. mHealth for clinical decision-making in sub-saharan Africa: a scoping review. JMIR Mhealth Uhealth 2017 Mar 23;5(3):e38 [FREE Full text] [doi: 10.2196/mhealth.7185] [Medline: 28336504]
44. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]
45. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The district health information system (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag 2019 May;48(2):62-75. [doi: 10.1177/1833358318777713] [Medline: 29898604]
46. Gras G. Use of telemedicine in the management of infectious diseases. Med Mal Infect 2018 Jun;48(4):231-237. [doi: 10.1016/j.medmal.2018.01.005] [Medline: 29452936]
47. Rowbotham S, Chisholm A, Moschogianis S, Chew-Graham C, Cordingley L, Wearden A, et al. Challenges to nurse prescribers of a no-antibiotic prescribing strategy for managing self-limiting respiratory tract infections. J Adv Nurs 2012 Dec;68(12):2622-2632. [doi: 10.1111/j.1365-2648.2012.05960.x] [Medline: 22364215]
48. Williams SJ, Halls AV, Tonkin-Crine S, Moore MV, Latter SE, Little P, et al. General practitioner and nurse prescriber experiences of prescribing antibiotics for respiratory tract infections in UK primary care out-of-hours services (the UNITE study). J Antimicrob Chemother 2018 Mar 1;73(3):795-803 [FREE Full text] [doi: 10.1093/jac/dkx429] [Medline: 29190384]
49. Chang J, Xu S, Zhu S, Li Z, Yu J, Zhang Y, et al. Assessment of non-prescription antibiotic dispensing at community pharmacies in China with simulated clients: a mixed cross-sectional and longitudinal study. Lancet Infect Dis 2019 Dec;19(12):1345-1354. [doi: 10.1016/S1473-3099(19)30324-X] [Medline: 31588042]
50. Morgan DJ, Okeke IN, Laxminarayan R, Perencevich EN, Weisenberg S. Non-prescription antimicrobial use worldwide: a systematic review. Lancet Infect Dis 2011 Sep;11(9):692-701 [FREE Full text] [doi: 10.1016/S1473-3099(11)70054-8] [Medline: 21659004]
51. Grigoryan L, Germanos G, Zoorob R, Juneja S, Raphael JL, Paasche-Orlow MK, et al. Use of antibiotics without a prescription in the US population: a scoping review. Ann Intern Med 2019 Jul 23:- epub ahead of print. [doi: 10.7326/M19-0505] [Medline: 31330541]
52. Birgand G, Castro-Sánchez E, Hansen S, Gastmeier P, Lucet JC, Ferlie E, et al. Comparison of governance approaches for the control of antimicrobial resistance: analysis of three European countries. Antimicrob Resist Infect Control 2018;7:28 [FREE Full text] [doi: 10.1186/s13756-018-0321-5] [Medline: 29468055]
53. Anderson M, Schulze K, Cassini A, Plachouras D, Mossialos E. A governance framework for development and assessment of national action plans on antimicrobial resistance. Lancet Infect Dis 2019 Nov;19(11):e371-e384. [doi: 10.1016/S1473-3099(19)30415-3] [Medline: 31588040]
54. Government of UK. 2019. UK 5-year Action Plan for Antimicrobial Resistance 2019 to 2024 URL: https://www.gov.uk/government/publications/uk-5-year-action-plan-for-antimicrobial-resistance-2019-to-2024 [accessed 2020-04-28]

Abbreviations

AMR: antimicrobial resistance
CDSS: clinical decisions support system
HIC: high-income countries
LMIC: low- and middle-income countries
SPILF: French Infectious Diseases Society
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by C Luz, S Khairat; comments to author 03.03.20; revised version received 13.03.20; accepted 31.03.20; published 20.07.20.

Please cite as:
Peiffer-Smadja N, Poda A, Ouedraogo AS, Guiard-Schmid JB, Delory T, Le Bel J, Bouvet E, Lariven S, Jeanmougin P, Ahmad R, Lescure FX
Paving the Way for the Implementation of a Decision Support System for Antibiotic Prescribing in Primary Care in West Africa: Preimplementation and Co-Design Workshop With Physicians
J Med Internet Res 2020;22(7):e17940
URL: https://www.jmir.org/2020/7/e17940
doi:10.2196/17940
PMID:32442155

XSL•FO
RenderX

©Nathan Peiffer-Smadja, Armel Poda, Abdoul-Salam Ouedraogo, Jean-Baptiste Guiard-Schmid, Tristan Delory, Josselin Le Bel, Elisabeth Bouvet, Sylvie Lariven, Pauline Jeanmougin, Raheelah Ahmad, François-Xavier Lescure. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study

Yang Xiang[1], PhD; Hangyu Ji[2], MD; Yujia Zhou[1], MBBS; Fang Li[1], PhD; Jingcheng Du[1], PhD; Laila Rasmy[1], MSc; Stephen Wu[1], PhD; W Jim Zheng[1], PhD; Hua Xu[1], PhD; Degui Zhi[1], PhD; Yaoyun Zhang[1], PhD; Cui Tao[1], PhD

[1]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[2]Division of Gastroenterology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

Corresponding Author:
Cui Tao, PhD
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
UCT 600, 7000 Fannin Street
Houston, TX, 77030
United States
Phone: 1 713 500 3981
Email: Cui.Tao@uth.tmc.edu

Abstract

Background: Asthma exacerbation is an acute or subacute episode of progressive worsening of asthma symptoms and can have a significant impact on patients' quality of life. However, efficient methods that can help *identify personalized risk factors and make early predictions* are lacking.

Objective: This study aims to use advanced deep learning models to better predict the risk of asthma exacerbations and to explore potential risk factors involved in progressive asthma.

Methods: We proposed a novel time-sensitive, attentive neural network to predict asthma exacerbation using clinical variables from large electronic health records. The clinical variables were collected from the Cerner Health Facts database between 1992 and 2015, including 31,433 adult patients with asthma. Interpretations on both patient and cohort levels were investigated based on the model parameters.

Results: The proposed model obtained an area under the curve value of 0.7003 through a five-fold cross-validation, which outperformed the baseline methods. The results also demonstrated that the addition of elapsed time embeddings considerably improved the prediction performance. Further analysis observed diverse distributions of contributing factors across patients as well as some possible cohort-level risk factors, which could be found supporting evidence from peer-reviewed literature such as respiratory diseases and esophageal reflux.

Conclusions: The proposed neural network model performed better than previous methods for the prediction of asthma exacerbation. We believe that personalized risk scores and analyses of contributing factors can help clinicians better assess the individual's level of disease progression and afford the opportunity to adjust treatment, prevent exacerbation, and improve outcomes.

(J Med Internet Res 2020;22(7):e16981) doi:10.2196/16981

KEYWORDS

asthma; deep learning; electronic health records; health risk appraisal

Introduction

Background

Asthma is a common and serious health problem that affects 235 million people worldwide [1] and an estimated 26.5 million people (8.3% of the US population) in the United States [2]. Asthma takes a significant toll on the population, which imposes an unacceptable burden on health care systems. In 2013, the total annual cost of asthma was US $81.9 billion in the United States [3]. If not well controlled or stimulated by specific risk

XSL•FO
RenderX

factors, asthma may develop into exacerbations (asthma attacks), which are acute or subacute episodes characterized by a progressive increase in one or more typical symptoms of asthma (dyspnea, coughing, wheezing, and chest tightness) [4]. In 2016, 12.4 million current asthmatics (46.9%) in the United States had at least one asthma exacerbation in the previous year [2]. Exacerbations of asthma can be severe and require immediate medical interventions, either as an emergency department (ED) visit or admission to the hospital [5]. Serious asthma exacerbations may even result in death [6]. Therefore, it is of practical significance to make early predictions such that interventions can be carried out in advance to reduce the probability of an exacerbation.

Investigations on risk factor analysis or prediction for asthma exacerbation have been respectable, in which the mainstream adopts traditional statistical methods, such as logistic regression [7-10], proportional hazards regression [11], and generalized linear mixed models [12]. However, most of them have only explored a small group of candidate risk factors and are usually hard to extend to other data sets and make personalized predictions difficult [13,14]. With the explosion of health care data in recent years, machine learning methods have taken a nontrivial place in this domain, benefiting from their general superiority over statistical methods in processing larger numbers of variables and flexibility in modeling more complex correlations [15]. Typical models include naïve Bayes [16], Bayesian networks [16-19], artificial neural networks [17], Gaussian process [17], and support vector machines [16,17]. However, although different attempts have been made, there are still several deficiencies in traditional machine learning methods. For example, ignoring temporal dependencies between variables might not provide a meaningful risk estimation of future exacerbations for individual patients [14]. Furthermore, most approaches only concentrate on the quantitative evaluation of prediction performance, but lack further attention to personalized risk factors [20].

Recent revolutions in health artificial intelligence started from deep learning, which has an upper hand on health care predictions because of its flexibility in dealing with longitudinal data [21], powerful learning capabilities [22], and ability to alleviate the problem of data irregularity [23]. One of the most popular architectures is recurrent neural networks (RNNs), which make predictions according to the sequence of historical events. Dozens of successes have been achieved in applying deep learning to disease predictions [24], mostly using variants of RNNs with distinct network components, for example, by adding an attention mechanism to evaluate the weights of each variable [25-29] or by using special configurations to tackle the problem of time decays [23,25,27,30-32]. Typical prediction tasks include the prediction of diabetes mellitus [23], Parkinson disease [29,33], chronic heart failure [26], sepsis [34], mortality, and readmission [25]. However, deep learning–based studies on the prediction of asthma exacerbation remain lacking. Do et al [35] proposed a protocol for the prediction based on RNNs and reinforcement learning but did not test the method on real-world data.

Objectives

Inspired by previous studies, we applied long short-term memory (LSTM) [33], a popular RNN variant commonly used by previous predictive models [23-25,29,34] as the main framework for asthma exacerbation prediction, which can mitigate the gradient vanishing problem in RNNs. We proposed the time-sensitive, attentive neural network (TSANN), which employs a self-attention mechanism [36] to help model the context of both visit-level and code-level variables. Meanwhile, to incorporate the impact of elapsed time, we projected the relative time of each clinical variable into a low-dimensional space and combined it with the code representations. Using the attention weights of the TSANN, data analysis was then conducted to investigate personalized and cohort-level risk factors.

There are major differences between TSANN and recent state-of-the-art deep learning–based clinical predictive models such as time-aware LSTM (TLSTM) [23], Reverse Time Attention model (RETAIN) [27], and Attention-based Time-aware Disease Progression (ATTAIN) [32]. First, the model structures are different. Compared with TLSTM and ATTAIN, which only include 1 layer of RNN, our two-layer architecture enables us to analyze the relative importance of each event within each visit. Although RETAIN also has 2 layers of attention, it does not have explicit hierarchical structures as TSANN. Instead, an additional inference step is required to obtain the contribution of each variable. Second, TSANN uses a different approach to model the elapsed time. RETAIN, TLSTM, and ATTAIN feed the time elapsed into a decay function as a single value and multiply it with the network memory. In comparison, the elapsed time embeddings in TSANN are more analogous to position embeddings in natural language processing, which were introduced to model the relative distance between words by learning multidimensional and semantic representations to facilitate certain tasks such as relation classification [37] and neural language modeling [36,38,39]. By using time embeddings, we assume that time is no longer a single value as it was used in previous methods, but it can represent more complex patterns together with clinical variables such as varying lengths of correlations between variables.

The primary aims of this study were (1) to propose a novel predictive model with better performance and (2) to add the transparency of the model by visualizing contributing factors at both the individual and cohort levels. Furthermore, the proposed model can potentially be applied to other clinical problems. Deep learning models are usually scalable. Although focusing on asthma exacerbation for this specific project, the proposed approach can also be adopted in risk predictions for other chronic diseases. We hope that the associated pipeline of deep learning–based predictive modeling, including data collection, model training, model evaluation, and risk factor analysis, can help the clinical community better understand the underlying mechanisms of disease progression and assist in decision making.

XSL•FO
RenderX

Methods

Problem Statement

Given a sequence of historical clinical variables in patients with asthma, we aimed to evaluate the risk of developing asthma exacerbation in the designated time window. Meanwhile, personalized contributing factors are to be identified to facilitate the evaluation of disease progression and make early interventions.

Database

The study used Cerner Health Facts, a Health Insurance Portability and Accountability Act–compliant database collected from multiple enrolled clinical facilities, containing mostly inpatient data. Data in Health Facts were extracted directly from the electronic health records (EHRs) from hospitals with which Cerner has a data use agreement. Encounters may include the pharmacy, clinical and microbiology laboratory, admission, and billing information from affiliated patient care locations. All personal identifying information of the patients was anonymized. In this study, we primarily focused on the impact of clinical factors on asthma exacerbation; therefore, we extracted diagnoses, medications, and demographic characteristics such as gender, race, and age from the database as clinical variables or clinical events. The University of Texas Health Science Center (UTHealth) had agreements with Cerner to use these data for research purposes. The institutional review board at the UTHealth approved the study protocol.

Study Design

We conducted a retrospective study to predict the risk of asthma exacerbation. Patients' records between 1992 and 2015 were extracted from the Cerner database. For clarity, we defined several terms in advance (Table 1).

We built 2 cohorts as simulations for both the real-world application scenario (*early prediction*) and the model evaluation (called *next-visit prediction* in many previous studies [27,32]). In early prediction, we could not foresee when the exacerbation would happen but could only evaluate the future risk at each visit. In our study, we selected the fifth visit from the asthma index as the prediction date (*testing set A*) according to the average number of visits (5.78, SD 6.04) between asthma index and exacerbation among the patients. The detailed steps for the cohort selection are listed in Multimedia Appendix 1 [4,25,40-55]. In next-visit prediction, we simply set the penultimate visit as the prediction date (*testing set B*). We set testing set A as our primary evaluation set as it was much closer to the realistic diagnostic situation.

The TSANN model was trained to evaluate the risk of asthma exacerbation given the observed time window. The main outcomes of the proposed method are (1) a score that measures the risk of asthma exacerbation for each patient and (2) visualization of the results, including a personalized heatmap identifying the importance of each clinical variable in the observed time window, cohort-level risk factors, and their temporal distributions among patients. On the basis of the outcomes, further data mining or clinical trials can be carried out for validation. For example, cohort-level factors will help data scientists reduce labor and expertise in collecting candidate risk factors from the literature before conducting a regression analysis. Patient-level factors will facilitate physicians and patients in better understanding disease progression. The workflow of this research is shown in Figure 1.

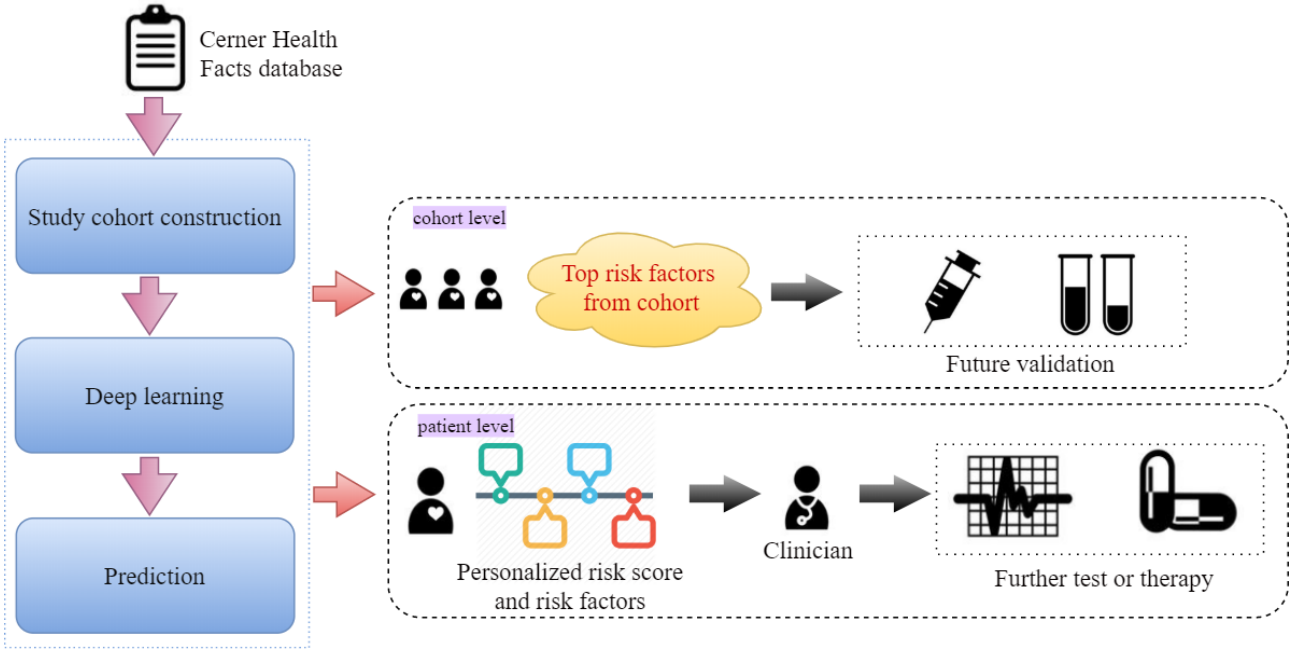
Table 1. Defined terms for asthma exacerbation prediction.

Term	Definition
Index date	The date of the first diagnosis of asthma in a patient's EHR[a]
Exacerbation date	The date of the first diagnosis of asthma exacerbation after the index date
Case group	Patients with asthma and later asthma exacerbations within 365 days and satisfying the inclusion and exclusion criteria
Control group	Patients with asthma but without exacerbations within 365 days and satisfying the inclusion and exclusion criteria
Prediction date	Training set: for the case group, the visit date before the exacerbation date; for the control group, the penultimate visit date within 365 days: • Testing set A: the fifth visit starting from the index date • Testing set B: defined analogously to the training set
Observed time window	The time window between the index date and the prediction date

[a]EHR: electronic health record.

XSL•FO
RenderX

Figure 1. The workflow of the prediction and risk analysis of asthma exacerbation.



Selection of Study Subjects

The subjects in the study were patients with a diagnosis of asthma. The inclusion and exclusion criteria derived from previous studies [4,56] were as follows.

Inclusion Criteria

The subjects in the study were patients with at least one record of asthma diagnosis. The definitions of asthma and exacerbation are as follows.

Asthma

1. Asthma diagnosis codes were provided according to the International Classification of Disease Code (ICD; ICD-9 code 493.xx or ICD-10 code J45.xx). This is the first occurrence of asthma in the patient's EHR.
2. At least one of the asthma medications was prescribed on the asthma diagnosis date (the index date). Asthma medications include short-acting beta agonists, inhaled corticosteroids (ICS), long-acting beta agonists (LABA), leukotriene receptor antagonists, anticholinergics, and ICS/LABA combinations.

Asthma Exacerbation

1. Asthma (ICD-9 code 493.xx or ICD-10 code J45.xx) was given as a primary diagnosis for an ED visit or hospitalization.
2. At least one oral corticosteroid treatment was received.

Exclusion Criteria

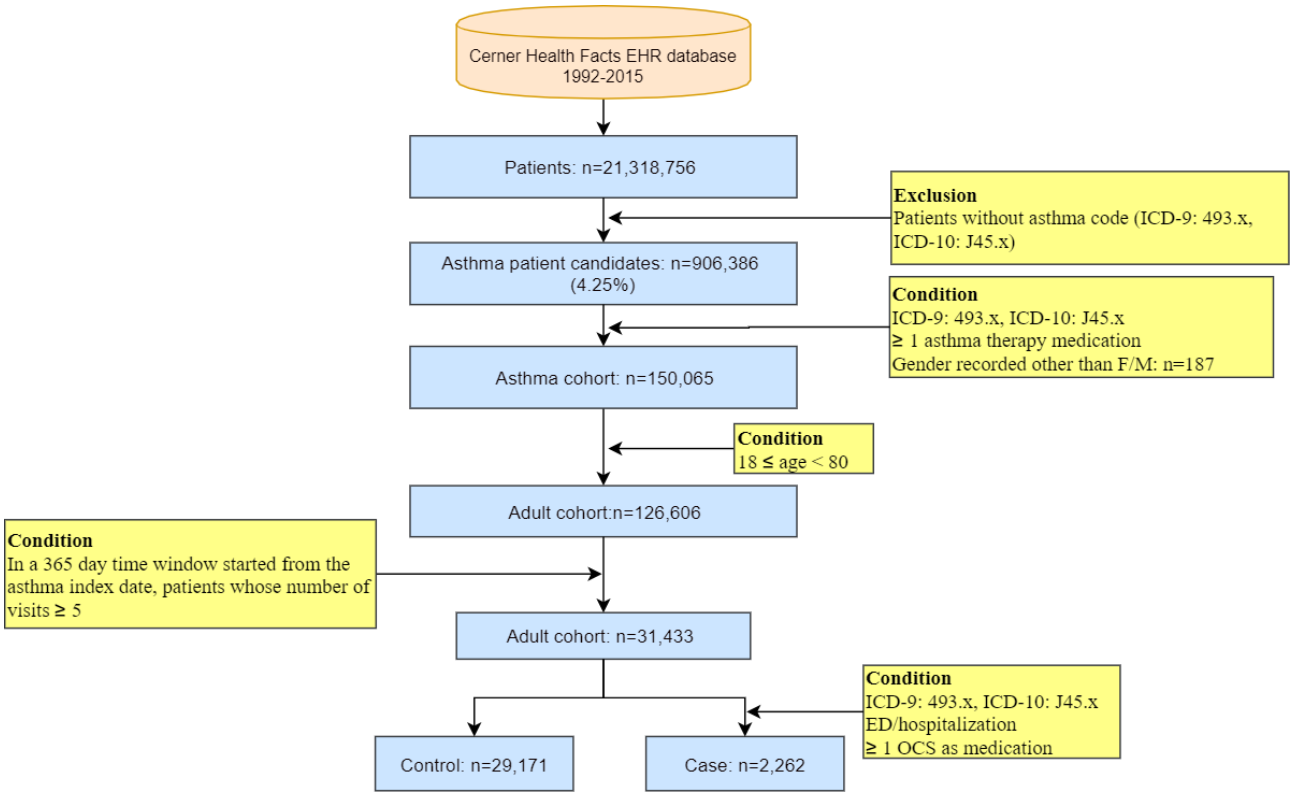
To allow the data to better fit for machine learning models, we excluded the following patients:

1. Those with missing or unclear time information (eg, with a wrongly recorded format of time stamps)
2. Those with a gender other than male or female
3. Those whose number of visits is <5 in the observed time window

This study only focused on adult patients aged between 18 and 80 years. In the end, 31,433 individuals remained, including 2262 cases and 29,171 controls (case by control ratio approximately 1:13). The cohort selection process is shown in Figure 2. A detailed descriptive analysis of the cohort is shown in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. Cohort selection process for the study of asthma exacerbation. ED: emergency department; EHR: electronic health record; ICD: International Classification of Disease Code; OCS: oral corticoids.



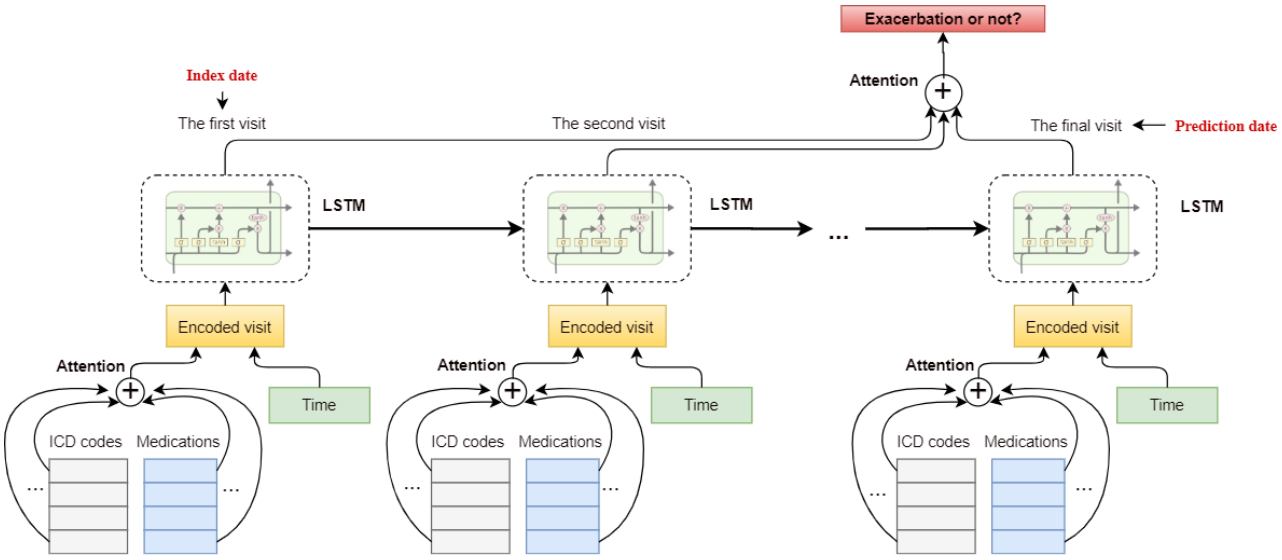
Time-Sensitive Attention Neural Network

Model Overview

TSANN takes the whole sequence of clinical variables in the observed time window as inputs and outputs the probability of asthma exacerbation (Figure 3). The architecture of TSANN is based on LSTM and strengthened by the addition of hierarchical attention and elapsed time embeddings.

For each visit, multiple clinical variables were encoded in the input layer and averaged through the code-level attention mechanism. The elapsed time embedding is attached to each visit as complementary information to indicate the time interval between the date of each visit and the prediction date. LSTM then accepts the sequence of encoded visits as inputs and outputs further encodings for each visit. The visit-level attention layer is then applied to the outputs of the LSTM to summarize all the visits for each patient. Finally, by feeding the output of visit-level attention into the Softmax function, a probability indicating the risk of disease onset is generated.

Figure 3. Overview of the time-sensitive attentive neural network model for asthma exacerbation prediction. ICD: International Classification of Disease Code; LSTM: long short-term memory.



XSL•FO
RenderX

Input

The inputs of the model consist of 2 types of features. One type is clinical concepts (we use *clinical concepts* and *clinical variables* interchangeably), including ICD codes, medications, and demographic features. All ICD-10 codes were converted into ICD-9 based on predefined mappings [57] because very few diagnosis codes in our data set were encoded by ICD-10 as the data collection time range is between 1992 and 2015, but the implementation of ICD-10 started in October 2015. All medications were normalized to their generic names. The demographic features included age, gender, and race, which were only taken as inputs on the prediction date. Using a projection matrix (V_c: concept vocabulary size and D_c: concept embedding dimension), we mapped each clinical concept into a concept-embedding vector:



where C_{ij} is the generated concept-embedding vector and  is the one hot vector denoting the existence of clinical concept j in visit i.

The other feature type is time features, which indicates the occurrence time for each clinical variable. Intuitively, variables with different time stamps would behave differently in prediction. For instance, in many cases, a clinical event that happened several days ago would play a more important role than one that happened several months ago. Meanwhile, due to the nature of data irregularity and deficiency of EHRs, successive visits always have diverse time intervals [23], which makes it indispensable to consider the time elapsed when conducting predictive modeling.

Elapsed time embeddings were introduced to represent the relative time gap for each clinical concept. Specifically, taking the time of the prediction date T_0 as a pivot, the time attribute of each clinical concept is the absolute difference between its occurrence time T_i and T_0, that is, the relative time gap T_0 -T_i. As the observed time window has an upper bound of 365 days, the vocabulary size V_t of the time embeddings was set as 365.

We applied a matrix  to project each time value to an m-dimension vector. Unlike the clinical concept embeddings, elapsed time embeddings are fed into the model after the code-level attention and assigned to each visit. The equation to obtain the elapsed time embedding for each visit is analogous to that for concept embeddings, where: 



Specifically, the minimum time unit in this study was set as *day*.

Code-Level Attention

Attention is a mechanism specifically designed for deep neural networks that acts as an information filter; meanwhile, it can alleviate information loss when dealing with long sequences. It selects important sequence steps by assigning them different weights [58,59]. Through attention, each clinical concept is assigned a weight such that important concepts would have larger weights than the others. We adopted the attention mechanism from Yang et al [60], in which the weight of each variable is generated according to the sequence and context vector. Specifically, given the set of codes  in the ith visit, the encoded representation for visit v_i can be generated by:







where W_v and b_v are the weight and bias for matrix transformation, u_{ij} is the attention vector for each code j in v_i, u_v is the context vector for v_i, which is randomly initialized and updated during training, and $_{ij}$ is the attention weight for the concept Vij based on which we can generate its final weight. By assigning time embeddings to the ith visit v_i, the representation of each visit is updated as , where  denotes the matrix concatenation.

Visit-Level Attentive LSTM Layer

Taking the encoded representation of each visit as input, LSTM models the sequential information in the observed time window and obtains the summarization at the final step (the prediction date). The advantage of LSTMs over traditional RNNs is that they can alleviate the gradient vanishing problem and are thus able to retain longer *memories* from prior time stamps [61,62]. LSTMs are implemented by several matrix multiplications and nonlinear transformations that aim to mimic the memory mechanism of human brains, which are called *gates*, signifying that the network can select effective information and abandon useless information. The equations of the LSTMs are as follows:













where Ws and bs are the weights and biases for different gates or cells (f_t: forget gate, i_t: input gate, C_t: memory cell, o_t: output gate, and h_t: hidden cell) and σ is the activation function, such as *Tanh* or *sigmoid*.

By assigning attention weights to the outputs of LSTM from each step, we can weight each visit in the observed time window and obtain a summary of these visits as r_p:

XSL•FO
RenderX







where W_p and b_p are the weight and bias for matrix transformation, u_i is the attention vector for each visit i given v_i, u_p is the context vector, and $_j$ is the attention weight for each visit v_j. This process can be seen as a simulation of the diagnosis procedure of a clinic visit, during which a physician would look back into a patient's EHR, measure the impacts of each historical clinical event, and make the final decision.

Output

The visit-level attention layer compresses all the information in the observed time window into a fixed-length vector r_p. The output of attention goes through a fully connected layer with nonlinear activation. Finally, a Softmax function is applied to generate the prediction probability, *P:*



where *P* is used as the score to evaluate the risk of developing asthma exacerbation.

Evaluation

Area under the receiver operating curve (AUC) is widely used as an evaluation metric for predictive models, which reflects a balance between sensitivity and specificity [63]. According to the prediction probability *P* (between 0 and 1) for each instance, the AUC value is generated by setting different cutoffs. The methods listed in Table 2 were compared in our experiments.

Table 2. The methods used for comparisons.

Method	Note
LR[a]	A popular conventional machine learning algorithm [64], usually serving as a strong baseline in predictive modeling [27]. The input of LR is a fixed-length feature vector that denotes the frequencies of each variable. For LR considering time, we associate each variable with its time stamp and expand the vocabulary. We did not use *day* as the time unit as it would have introduced a greater number of variables (ie, 12,390×365 [the code vocabulary size×the maximum number of days]), which would have been too sparse and difficult for computation. Instead, we set month as the time unit, and finally, 148,680 distinct clinical variables were generated. We employed the Synthetic Minority Oversampling Technique [65] to help alleviate the problem of data imbalance
MLP[b]	The MLP model used in this study contains 1 input layer and 1 Softmax layer [66]. The representations of all the codes were averaged on each dimension after being projected to the embedding space for each patient
LSTM[c]	The basic LSTM algorithm, taking the sequence of the clinical variables as input ordered by time. The variables in each visit are averaged
ALSTM[d]	Comprising 1 layer of LSTM and 1 layer of attention
TLSTM[e] [23]	The time-aware LSTM model, which is one of the state-of-the-art predictive models. In TLSTM, the time gap is used to compute the information decay in the LSTM unit
RETAIN[f] [27]	A two-layer attention model, which is another state-of-the-art model for the prediction of disease onset. In RETAIN, the time features are not embedded as vectors but real values denoting the gaps from the first visit
ATTAIN[g] [32]	A modification of TLSTM with special types of attention mechanisms added (flexible attention). It also uses a similar time decay function as RETAIN. We implemented it ourselves using TensorFlow
TSANN[h]-I	The proposed TSANN model but with the second attention layer removed. Prediction is based on the final state of LSTM
TSANN-I-step	Apply the time-encoding method from Song et al [39] on TSANN-I. In TSANN-I-step, although time was also encoded using a vector, it only showed the order of each visit, for example, 1, 2, 3 for consecutive visits, but not the actual elapsed time
TSANN-II	A complete version of the proposed TSANN model

[a]LR: logistic regression.

[b]MLP: multilayer perceptron.

[e]LSTM: long short-term memory.

[d]ALSTM: attention long short-term memory.

[e]TLSTM: time-aware long short-term memory.

[f]RETAIN: Reverse Time Attention model.

[g]ATTAIN: Attention-based Time-Aware Disease Progression.

[h]TSANN: time-sensitive attentive neural network.

For evaluation, we first split the data into a training set and a held-out testing set with a ratio of 8:2. Furthermore, five-fold cross-validation was performed on the training data set for parameter tuning. During cross-validation, a grid search was applied to tune the hyperparameters including learning rate (0.0005, 0.001, 0.005, 0.01), l2 penalty (0.0001, 0.0005, 0.001), batch size (32, 64, 128), activation function for LSTM (ReLU [40] and Leaky_ReLU [41]), whether to add batch normalization

XSL•FO
RenderX

[45], and the optimizer selection between RMSprop [42] and Adam [43]. We then averaged the AUCs of each epoch (up to 30 epochs) across five folds to obtain the best training epoch. The optimal hyperparameters were adopted to retrain the model on the entire training set and produce the AUC on the testing set. Finally, the hyperparameters for the model TSANN-I, which has the best AUC value, were as follows: batch size=32, concept embedding dimension=100, time embedding dimension=20, Adam as the optimizer with learning rate=0.001, l2 penalty=0.0001 for all parameters, Leaky_ReLU as the activation function, and adding batch normalization before Softmax. The codes for RETAIN and TLSTM were reused from the respective studies. All other deep learning models were implemented with TensorFlow [44] and trained on Nvidia Tesla V100, Quadro P6000, and Titan XP GPUs. We shared our code on GitHub to facilitate other researchers [67].

Results

AUC Values

AUC values with (+time) and without the consideration of time (–time) on testing set A (the primary evaluation set) are shown in Table 3. In the table, for TLSTM, we only considered a *+time* version as it is defined as a time-aware variant of LSTM, and for multilayer perceptron (MLP), LSTM, attention long short-term memory (ALSTM), TSANN-I, TSANN-I-step, and TSANN-II, we used the elapsed time embeddings introduced in this study to include time.

Table 3. Area under the curve (AUC) values by different models (–time: time information was excluded and +time: time information was included).

Method	AUC[a]–time	AUC+time
LR[b]	0.6447	0.6773
MLP[c]	0.6545	0.6753
LSTM[d]	0.6045	0.6567
ALSTM[e]	0.6346	0.6714
TLSTM[f]	—	0.6548
ATTAIN[g]	0.6119	0.6597
RETAIN[h]	0.6455	0.6882
TSANN[i]-I	0.6692	*0.7003* [j]
TSANN-I-step	0.6463	—
TSANN-II	*0.6827*	0.6855

[a]AUC: area under the receiver operating curve.

[b]LR: logistic regression.

[c]MLP: multilayer perceptron.

[d]LSTM: long short-term memory.

[e]ALSTM: attention long short-term memory.

[f]TLSTM: time-aware long short-term memory.

[g]ATTAIN: Attention-based Time-aware Disease Progression.

[h]RETAIN: Reverse Time Attention model.

[i]TSANN: time-sensitive attentive neural network.

[j]The optimal value for each column is italicized.

When comparing vertically (different rows) and considering time information, we noticed that TSANN-I achieved the optimal AUC value, improving the strongest baseline (RETAIN) by 1.21% (the difference was significant according to the Wilcoxon test with P=.03). Among other methods, TSANN-II achieved a performance comparable with that of RETAIN. The conventional machine learning method logistic regression (LR) behaved better than some deep learning methods but was worse than RETAIN, TSANN-I, and TSANN-II. TSANN-I-step, which only used time embeddings to denote the relative position of each visit, did not produce good results. Although TLSTM and ATTAIN performed well on other tasks, they did not obtain satisfactory results on our data. For results without time, TSANN-I and -II performed much better than others, with a maximum improvement of 2.82%.

When comparing the results horizontally (–/+ time), considerable improvements were observed after adding time information on most methods; for example, TSANN-I obtained a 3.11% improvement. Surprisingly, TSANN-II, when integrating time embeddings, did not improve considerably.

Better performances by TSANN models and considerable improvements after adding the time information could also be observed on testing set B (we did not list the results here but showed them in Multimedia Appendix 1 as it is not our primary evaluation set). As expected, the general results on testing set

XSL•FO
RenderX

B were better than those on testing set A, as a sample conveys more complete information.

Patient-Level Risk Factors

In this study, a heatmap was used for each patient's EHR for the visualization of highly associated variables or possible risk factors. The heatmap illustrates how each variable behaves in each visit during the progression of asthma. Each grid in the heatmap is colored based on the attention weights derived from the model. The darker an area, the more important the clinical variable and the higher the association it has with exacerbation. For example, Figure 4 shows a case where the symptoms of *hypoxemia*, *shortness of breath*, and *wheezing* (799.02, 786.05, and 786.07 in ICD 9, respectively) were recognized as highly associated variables. A possible explanation might be that the patient's status of hypoxemia worsened the condition of asthma following symptoms in the breath, and asthma exacerbation was then diagnosed.

These highly associated variables can either be signs of asthma worsening or be triggers for exacerbation, which requires further confirmation by domain experts. Signs including symptoms and treated medications may convey important clues for disease progression and will help clinicians in making final diagnoses, whereas the triggers behaving as personalized risk factors will potentially benefit early interventions. In addition, each heatmap is associated with a probability score derived from equation 15, indicating the risk of the patient in developing an exacerbation (the top row in Figure 4, *predicting 1* indicates *predicting exacerbation*).

Figure 4. An example of a heatmap with highly associated clinical variables, such as hypoxemia (D_799.02), shortness of breath (D_786.05), and wheezing (D_786.07).



Cohort-Level Risk Factors

We also discovered common highly associated variables at the cohort level using the normalized multiplication of the visit-level and code-level attention weights. We then asked the physician to help distinguish the type of these variables according to their expertise and literature lookup. Some of these variables can be confirmed as risk factors, for example, poor control of respiratory diseases [4] and gastroesophageal reflux disease [44], although others need further validations, such as chest pain, migraine, and use of some medications. The top-ranked variables derived from the model are shown in Table 4. The details of the method and the explanations of these factors are described in Multimedia Appendix 1.

Apart from demonstrating the list of cohort-level factors, using the weights generated by equation 2 and equation 3 in Multimedia Appendix 1, we can also visualize how each clinical variable contributes across time; for example, a variable may behave distinctly among individuals with different action times or different incidences. Figures 5 and 6 present 2 examples in which the time distributions for the clinical variables are displayed through scatters. In these scatters, each circle represents a patient where its size and color depth denote the importance of the corresponding variable. In the figures, the x-axis represents the time gap between the occurrence date of the variable and the prediction date, whereas the y-axis is employed merely for cosmesis. We randomly selected a maximum of 2000 patients to plot this figure.

Figures 5 and 6 were derived from an ICD code (530.81) and a medication (fentanyl), respectively. We observed different effective time ranges for these 2 factors, where the first factor tends to distribute more intensively between the previous 250 to 50 days, whereas the second factor focuses more intensively on the previous 100 days. We hope that these visualizations can help determine the temporal distributions of highly associated factors to aid asthma control.

XSL•FO
RenderX

Table 4. Clinical variables with the top-ranked weights (/N stands for the clinical variable presented in N months before the prediction date).

Sr. No.	ICD[a]-9/occurrence time	Medication/occurrence time
1	493.9×asthma/0-5[b] (meaning diagnosed with asthma multiple times before exacerbation)	Methylprednisolone/0, 1[c]
2	786.07 wheezing/0-2[d]	Prednisone/0, 1, 2[c]
3	496.0 chronic airway obstruction not elsewhere classified/0, 1[e]	Ipratropium/0, 1, 2[c]
4	530.81 esophageal reflux/0[b]	Midazolam/0, 1, 2[d]
5	V46.2 dependence on supplemental oxygen/0[d]	Hydromorphone/0-2[e]
6	787.02 nausea alone/0[d]	Heparin/0, 1[d]
7	786.50 unspecified chest pain/0[d]	Acetaminophen-oxycodone/0[b]
8	V08 HIV infection status/0[e]	Fentanyl/0[e]
9	786.59 other chest pain/0[d]	Methylprednisolone/2-4[e]
10	786.05 shortness of breath/0[d]	Glycopyrrolate/0[b]
11	V58.69 long-term (current) use of other medications/0[e]	Lidocaine/0[d]
12	784.0 headache/0[e]	Dexamethasone/0[d]
13	346.90 migraine, unspecified, without mention of intractable migraine without mention of status migrainosus/0[e]	Promethazine/0[d]
14	V58.66 long-term (current) use of aspirin/0[b]	Atorvastatin/0[d]
15	491.21 obstructive chronic bronchitis with (acute) exacerbation/0[e]	Furosemide/0[c]

[a]ICD: International Classification of Disease Code.

[b]Identified possible risk factors of asthma exacerbations by the domain expert. The authors regard these as containing valuable information.

[c]These medications can be used to treat asthma or control the symptoms of asthma. In this study, it was difficult to determine whether these medications are risk factors as we were unable to investigate the dosage of these medications in the current study. Inappropriate medication use, short-acting beta agonists/inhaled corticosteroids, could also lead to asthma exacerbations.

[d]These factors were symptoms, comorbidities, or combined medications. We believe they were not risk factors for asthma exacerbations.

[e]It could hardly be determined whether these factors caused asthma exacerbations, but they demonstrated high associations. The authors regard these as containing valuable information.

XSL•FO
RenderX

Figure 5. Time distribution of the contribution of the clinical variable gastroesophageal reflux disease is denoted by ICD-9:530.81. ICD: International Classification of Disease Code.

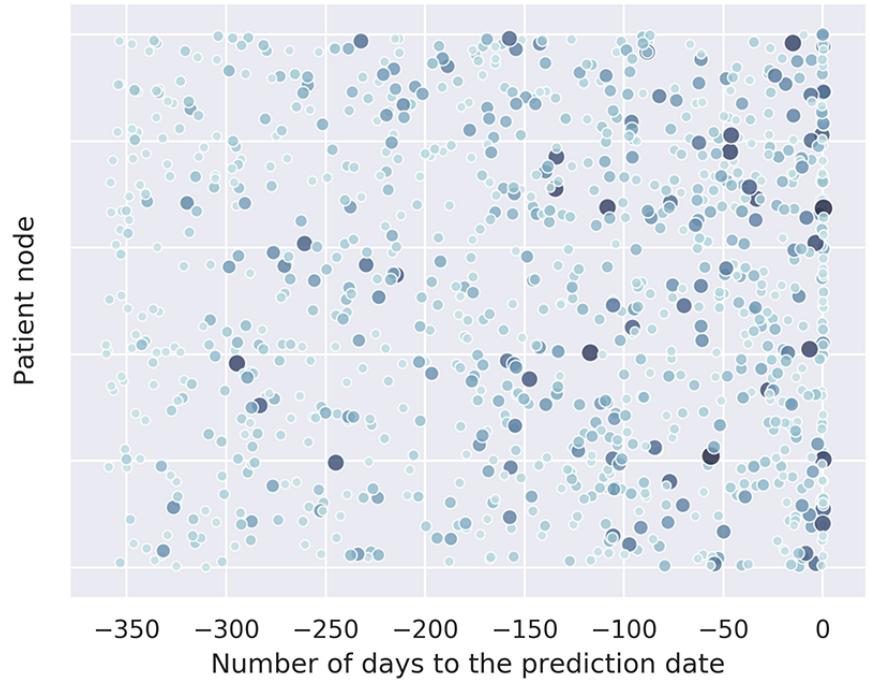
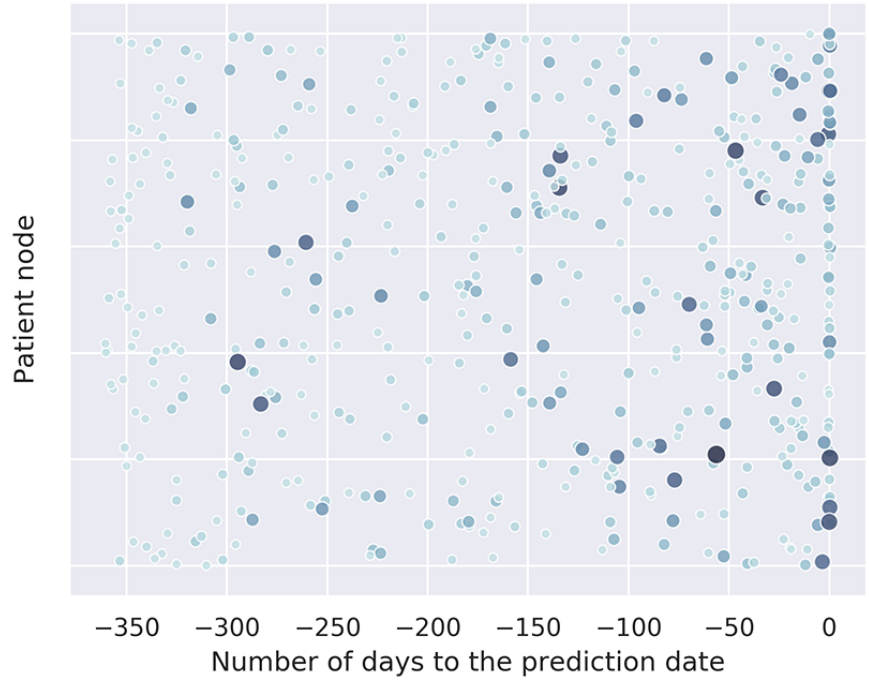


Figure 6. The time distribution of the contribution of the clinical variable fentanyl.



Discussion

Principal Findings

Our proposed method obtains the optimal AUC value on the prediction task, with hierarchical attention and elapsed time embeddings as its booster. The visualization also provides useful tracks for a better understanding of disease progression. The primary outcomes of this study are as follows: (1) a new state-of-the-art predictive model for asthma exacerbation prediction was proposed and validated and (2) a reasonable pipeline of disease risk prediction and factor analysis was introduced. Some of the identified risk factors can be validated from the literature, which shows the effectiveness of the method, whereas some other factors, although supportive shreds of evidence were seldom reported in previous studies, offer meaningful insights for further research. The discussions in this section are primarily based on the results of testing set A without additional comments.

Model Performance

TSANN-I and TSANN-II have the capacity to capture nonlinearities and learn more complex dependency relationships between variables, benefiting from the structure of hierarchical

XSL•FO
RenderX

attention and the addition of elapsed time embeddings. It is difficult for LR and MLP to learn temporal dependencies between variables according to their natural structures, which makes them perform worse than TSANNs. However, they obtained better AUCs than LSTM and ALSTM on testing set A, partly because the truncation of EHRs by 5 visits weakens their advantages in modeling longer sequences. Although TLSTM and ATTAIN also integrated the time information, they did not obtain satisfactory results as other methods. It is likely that the combination of the single-layer deep learning structure and the numerical time decay function is insufficient in dealing with more complex temporal patterns on this data set and even confuses the classifiers. RETAIN also has a hierarchical attention structure, as introduced in the Introduction section. However, one of its attention mechanisms is applied to each code-embedding dimension, which is different from ours and requires an additional inference step for interpretation. In addition, our addition of time embeddings enhanced the flexibility of modeling the time information, which contributed considerably to the performance.

A typical characteristic of the EHR data is irregularity, which means that the time gaps between clinic visits are irregular and the visits are often sparsely distributed along the timeline and sometimes are even missing. Thus, the predictive model is responsible for serializing the visits for each patient with consideration of time elapses between continuous visits and reduces the effect of missing data. The comparisons between results with and without time information in Table 3 demonstrate the effectiveness of considering time elapses in this cohort. It might be inferred that the prediction of asthma exacerbation is quite time sensitive and most of the critical risk factors should have been time stamped. For instance, even for a visit just before the prediction date, if its occurrence is several months earlier, its impact would be reduced. Similar cases can also be found in Baytas et al [23], who reported an improvement of 6% from LSTM to TLSTM. In comparison, for TSANN-I-step, although time embeddings were also used, they were only used to denote the relative position of each visit in the sequence but lacked the ability to represent time decays, which can hardly obtain satisfying results here. However, adding time to TSANN-II did not improve much as in other methods, a possible reason might be that the addition of visit-level attention weakens the contribution of time embeddings.

Risk Factors

As mentioned earlier, the factors identified by this method can be roughly divided into possible risk factors and highly associated factors, for which some pieces of evidence can be found in the literature. Besides, there were still several candidate factors proposed by our model that were seldom reported, for example, HIV infection, or we could not confirm their associations, for example, abdominal pain. One possible reason is that we only considered structured data but not textual information (ie, clinical notes); therefore, that disease or symptom may not be detailed enough to understand given only a code (ie, we know abdominal pain but do not know in which part). Furthermore, according to the AUC values of the model, the results may not be precise enough and still need to be improved. Overall, our method is completely data driven, without any predefined candidate risk factors by experts, which is different from most studies based on regression analysis [7]. We expect that our method can provide compensational information and some new findings can be further validated by clinicians or researchers.

Error Analysis

We analyzed patient samples that are likely to be false-positives or false-negatives according to the prediction probability (as we did not require an output label but only a probability indicating the risk of each patient). One possible reason for the likely false-negatives (ground truth is case, but the predicted probability for case is quite low) is the data missing problem. For example, patient A had some respiratory symptoms such as asthma, shortness of breath, and chronic airway obstruction about 6 months before the prediction date; however, all the diagnosis codes were related to heart disease and hypertension. Therefore, it is likely that some symptoms that might serve as better indications were missing. On the other hand, one explanation for some false-positives (ground truth is control, but the predicted probability for control is quite low) is the difficulty in evaluating the severity of certain diseases or symptoms. For example, patient B had continuous respiratory symptoms such as chronic airway obstruction, but without any laboratory test values or knowledge of the drugs, it is difficult to determine whether these symptoms worsened or were already well controlled. To mitigate these factors, it is desirable to integrate more variables and background knowledge into the model in the future.

Comparison With Prior Work

One of the advantages of machine learning over statistical analysis is that it can make predictions on unseen samples [15], and it might be much easier to be deployed in real-world applications. Although many studies have focused on asthma exacerbation prediction, the majority of them belong to statistics, as they did not test their model on held-out data sets [68,69]. Among other machine learning studies, multiple conventional models have been explored, such as classification and regression tree [11,70], random forest [71], LR [72], and support vector machines [16]. However, none of these previous studies used deep learning as we know. Compared with these conventional machine learning methods, our deep learning–based method has multiple advantages. First, no feature engineering is needed, which will extremely reduce the laborious cost and expertise at the first step, for example, in comparison, Luo et al [20] included 235 features designed by multiple clinical experts as inputs that might cost a lot, but we input all the clinical codes to the model and kept their original formats without any feature selection. Second, LSTM structures can integrate any temporal patterns; thus, dependencies between variables can be easily modeled. Third, deep learning methods usually obtain better performances compared with conventional machine learning methods because of their capacities in modeling complex data structures [22,23,27], which was also proved in our experiments (compared with LR).

Furthermore, compared with previous studies, we are the first to make comprehensive visualization and personalization over the associated factors. One paper mentioned *personalize*, but it

XSL•FO
RenderX

only discussed it as a future possibility [16]. In comparison, our method showed not only cohort-level factors but also temporal-based personalized risks and factors, which would greatly facilitate precise medicine. Meanwhile, as we did not limit our input to the predefined factors, we were able to find new potential risk factors. However, one drawback of deep learning–based methods compared with the previous shallow methods is the lack of interpretability from some perspectives, for example, they can hardly report statistical evaluation measures such as *P* values and CIs, which might need further exploration.

Limitations and Future Work

Using deep learning, we offered a novel means of identifying possible risk factors and predicting the risk of asthma exacerbation. However, this study has some limitations. First, for the model interpretation part, how multiple clinical variables interact with each other needs further exploration; simply considering each variable independently but ignoring the dependency patterns between them might be insufficient for interpretation, for example, the prescription of a drug might be closely associated with a disease or symptom. Second, structured EHRs have their own drawbacks, such as data irregularity, sparsity, and noise. Thus, some potential risk factors for asthma exacerbations might not be recorded or might even be incorrectly recorded in EHRs. As a result, information integrity cannot be guaranteed. We may need to find ways to make the data complete and more reliable, such as including information from textual reports or patient surveys. Third, it is still difficult for computer programs alone to distinguish between asthma symptoms and risk factors, and knowledge injection is needed in the future. Finally, the performance of the model still has room for improvement. It might be boosted further by designing more powerful structures or including background knowledge.

Conclusions

In this paper, we proposed an attentive deep learning–based model for asthma exacerbation prediction and employed elapsed time embeddings to model the time decays. By leveraging the weights of the model, we not only generated personalized heatmaps and specific risk scores at the individual level but also identified possible risk factors at the cohort level. Compared with previous studies, our model is effective in modeling time information and obtains better overall AUCs. As the model is completely data driven and relies little on feature engineering, it can easily be generalized to other prediction tasks. To the best of our knowledge, this is the first study to predict asthma exacerbation risks using a deep learning model that includes elapsed time embeddings. Some of the top-ranked risk factors identified have gained supporting evidence from previous medical studies, which proved that our method has good reliability and accuracy.

Acknowledgments

The authors thank Dr Irmgard Willcockson for proofreading. In addition, the authors thank Cerner for providing the valuable Health Facts electronic medical record data. The authors gratefully acknowledge the support of Nvidia Corporation with the donation of the Quadro P6000 and Titan XP GPUs used for this research.

This research was partially supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011829, the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number 1R01AI130460, National Center for Advancing Translational Sciences of the National Institutes of Health under award number U01TR02062, and the Cancer Prevention Research Institute of Texas Training Grant #RP160015.

Authors' Contributions

CT conceived the research project. YX, HJ, and CT designed the pipeline and method. YX implemented the deep learning model of the study and prepared the manuscript. HJ completed the clinical part of the manuscript. WZ and HX provided valuable suggestions on cohort selection and experimental design. Y Zhou and Y Zhang extracted, cleaned the data, and performed statistics. LR helped to reorganize the data and performed normalizations for the revised version. FL, JD, SW, DZ, and CT proofread the paper and provided valuable suggestions. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 145 KB - jmir_v22i7e16981_app1.docx]

References

1. Asthma. World Health Organization. 2017. URL: http://www.who.int/mediacentre/factsheets/fs307/en/ [accessed 2020-06-29]
2. Most Recent Asthma Data. Center for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/asthma/most_recent_data.htm [accessed 2020-06-29]
3. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. Ann Am Thorac Soc 2018 Mar;15(3):348-356. [doi: 10.1513/AnnalsATS.201703-259OC] [Medline: 29323930]

XSL•FO
RenderX

4. Pocket Guide for Asthma Management and Prevention. Global Initiative for Asthma. 2018. URL: https://ginasthma.org/wp-content/uploads/2018/03/wms-GINA-main-pocket-guide_2018-v1.0.pdf [accessed 2020-06-29]
5. Wark PA, Gibson PG. Asthma exacerbations 3: pathogenesis. Thorax 2006 Oct;61(10):909-915 [FREE Full text] [doi: 10.1136/thx.2005.045187] [Medline: 17008482]
6. Levy ML. The national review of asthma deaths: what did we learn and what needs to change? Breathe (Sheff) 2015 Mar;11(1):14-24 [FREE Full text] [doi: 10.1183/20734735.008914] [Medline: 26306100]
7. Fleming L. Asthma exacerbation prediction: recent insights. Curr Opin Allergy Clin Immunol 2018 Apr;18(2):117-123. [doi: 10.1097/ACI.0000000000000428] [Medline: 29406359]
8. Azizpour Y, Delpisheh A, Montazeri Z, Sayehmiri K, Darabi B. Effect of childhood BMI on asthma: a systematic review and meta-analysis of case-control studies. BMC Pediatr 2018 Apr 26;18(1):143 [FREE Full text] [doi: 10.1186/s12887-018-1093-z] [Medline: 29699517]
9. Martin A, Bauer V, Datta A, Masi C, Mosnaim G, Solomonides A, et al. Development and validation of an asthma exacerbation prediction model using electronic health record (EHR) data. J Asthma 2019 Aug 8:1-8. [doi: 10.1080/02770903.2019.1648505] [Medline: 31340688]
10. Kimura H, Konno S, Makita H, Taniguchi N, Shimizu K, Suzuki M, Hi-CARAT investigators. Prospective predictors of exacerbation status in severe asthma over a 3-year follow-up. Clin Exp Allergy 2018 Sep;48(9):1137-1146. [doi: 10.1111/cea.13170] [Medline: 29781543]
11. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. Am J Respir Crit Care Med 1998 Apr;157(4 Pt 1):1173-1180. [doi: 10.1164/ajrccm.157.4.9708124] [Medline: 9563736]
12. Stanford RH, Nagar S, Lin X, O'Connor RD. Use of ICS/LABA on asthma exacerbation risk in patients within a medical group. J Manag Care Spec Pharm 2015 Nov;21(11):1014-1019 [FREE Full text] [doi: 10.18553/jmcp.2015.21.11.1014] [Medline: 26521113]
13. Blaiss MS, Steven GC, Bender B, Bukstein DA, Meltzer EO, Winders T. Shared decision making for the allergist. Ann Allergy Asthma Immunol 2019 May;122(5):463-470 [FREE Full text] [doi: 10.1016/j.anai.2018.08.019] [Medline: 30201469]
14. Loymans RJ, Debray TP, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Schermer TR, et al. Exacerbations in adults with asthma: a systematic review and external validation of prediction models. J Allergy Clin Immunol Pract 2018;6(6):1942-52.e15. [doi: 10.1016/j.jaip.2018.02.004] [Medline: 29454163]
15. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018 Apr;15(4):233-234 [FREE Full text] [doi: 10.1038/nmeth.4642] [Medline: 30100822]
16. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. Ann N Y Acad Sci 2017 Jan;1387(1):153-165 [FREE Full text] [doi: 10.1111/nyas.13218] [Medline: 27627195]
17. Dexheimer J, Brown L, Leegon J, Aronsky D. Comparing decision support methodologies for identifying asthma exacerbations. Stud Health Technol Inform 2007;129(Pt 2):880-884. [Medline: 17911842]
18. Sanders D, Aronsky D. Detecting asthma exacerbations in a pediatric emergency department using a Bayesian network. AMIA Annu Symp Proc 2006:684-688 [FREE Full text] [Medline: 17238428]
19. Spyroglou II, Spöck G, Rigas AG, Paraskakis EN. Evaluation of Bayesian classifiers in asthma exacerbation prediction after medication discontinuation. BMC Res Notes 2018 Jul 31;11(1):522 [FREE Full text] [doi: 10.1186/s13104-018-3621-1] [Medline: 30064478]
20. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. JMIR Med Inform 2020 Jan 21;8(1):e16080 [FREE Full text] [doi: 10.2196/16080] [Medline: 31961332]
21. Bae S, Choi I, Kim N. Acoustic Scene Classification Using Parallel Combination of LSTM and CNN. In: Detection and Classification of Acoustic Scenes and Events. 2016 Presented at: DCASE'16; September 3, 2016; Budapest, Hungary URL: http://www.cs.tut.fi/sgn/arg/dcase2016/documents/challenge_technical_reports/DCASE2016_Bae_1025.pdf
22. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
23. Baytas I, Xiao C, Zhang X, Wang F, Jain A, Zhou J. Patient Subtyping via Time-Aware LSTM Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: 10.1145/3097983.3097997]
24. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 1;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
25. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
26. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: 10.1145/3097983.3098088]

XSL•FO
RenderX

27. Choi E, Bahadori M, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. ArXiv 2016:- epub ahead of print(1608.05745) [FREE Full text]
28. Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access 2019;7:134236-134244. [doi: 10.1109/access.2019.2928579]
29. Xu Y, Biswal S, Deshpande S, Maher K, Sun J. RAIM: Recurrent Attentive and Intensive Model of Multimodal Patient Monitoring Data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD'18; August 19-23, 2018; London, UK. [doi: 10.1145/3219819.3220051]
30. Wu S, Liu S, Sohn S, Moon S, Wi C, Juhn Y, et al. Modeling asynchronous event sequences with RNNs. J Biomed Inform 2018 Jul;83:167-177 [FREE Full text] [doi: 10.1016/j.jbi.2018.05.016] [Medline: 29883623]
31. Huang Y, Yang X, Xu C. Time-Guided High-Order Attention Model of Longitudinal Heterogeneous Healthcare Data. In: Pacific Rim International Conference on Artificial Intelligence. 2019 Presented at: PRICAI'19; August 23-30, 2019; Cuvu, Yanuka Island, Fiji. [doi: 10.1007/978-3-030-29908-8_5]
32. Zhang Y, Yang X, Ivy J, Chi M. ATTAIN: Attention-based Time-Aware LSTM Networks for Disease Progression Modeling. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligen. 2019 Presented at: IJCAI'19; August 10-16, 2019; Vienna, Austria. [doi: 10.24963/ijcai.2019/607]
33. Graves A, Mohamed A, Hinton G. Speech Recognition With Deep Recurrent Neural Networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing. 2013 Presented at: ICASSP'13; May 26-31, 2013; Vancouver, BC, Canada. [doi: 10.1109/icassp.2013.6638947]
34. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017 Oct 1;89:248-255. [doi: 10.1016/j.compbiomed.2017.08.015] [Medline: 28843829]
35. Do QT, Doig AK, Son TC. Deep Q-learning for predicting asthma attack with considering personalized environmental triggers' risk scores. Conf Proc IEEE Eng Med Biol Soc 2019 Jul;2019:562-565. [doi: 10.1109/EMBC.2019.8857172] [Medline: 31945961]
36. Vaswani A, Shazeer N, Parmar N. Attention Is All You Need. NIPS Proceedings. 2017. URL: https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf [accessed 2020-07-14]
37. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation Classification via Convolutional Deep Neural Network. In: 25th International Conference on Computational Linguistics: Technical Papers. 2014 Presented at: COLING'14; August 23-29, 2014; Dublin, Ireland.
38. Amazon Simple Storage Service Endpoints and Quotas - AWS. URL: https://s3-us-west-2 [accessed 2020-06-29]
39. Song H, Rajan D, Thiagarajan J, Spanias A. Attend and diagnose: clinical time series analysis using attention models. ArXiv 2018:- epub ahead of print [FREE Full text]
40. Nair V, Hinton G. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning. 2010 Presented at: ICML'10; June 21-24, 2010; Haifa, Israel URL: https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf
41. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. ArXiv 2015:- epub ahead of print [FREE Full text]
42. Hinton G, Srivastava N, Swersky K. Neural networks for machine learning. Coursera, video lectures 2012;264:1. [doi: 10.1002/da.20889] [Medline: 22307921]
43. Kingma D, Ba J. Adam: a method for stochastic optimization. ArXiv 2014:- epub ahead of print [FREE Full text]
44. Abadi M, Agarwal A, Barham P. Tensorflow: large-scale machine learning on heterogeneous distributed systems. ArXiv 2016:- epub ahead of print [FREE Full text]
45. Ioffe S, Szegedy C. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32 nd International Conference on Machine Learning. URL: http://proceedings.mlr.press/v37/ioffe15.pdf [accessed 2020-07-14]
46. Ghosh N, Choudhury P, Subramani E, Saha D, Sengupta S, Joshi M, et al. Metabolomic signatures of asthma-COPD overlap (ACO) are different from asthma and COPD. Metabolomics 2019 Jun 04;15(6):87. [doi: 10.1007/s11306-019-1552-z] [Medline: 31165288]
47. Naik RD, Vaezi MF. Extra-esophageal gastroesophageal reflux disease and asthma: understanding this interplay. Expert Rev Gastroenterol Hepatol 2015 Jul;9(7):969-982. [doi: 10.1586/17474124.2015.1042861] [Medline: 26067887]
48. Solidoro P, Patrucco F, Fagoonee S, Pellicano R. Asthma and gastroesophageal reflux disease: a multidisciplinary point of view. Minerva Med 2017 Aug;108(4):350-356. [doi: 10.23736/S0026-4806.17.05181-3] [Medline: 28378957]
49. Peng YH, Chen KF, Liao WC, Hsia TC, Chen HJ, Yin MC, et al. Association of migraine with asthma risk: a retrospective population-based cohort study. Clin Respir J 2018 Mar;12(3):1030-1037. [doi: 10.1111/crj.12623] [Medline: 28268255]
50. Dirican N, Demirci S, Cakir M. The relationship between migraine headache and asthma features. Acta Neurol Belg 2017 Jun;117(2):531-536. [doi: 10.1007/s13760-017-0764-0] [Medline: 28258562]
51. Savelloni J, Gunter H, Lee KC, Hsu C, Yi C, Edmonds KP, et al. Risk of respiratory depression with opioids and concomitant gabapentinoids. J Pain Res 2017;10:2635-2641 [FREE Full text] [doi: 10.2147/JPR.S144963] [Medline: 29180889]
52. Parmar MS. Exacerbation of asthma secondary to fentanyl transdermal patch. BMJ Case Rep 2009;2009:- [FREE Full text] [doi: 10.1136/bcr.10.2008.1062] [Medline: 21686475]

XSL•FO
RenderX

53. Henderson AJ, Shaheen SO. Acetaminophen and asthma. Paediatr Respir Rev 2013 Mar;14(1):9-15; quiz 16. [doi: 10.1016/j.prrv.2012.04.004] [Medline: 23347656]
54. Web MD. URL: https://www.webmd.com/asthma/news/20091105/acetaminophen-may-be-linked-to-asthma-risk#1
55. Morales DR, Guthrie B, Lipworth BJ, Jackson C, Donnan PT, Santiago VH. NSAID-exacerbated respiratory disease: a meta-analysis evaluating prevalence, mean provocative dose of aspirin and increased asthma morbidity. Allergy 2015 Jul;70(7):828-835. [doi: 10.1111/all.12629] [Medline: 25855099]
56. Bai TR, Vonk JM, Postma DS, Boezen HM. Severe exacerbations predict excess lung function decline in asthma. Eur Respir J 2007 Sep;30(3):452-456 [FREE Full text] [doi: 10.1183/09031936.00165106] [Medline: 17537763]
57. Mapping Between ICD-10 and ICD-9. Ministry of Health NZ. URL: https://www.health.govt.nz/nz-health-statistics/data-references/mapping-tools/mapping-between-icd-10-and-icd-9 [accessed 2019-01-01]
58. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. ArXiv 2014:- epub ahead of print [FREE Full text]
59. Xiang Y, Chen Q, Wang X, Qin Y. Answer selection in community question answering via attentive neural networks. IEEE Signal Process Lett 2017 Apr;24(4):505-509. [doi: 10.1109/lsp.2017.2673123]
60. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical Attention Networks for Document Classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies at: Proceedings of the. 2016 Presented at: ACL'16; June 12-17, 2016; San Diego, California. [doi: 10.18653/v1/n16-1174]
61. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
62. Zia T, Zahid U. Long short-term memory recurrent neural network architectures for Urdu acoustic modeling. Int J Speech Technol 2018 Nov 8;22(1):21-30. [doi: 10.1007/s10772-018-09573-7]
63. Mandic S, Go C, Aggarwal I, Myers J, Froelicher VF. Relationship of predictive modeling to receiver operating characteristics. J Cardiopulm Rehabil 2008;28(6):415-419. [doi: 10.1097/hcr.0b013e31818c3c78]
64. Hosmer JD, Lemeshow S, Sturdivant R. Applied Logistic Regression. New York, USA: John Wiley & Sons; 2013.
65. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 1;16:321-357. [doi: 10.1613/jair.953]
66. Pal S, Mitra S. Multilayer perceptron, fuzzy sets, and classification. IEEE Trans Neural Netw 1992;3(5):683-697. [doi: 10.1109/72.159058] [Medline: 18276468]
67. GitHub. URL: https://github.com/o0windseed0o/asthma_prediction [accessed 2020-07-14]
68. Sato R, Tomita K, Sano H, Ichihashi H, Yamagata S, Sano A, et al. The strategy for predicting future exacerbation of asthma using a combination of the asthma control test and lung function test. J Asthma 2009 Sep;46(7):677-682. [doi: 10.1080/02770900902972160] [Medline: 19728204]
69. Eisner MD, Yegin A, Trzaskoma B. Severity of asthma score predicts clinical outcomes in patients with moderate to severe persistent asthma. Chest 2012 Jan;141(1):58-65. [doi: 10.1378/chest.11-0020] [Medline: 21885725]
70. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty of half full? J Asthma 1999 Jun;36(4):359-370. [doi: 10.3109/02770909909068229] [Medline: 10386500]
71. Xu M, Tantisira KG, Wu A, Litonjua AA, Chu J, Himes BE, et al. Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers. BMC Med Genet 2011 Jun 30;12:90 [FREE Full text] [doi: 10.1186/1471-2350-12-90] [Medline: 21718536]
72. Loymans RJ, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Assendelft WJ, Schermer TRJ, et al. Identifying patients at risk for severe exacerbations of asthma: development and external validation of a multivariable prediction model. Thorax 2016 Sep;71(9):838-846. [doi: 10.1136/thoraxjnl-2015-208138] [Medline: 27044486]

Abbreviations

ALSTM: attention long short-term memory
ATTAIN: Attention-based Time-aware Disease Progression
AUC: area under the receiver operating curve
EHR: electronic health record
ICD: International Classification of Disease Code
ICS: inhaled corticosteroids
LR: logistic regression
LSTM: long short-term memory
RETAIN: Reverse Time Attention model
RNN: recurrent neural network
SABA: short-acting beta agonist
TLSTM: time-aware long short-term memory

XSL•FO
RenderX

TSANN: time-sensitive attentive neural network

Edited by G Eysenbach; submitted 09.11.19; peer-reviewed by Y Hafeez, S Chen, X Liu; comments to author 30.12.19; revised version received 02.03.20; accepted 13.05.20; published 31.07.20.

Please cite as:
Xiang Y, Ji H, Zhou Y, Li F, Du J, Rasmy L, Wu S, Zheng WJ, Xu H, Zhi D, Zhang Y, Tao C
Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study
J Med Internet Res 2020;22(7):e16981
URL: https://www.jmir.org/2020/7/e16981
doi:10.2196/16981
PMID:32735224

©Yang Xiang, Hangyu Ji, Yujia Zhou, Fang Li, Jingcheng Du, Laila Rasmy, Stephen Wu, W Jim Zheng, Hua Xu, Degui Zhi, Yaoyun Zhang, Cui Tao. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Usability Evaluation of a Facebook-Based Intervention Program for Childhood Cancer Patients: Mixed Methods Study

Bu Kyung Park[1], PhD; Ji Yoon Kim[2], PhD; Valerie E Rogers[3], PhD

[1]College of Nursing, Research Institute of Nursing Science, Kyungpook National University, Daegu, Republic of Korea

[2]Department of Pediatrics, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[3]School of Nursing (retired), University of Maryland Baltimore, Baltimore, MD, United States

Corresponding Author:
Bu Kyung Park, PhD
College of Nursing
Research Institute of Nursing Science
Kyungpook National University
680 Gukchaegosang-ro, Jung-gu
Daegu, 41944
Republic of Korea
Phone: 82 53 420 4929
Fax: 82 53 421 2758
Email: bukpark@knu.ac.kr

Abstract

Background: Childhood cancers previously considered to be incurable now have 5-year survival rates up to 84%. Nevertheless, these patients remain at risk of morbidity and mortality from therapy-related complications. Thus, patient education and self-management strategies for promoting a healthy lifestyle are of tantamount importance for improving short- and long-term health outcomes. A Facebook-based “Healthy Teens for Soaam” (a Korean term for childhood cancers) program was developed to help improve knowledge and self-management practices of teens with cancer related to their disease and treatment.

Objective: The two-fold purpose of this usability study was (1) to describe the process of developing an 8-week Facebook-based intervention program for teens with cancer, and (2) to evaluate its usability to refine the program.

Methods: Multiple phases and methods were employed to develop and evaluate the usability of the program. Study phases included: (1) needs assessment through focus group interviews and qualitative content analysis, (2) development of module content, (3) expert review and feedback on module content, (4) Facebook-based program development, (5) usability evaluation by heuristic evaluation, (6) usability evaluation by targeted end-user testing, and (7) modification and final version of the program. Usability of the final version was confirmed through feedback loops of these phases.

Results: Based on 6 focus group discussion sessions, it was determined that teens with cancer were interested in seeing stories of successful childhood cancer cases and self-management after discharge, and preferred multimedia content over text. Therefore, each Facebook module was redesigned to include multimedia materials such as relevant video clips tailored for teens. Usability assessed by heuristic evaluation and user testing revealed several critical usability issues, which were then revised. Potential end users tested the final program and perceived it to be usable and useful for teens with cancer.

Conclusions: To our knowledge, “Healthy Teens for Soaam” is the first Facebook-based intervention program for teens with cancer. We actively worked with current childhood cancer patients and survivors to develop and improve this program, achieved good usability, and met the expressed needs and preferences of target end users. This 8-week Facebook-based educational program for teens with cancer, developed as the first step of an upcoming intervention study, will be useful for improving knowledge and self-management strategies of teens.

(J Med Internet Res 2020;22(7):e18779) doi:10.2196/18779

KEYWORDS

pediatric cancer patients; childhood cancer; social network site; Facebook; usability

XSL•FO
RenderX

Introduction

Survival rates associated with childhood cancers continue to improve, partly due to advances in diagnostic techniques and treatment modalities along with clinical research [1]. According to the Surveillance, Epidemiology, and End Results Cancer Statistics Review, the 5-year survival rate of children with leukemia 0-19 years of age in the United States is 84.1% and the 5-year survival rate of patients with childhood brain and central nervous system cancers is 74.8% [2]. In South Korea, the 5-year survival rate among patients with childhood cancers overall was 81.7% in 2014 [3].

Despite improvements in survival, childhood cancer survivors still face a high risk of therapy-related complications or adverse effects that persist or arise after completion of treatment [4]. Childhood cancer survivors are at significantly increased risk of relapse, second malignancy, and long-term late effects, including cardiovascular and pulmonary dysfunction, endocrine disorders such as metabolic syndrome, and others [5,6]. Specifically, childhood cancer survivors of 5 years or greater have at least one chronic health condition [7]. A study on the late effects of childhood cancer treatment found that the cumulative incidence of a chronic health condition among long-term childhood cancer survivors was 99.9% by 50 years of age, with a nearly 2-fold greater cumulative burden among survivors compared to matched community controls [5].

Some morbidities are modifiable through preventive healthy behaviors such as physical activity, good nutrition, obesity prevention, and avoidance or cessation of smoking, as well as identification of characteristics that may modify morbidities or adherence to healthy behaviors to better target interventions [8,9]. For example, one study showed that childhood cancer survivors of older age and lower socioeconomic status were less frequently engaged in preventive healthy behaviors [10]. Another study reported that a greater proportion of female survivors smoked compared to teens without cancer [11]. Although childhood cancer survivors have been found to engage in unhealthy behaviors (eg, tobacco, alcohol, drug use, and sexual behaviors) and to be nonadherent to national health behavior guidelines at rates similar to those of their healthy siblings and teens without cancer [12,13], these behaviors are likely to be more consequential for individuals with organ damage secondary to cancer treatment. Moreover, late effects of treatment such as obesity, cancer-related pain, and sensory impairments have been significantly associated with increased risk of comorbid symptoms [14]. Increased comorbidities are associated with decreased quality of life, and with an increased risk of hospitalization and mortality [15]. Thus, the importance of patient education and promoting self-care for a healthy lifestyle for teens with cancer is increasingly becoming recognized to help mitigate complications of cancer and its treatment.

Children currently undergoing cancer treatment, as well as survivors, require follow-up care for the rest of their lives; thus, the Children's Oncology Group (COG) has emphasized the importance of regular medical follow-up and has developed the "COG long-term follow-up guidelines for survivors of childhood, adolescent, and young adult cancers" [16]. In addition, various interventions to improve long-term outcomes for teens with cancer have been developed worldwide. For example, the Nursing Discipline of the COG has developed key principles and recommendations for patient and family education practices [17] as well as interventions for teens with cancer, such as nutrition and cooking workshops [18], nutrition and body weight changes [19], systematic intervention for psychological preparation for radiotherapy treatment [20], mobile health intervention to improve adherence and quality of life [21], and a web-based physical activity intervention [22]. However, few intervention studies involving teens with cancer have been conducted in Korea. Existing examples include web-based patient safety education [23], an art intervention for siblings of children with cancers [24], and educational interventions to enhance adherence to prophylactic treatment [25]. Intervention programs for preventing long-term late effects of childhood cancers and their treatment are glaringly absent. In particular, there is a crucial need for interventions that promote self-management strategies for achieving a healthy lifestyle among teens with cancer. This is important, as they will likely survive into adulthood and need to develop the knowledge and skills necessary to prevent or mitigate the late effects of cancers, and thereby improve their health-related quality of life.

Knowledge about their disease, the importance of treatment adherence, and strategies to improve adherence could be improved through innovative health interventions designed specifically for teens [21]. Among teens with cancer, inadequate information was identified as one of the barriers to compliance with chemotherapy [26]. Adult survivors of childhood cancers were found to lack detailed knowledge of their treatment history and risk for late effects [27]. Therefore, further efforts are needed to educate and empower teens with cancer to gain appropriate knowledge and assume responsibility for their health management [28].

Social media has been increasingly utilized as a platform for delivering health interventions in childhood cancer care. For example, Watson [29] reported that oncology health care professionals utilized social media to listen, learn, engage, and cocreate to advance cancer care. Teens with cancer have also turned to social media for information about cancer or to interact with peers and others about their diagnosis and its impact [30]. Importantly, the effectiveness of using social media platforms such as Facebook has been reported for this tech-savvy generation. For example, a previous study [31] found that a randomized trial of a Facebook-based physical activity intervention among young adult cancer survivors significantly increased their physical activity. Therefore, we developed an educational program for teens with cancer on Facebook called "Healthy Teens for *Soaam*" (*soaam* is the word for pediatric cancers in Korean), as the first step of an intervention that will deliver educational materials with the aim of increasing knowledge and improving self-care practices related to their disease and treatment. The purpose of this usability study was (1) to describe the process of developing this 8-week modular intervention program for teens with cancer and (2) to evaluate its usability to refine the program.

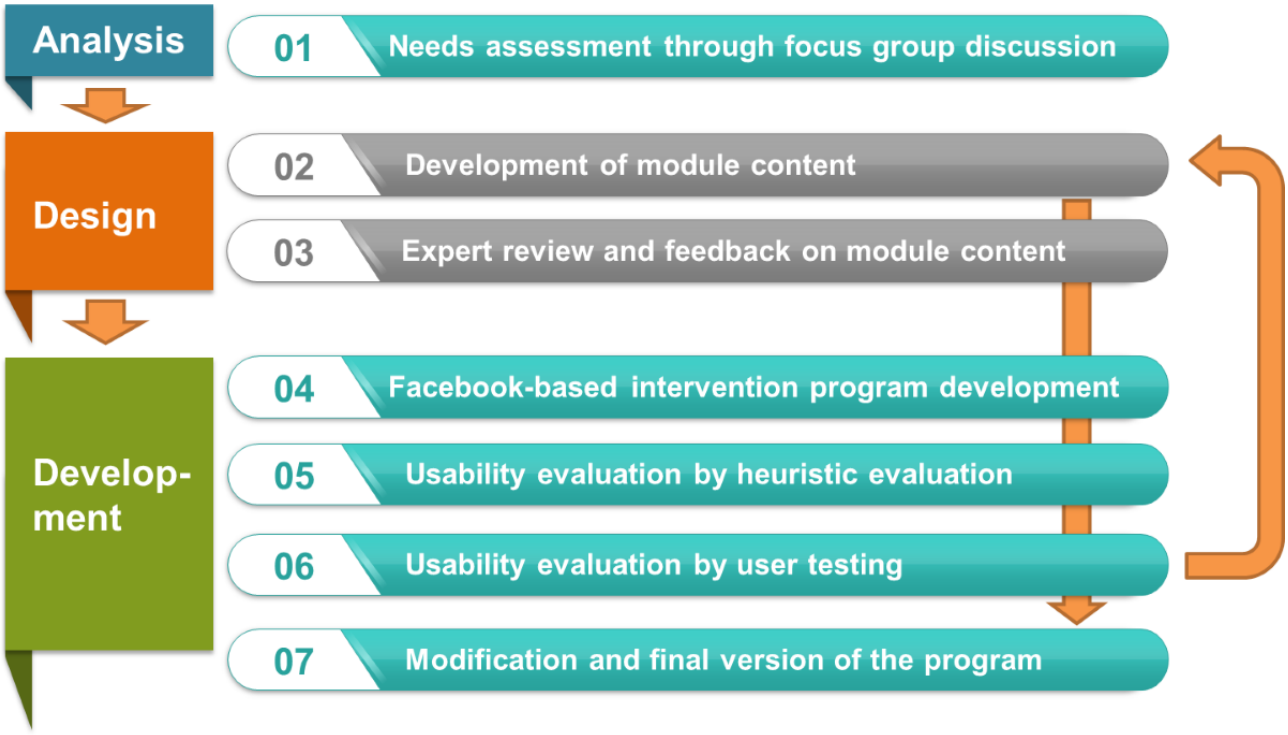
XSL•FO
RenderX

Methods

Study Design

This study was reviewed and approved by the principal investigator's institutional review boards (1041078-201606-HRSB-122-01 and 20200026). The study was conducted in 7 phases (Figure 1) and was guided by the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) instructional design approach [32]. This article describes the first 3 phases of the ADDIE approach.

Figure 1. Flow chart of "Healthy Teens for Soaam" program development.



Phase 1: Needs Assessment Through Focus Group Discussion and Qualitative Content Analysis

Semistructured focus group discussions were conducted with 12 teens with cancer. Participants were recruited from childhood cancer self-support groups in Korea using convenience and snowball sampling methods. A flyer was posted to the online platforms of self-support groups on popular Korean online platforms (Naver Café, KakaoTalk, and Band), and interested volunteers contacted the researchers by phone or email. Interested participants also suggested other potential participants to the researcher.

A total of 6 in-person focus group discussion sessions were conducted, during which participants were accompanied by their parent or legal guardian. Each session included 2 to 5 participants, and each lasted for about 60 to 90 minutes. Focus group discussions were facilitated by the principal investigator (BP) and a research assistant. First, the principal investigator explained the overview of the program, distributed handout materials detailing program contents, and asked participants and parents to provide their opinions regarding the contents. For example, they were explicitly prompted to provide suggestions regarding the addition, removal, or emphasis of content. Besides answering questions, participants and parents were encouraged to share their personal experiences with childhood cancer using prompts such as "If you have anything you want to share with us even though it is not directly related to the contents, don't hesitate to tell us." Focus group discussions were recorded and transcribed verbatim for qualitative content analysis using NVivo 11 (QRS International, Burlington, MA, USA).

Focus Group Discussion Data Analysis

Qualitative content analysis was conducted using a combination of the content analysis method suggested by Krippendorff [33] and the inductive coding approach suggested by Elo and Kyngäs [34]. Coders read the transcripts and freely generated themes. They then developed categories from similar themes as the coding progressed. Major coding rules included the following: (1) a category was created if a list of at least three themes was generated under that category, and (2) multiple sentences from the same participant that referred to the same content were coded as one unit.

Qualitative content analysis was conducted by the two coders (the principal investigator [BP] with a PhD in nursing and the research assistant with a master's degree in nursing) to ensure reliability of the analysis. Coders independently analyzed the transcripts. Questions or disagreements regarding the themes or categories were discussed. Revision and refinement of themes and categories continued until content analysis was completed.

Based on results from focus group discussions, the program module content was organized using the following usability design methods. First, we used group information together (also known as card sorting) after the focus group discussions. We defined the important themes from the focus group discussions

XSL•FO
RenderX

and grouped them into several categories such as diagnosis, examination, treatment, social support, and return to society/school. Second, we developed a structure for 8 weekly modules. Third, we developed low-fidelity module prototypes such as paper prototypes. Finally, we established high-fidelity mockups using PowerPoint slides.

Phase 2: Development of Module Content

Main topics for the program module content were developed based on the results from the focus group discussion "needs assessment." Module content was written at a sixth-grade reading level using the Flesch–Kincaid readability test guidelines [35,36]. Content was developed with guidance from published research articles, medical and nursing textbooks, information available from health professional organization websites (eg, Korean Association for Children with Leukemia [37]), patient education materials in hospitals, and consultations with a panel of experts.

We incorporated usability principles for teen-friendly websites [38] into the program prototype in terms of content, appearance, and navigation. For example, we developed the content using PowerPoint 2016 (Microsoft Corp, Redmond, WA, USA) slides visualized on a single page without scrolling [38,39]. To improve appearance, we used large font sizes and avoided garish color schemes [38,39]. Moreover, slides for each module had their own theme to distinguish them from other modules. Although Facebook posts appear in reverse chronological order (ie, the most recent post appears on the top of users' News Feeds), this was not considered a potential problem for participants because the module content was designed to be updated weekly and potential target participants were familiar with Facebook navigation.

Each module included specific learning objectives, structured learning material for fundamentally important content with external web links for further information, multimedia such as video clips or games relevant to the topics, and discussion topics at the end of the modules. Participants were asked to respond to discussion topics by adding comments on Facebook. This served to monitor module participation and sharing of personal experiences for social support. Although not done during testing, in the implementation study, a program moderator (researcher) will provide feedback and encourage interactions among participants on discussion boards. Discussions included the following topics: "Let's talk about blood transfusion experiences. Have you ever received a blood transfusion? If yes, please tell us about the specific blood product you received, your feelings, and side effects," and "Have you had any pain while getting treatment in the hospital? Which one was the most painful, and how did you overcome the pain—pain medication, massage, and ice bag?"

Phase 3: Expert Review and Feedback on Module Content

The module content was reviewed by four experts currently working in pediatric hematology and oncology units in tertiary hospitals, who each had more than 10 years of experience: two nurses with master's degrees in pediatric nursing and two pediatric hematology and oncology physicians with PhDs. Reviewers were knowledgeable about the most recent evidence in this highly specialized field and reviewed the module contents for accuracy. They were asked to rate the ease of understanding and appropriateness of information for teens with cancer (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree). Based on their feedback, module contents were revised and finalized for mockups of the Facebook-based program.

Phase 4: Facebook-Based Intervention Program Development

Facebook Group was used to develop the 8-week intervention program, with the "private group" functionality utilized to protect participants' privacy. With a private group, only people who are invited by the group creator (principal investigator of this study) can join and see who is in the group and what they post [40]. We chose the private group option because, as established by the focus group discussions, many participants did not want to be widely known as cancer patients. Additionally, we customized the group's privacy options [41], for example, by selecting the "hide group" and "membership approval" options so that only members could find the group and that only the group creator could approve new members.

To deliver module content, we composed PowerPoint slides and then uploaded each slide to the group's Photos section. Participants could easily navigate the slides by clicking the "next" and "back" buttons. Considering that the purpose of this program is to deliver accurate and reliable educational content, only group administrators were granted posting permission, but participants were allowed to add comments to each post to share their opinions or contribute to discussions.

Phase 5: Usability Evaluation by Heuristic Evaluation

Two experts in the field of usability and human-computer interactions reviewed the Facebook-based program using Nielsen's heuristics [42,43] adapted and tailored for children's electronic learning (eLearning) program evaluation [44]. The two usability experts have PhDs in health care informatics and were currently teaching and conducting research on health care informatics in universities, one for 17 years and the other for 5 years. The criteria of Nielsen's heuristics tailored for eLearning in children were: (1) visibility of system status; (2) match between system and the real world; (3) user control and freedom; (4) consistent and standards; (5) error prevention; (6) recognition rather than recall; (7) flexibility and efficiency of use; (8) esthetic and minimalist design; (9) help users recognize, diagnose, and recover from errors; (10) help and documentation; (11) design attractive screen layout; (12) use appropriate hardware devices; (13) challenge the child; (14) evoke child mental imagery; and (15) support child curiosity [44]. Comments from the heuristic evaluation were analyzed employing content analysis using the inductive coding approach described in Phase 1. The program was revised according to feedback from the heuristic evaluation.

Phase 6: Usability Evaluation by User Testing

User testing was conducted using observation, the think-aloud method, voice and screen activity capture using Camtasia 9 (TechSmith, Okemos, MI, USA), and surveys. Participants for this phase were recruited through parents of teens belonging to

XSL•FO
RenderX

a childhood cancer self-support group in Korea using convenience and snowball sampling. A flyer was posted to the online platforms of self-support groups on KakaoTalk and Band, and interested volunteers contacted researchers by phone or email. Inclusion criteria were teens with cancer who were (1) aged 13 to 18 years old, (2) diagnosed with any type of childhood cancer, (3) received any type of cancer treatment (eg, radiation therapy, chemotherapy, stem cell transplant), and (4) at any stage of treatment (newly diagnosed through completed treatment). Participants arranged appointments for screening, consenting, and user testing. As the participants were adolescents, both informed consent from their parents or guardians and assent from the teens were obtained prior to participation.

Procedures

A total of 11 face-to-face, 1:1 user testing sessions were conducted. Testing was facilitated by the principal investigator in a private room at the study site. Before conducting user testing, the researcher gave instructions about the program and usability evaluation methods such as the think-aloud method and Camtasia screen-capture software. Participants were then provided with an information page that included the Facebook program link, user ID, and password. They were assigned a user testing task to review 2 out of the 8 modules, including logging in, navigating through instructions on the first page to modules, reading content, watching video clips, and adding comments. They were assured that they could stop at any time for any reason. Each session lasted 30 to 40 minutes.

During evaluation of the program, participants were encouraged to think out loud while their voices were recorded using Camtasia. During pauses from thought verbalization, participants were probed with the following prompts: “What do you think about this? What features do you think need improvement?” At the same time, the researcher documented comments from participants and her own observations using the user testing worksheet.

Measures

Before user testing, participants’ demographic characteristics as well as their internet and Facebook usage habits were queried and documented. After user testing, participants filled out a short survey that evaluated electronic health (eHealth) literacy and perceived usability. The eHealth literacy was assessed using the eHealth Literacy Scale (eHEALS) [45], which has 8 items and assesses an individual’s knowledge, comfort, and perceived skills for locating, evaluating, and applying eHealth information for health issues. The Korean version of eHEALS, known as K-eHEALS, was previously tested for reliability and validity [46]. Total eHEALS scores range from 8 to 40, with a higher score indicating better health information literacy. Cronbach α of the original scale was .88 [45] and was also .88 for the Korean version [46]. Cronbach α for this study was .71.

Perceived usability of the program was assessed using the Perceived Health Website Usability Questionnaire (PHWSUQ) [47]. This questionnaire has 12 items scored using a 7-point Likert scale. It measures three dimensions of usability: satisfaction, ease of use, and usefulness. Cronbach α values for the subscales ranged from .64 to .93, and was .85 for this study.

Data Analysis

Quantitative data from the questionnaires were analyzed with descriptive statistics using SPSS Statistics for Windows, version 25.0 (IBM Corp, Armonk, NY, USA). Audio recordings from user testing sessions were transcribed verbatim and analyzed employing the content analysis method described in Phase 1. Qualitative data, collected throughout observation and screen activity and recorded by Camtasia, provided a rich contextual background and a strong source of triangulation for developing themes and generating a comprehensive review of the program’s usability. Thus, screen recordings were reviewed when the transcript did not capture enough usable detail, which improved the accuracy of content analysis.

Phase 7: Modification and Final Version of the Program

Modifications regarding content and appearance issues were made on the PowerPoint slides, and the revised slides were uploaded to Facebook (the process was conducted in the order phase 2 to phase 4 to phase 7). If content revisions required expert opinions, relevant experts were contacted to confirm the accuracy of the new content (the process was conducted in the order phase 2 to phase 3 to phase 4 to phase 7).

Results

Phase 1: Summary of Focus Group Discussions

Participant characteristics are summarized in Table 1 [48].

Participant feedback regarding the outline of the program (Table 2) included (1) a desire for stories of successful cases and self-management at home, (2) a preference for multimedia content (eg, video clips) rather than text, (3) requests for more patient safety–related material, (4) an expressed need for detailed program objectives and instructions, and (5) appeals for more information on treatment, prognosis, and medical terminology. One teen expressed their support for the program as follows:

> *This is the big problem. The reason I’m so supportive of this education program (Healthy Teens for Soaam) is that there’s a limitation of internet and you narrow down the range a little bit….. I hope the contents go on in the future.*

Additionally, participants freely described and shared their experiences with cancer treatments, and the emerging themes and categories dealt with information needs, support systems, barriers to treatment, facilitators to treatment, return to social life, and health care system issues. Other representative comments from the focus group discussions are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Characteristics of focus group discussion participants (N=12).

Characteristic	Value
Age (years), mean (SD)	15.2 (5.3)
Sex, n (%)	
Male	6 (50)
Female	6 (50)
Diagnosis, n (%)	
Leukemia	6 (50)
Brain tumor	2 (17)
Aplastic Anemia	2 (17)
Other	2 (17)
Treatment, n (%)	
Completed with ongoing outpatient follow up	9 (75)
Ongoing chemotherapy	3 (25)

XSL•FO
RenderX

Table 2. Content analysis of focus group discussion transcriptions (N=12).

Theme	Frequency of units
Feedback on the program	
Add story of successful cases and self-management after discharge	9
Prefer multimedia (video clips) than text	10
Add patient safety issues	11
Need detailed program objectives and instructions	13
Add information on treatment, prognosis, medical terms	20
Information needs	
Lack of information or inaccurate information	7
Current sources of information (eg, other patients, hospital handout, searching the internet)	19
Useful information during treatment (personal experience)	20
Support system	
Current support from hospital, school, pediatric cancer associations, self-support groups, etc	10
Need systematic supporting system for patients and families	13
Useful support during treatment (personal experience)	15
Barriers to treatment	
Economic problems	5
Hospital-related issues (eg, lack of available beds for admission, manpower shortage, miscommunications between health care providers)	8
Side effects of treatment (eg, chemotherapy, transfusion, infection)	25
Emotional reactions at the time of diagnosis and during treatment	28
Facilitators to treatment	
Patients' insight on diagnosis	4
Empathy of health care providers	6
Social support from friends, other patients, family members	13
Return to social life	
School life	4
Lack of physical activity	6
Concerns related to infection prevention, weak immunity	10
Health care system issues	
Disabled child registration and welfare benefits	2
Inadequate and poor social welfare system	4
Health insurance	5

Phases 2 and 3: Facebook-Based 8-Week Intervention Program

An overview of the final version of the program is provided in Table 3. Content was revised and updated according to focus group discussion feedback. Participants and parents were interested, for example, in the topics of returning to school and the effects of chemotherapy on the adolescent's fertility; therefore, we added these topics to the modules. Additionally, as participants preferred multimedia over reading text, we searched for movies and Korean dramas depicting stories involving teens with cancer and included links to short video clips in the relevant module content. Animations explaining cancers and their treatment tailored for teenagers were also added for a better understanding of how treatment proceeds.

Four childhood cancer experts reviewed the content and provided feedback on whether the information was accurate and current, including the pictures and video clips from movies and Korean dramas. Two nurses reviewed relevant medical terminology from academic journals and textbooks, and noted that it differed from the terminology commonly used by participants and parents. For example, granulocyte colony-stimulating factor injections for neutropenia were usually referred to as "count shots" (ie, a shot given when the absolute neutrophil count is low) or "Grasin" among participants and

XSL•FO
RenderX

parents. Thus, we also used these words in the program content. Two physicians contributed information on the most recent treatment practices for childhood cancers in tertiary hospitals. For example, a previous version contained only bone marrow transplantation, but we added information about cord blood transplantation and peripheral blood stem cell transplantation.

Table 3. Overview of the Facebook-based Healthy Teens for Soaam program.

Module (Week)	Contents
Introduction	Program purpose, program schedule, how-to-use tutorial video clip
Module 1 (Week 1)	Pediatric cancers characteristics
	-Pediatric cancers statistics (risks and causes of pediatric cancers)
	-Pediatric cancers symptoms, early signs
	-International Childhood Cancer Day (February 15)
	-Psychosocial services and support for children and families
Module 2 (Week 2)	Types of pediatric cancers and their characteristics
	-Bone cancers, brain cancers, leukemia, hepatoblastoma, lymphoma, neuroblastoma, rhabdomyosarcoma, retinoblastoma, Wilms tumor, etc
Module 3 (Week 3)	Diagnostic tests
	-Imaging tests, CT[a] scans, MRI[b], ultrasound, blood tests
	-Bone marrow aspiration and biopsy
Module 4 (Week 4)	Pediatric cancers treatments 1
	-Hickman catheter insertion and management
	-Chemo-port insertion and management
	-Radiation treatment and chemotherapy
Module 5 (Week 5)	Pediatric cancers treatment 2
	-Treatment team and frequently used medical terminology
	-Caring tips for mouth care and oral mucositis
	-Caring tips for infection and bleeding
	-Symptoms of infection, ANC[c]
Module 6 (Week 6)	Pediatric cancers treatment 3
	-Managing side effects of radiation and chemotherapy
	-Pain, transfusion
	-Growth and development, fertility
	-Stem cell transplantation (BMT[d], CBT[e], PBSCT[f])
Module 7 (Week 7)	Back to school and society
	-Follow-up care after cancers treatment
	-Facilitating school reentry guide
	-Nutrition and daily activity
Module 8 (Week 8)	People around me: family and friends
	Review and summary

[a]CT: computed tomography.

[b]MRI: magnetic resonance imaging.

[c]ANC: absolute neutrophil count.

[d]BMT: bone marrow transplantation.

[e]CBT: cord blood transplantation.

[f]PBSCT: peripheral blood stem cell transplantation.

Phases 4 and 5: Revision After Heuristic Evaluation

Usability experts pointed out usability problems of the program among three categories of Nielsen's heuristics tailored for children's eLearning [44]: (3) user control and freedom, (7) flexibility and efficiency of use, and (10) help and documentation. For example, even though Facebook has a Help menu for user support, the experts suggested that there should be a Help section specific to the program. Therefore, we added a "Help and FAQs" section for dealing with problems encountered while using the program. However, some usability issues were not changeable on the ready-made Facebook platform; these included Facebook advertisements and the reverse chronological order of posts.

XSL•FO
RenderX

All four childhood cancer experts rated the content highly for its ease of understanding (4=strongly agree) and appropriateness for teens with cancer (4=strongly agree). They also commended the detailed instructions and the neat and consistent overall design of the PowerPoint slides.

Phases 6 and 7: Revisions After User Testing

User testing participant characteristics (Table 4) included a mean age of 16.7 years, and the majority had leukemia and were Facebook users (91%). Their mean K-eHEALS score suggested good eHealth literacy. The mean PHWSUQ score (reflecting the program's perceived usability) indicated that the program was perceived as usable. All 11 participants completed the user testing tasks and 9 participants were able to navigate through the program with little or no guidance. The mean testing time was 38.8 (SD 3.1) minutes. Representative quotes from user testing are shown in Multimedia Appendix 1.

Table 4. Characteristics of user testing participants (N=11).

Characteristic	Value
Age (years), mean (SD)	16.7 (1.1)
Sex	
Male	7 (64)
Female	4 (36)
Diagnosis	
Leukemia	8 (73)
Other	3 (27)
Facebook user	
Yes	10 (91)
No	1 (9)
K-eHEALS[a], mean (SD)	27.6 (1.3)
PHWSUQ[b], mean (SD)	
Total PHWSUQ	60.1 (2.8)
Satisfaction	32.9 (4.8)
Ease of use	15.4 (2.9)
Usefulness	17.8 (3.4)
Preferred frequency of content update, n (%)	
Every day	1 (9)
2-3 times per week	4 (6)
Once per week	6 (55)

[a]K-eHEALS: Korean version of the eHealth Literacy Scale (range 8-40).

[b]PHWSUQ: Perceived Health Website Usability Questionnaire (range 12-84 total; 6-42 satisfaction; 3-21 ease-of-use and usefulness).

The transcript of user testing content analysis identified 20 themes under four categories: program content, program appearance, navigation, and others, which mainly included feedback about the strengths of the program (Table 5). Participants mentioned that they gained some new knowledge from this program, but some perceived the content as difficult to understand or as having low readability because of the medical terminology. Regarding the appearance of the program, participants generally liked the layout, but three users noted, for example, that some text was hidden behind images or that low-resolution figures were difficult to decipher. Nine participants had no issues with navigation of Facebook, but two participants needed help from the researcher. Seven participants expressed their impressions about the discussion section, where they could share personal experiences with participants who had similar experiences. For example, one teen said: "Then if I don't know the details right now, I think it would be helpful for patients to learn this by Facebook myself."

Critical usability issues reported during user testing were reviewed, and, where necessary, revisions were made and confirmed through feedback loop of the phases (Figure 1). Regarding usability issues with "program content," one participant suggested a movie on a pediatric patient, which was not in our movie list. She mentioned that it had helped her a lot when she underwent treatment. In this case, childhood cancer experts reviewed the movie and included relevant video clips to the module. Some video clips and pictures that were not working or had low graphic resolution were replaced with other video clips and pictures. Usability issues with "program appearance" were mostly minor revisions such as typos, font size, or picture location issues, which were corrected directly on the PowerPoint slides, and then the revised slides were

XSL•FO
RenderX

reuploaded on Facebook. There were few usability issues with "navigation." Participants easily recovered from failed paths, with brief help from the researcher. However, for first-time users of Facebook, we added a short video tutorial of the program navigation. Multimedia Appendix 2 shows a screenshot of the final version of the "Healthy Teens for *Soaam*" program on Facebook.

Table 5. Content analysis of user testing transcriptions (N=11).a

Category, Themes	Type of feedback	Frequency of units
Program Content		
Reading level of elementary or middle school	Positive	2
Explanations on medical terminologies are useful	Positive	7
Video clips and pictures were helpful	Positive	14
Easy to understand, good readability	Positive	28
Helpful and useful educational content	Positive	34
Knowledge/information first learned from this program	Positive	35
Need revisions or irrelevant information	Negative	2
Some video clips and pictures were not working or had low graphic resolution	Negative	3
Need more information on specific topics	Negative	5
Content was not easy to understand, low readability (use of difficult medical terminology)	Negative	7
Program Appearance		
Liked layout (eg, font style, font size, sentence length, location of pictures and paragraphs)	Positive	23
Suggestions	Negative	3
Issues with layout (eg, picture sizes and locations)	Negative	6
Navigation		
Familiar with Facebook navigation	Positive	6
Got lost or failed path (needed help from the researcher)	Negative	2
Others (strengths of program)		
Reasons for not doing or following self-management at home	Positive	3
Provide online social support	Positive	4
Good sources of information: from health care professionals, school, the internet, family, self-experience	Positive	14
Improved intention (attention) to know about their disease	Positive	15
Can share personal experiences on treatment	Positive	17

[a]Representative comments from user testing are shown in Multimedia Appendix 1 (quotations from focus group discussions and user testing).

Discussion

Principal Findings

The aim of this study was to develop a Facebook-based intervention program for childhood cancer participants and to evaluate its usability to guide program refinements. Overall, the evaluation of "Healthy Teens for *Soaam*" revealed that the program was perceived as usable by our participants, who were representative of our target audience (teens with cancer). Seven phases of development and usability evaluation uncovered usability issues as well as areas to enhance user satisfaction, which were then modified accordingly.

Childhood cancer participants and families wanted a comprehensive online information source where they could find childhood cancers–related information. Preexisting sources of information (eg, other patients, hospital handouts, searching the internet) were not reliable and potentially provided limited or outdated information. Therefore, focus group discussion participants greatly valued the objectives and content of this program. Moreover, participants supported the discussion topics at the end of each module. Considering the characteristics of adolescence, peer groups are important for their socialization; however, hospital admissions and principles to prevent infection limit their opportunities to participate in peer group activities. Online interactions among teenagers with similar diagnoses provide the next-best opportunity for social support.

User testing by potential end users revealed that the program content was comprehensive. Previous research has found the two main predictors of noncompliance to chemotherapy to be

XSL•FO
RenderX

child resistance and inadequate information [26]. To overcome these barriers, this program provided tailored information to teens with the goals of improving their knowledge and reducing their resistance to treatment by explaining the rationale behind treatments and procedures, and may therefore improve treatment compliance among teens with cancers.

By adapting the Facebook platform to this program, we took advantage of the following key points: (1) potential end users are familiar with navigation of the program, (2) researchers can reduce costs for developing new online platforms, and (3) researchers may be able to improve treatment fidelity, as teenagers access Facebook almost every day and are continuously connected via the smartphone app. By contrast, the limitations of using Facebook [49] include: (1) design-related limitations (eg, limited freedom and options for text editing and background color schemes) and (2) the inability to control certain features such as advertisements and chronology of posts.

Next Steps

The next step of this "Health Teens for *Soaam*" program will be delivering this 8-week Facebook-based intervention for teens with cancer in South Korea. The purpose will be to increase knowledge and provide accurate, up-to-date information about self-management strategies for improving health and wellness. Through this intervention, we will be able to determine the effectiveness and efficiency of the Facebook-based intervention program.

A previous Facebook-based intervention program for young adult cancer survivors found that engagement with the Facebook program was variable, and investigators recommended that future research should explore how to promote sustained engagement in online social networking [31,50]. Our program includes discussion sections similar to the previous intervention [31,50], in that relevant discussion topics are included in each module. To foster interaction and social support between participants, a moderator will encourage and remind users of discussion sections. A review of the use of social media for teens with childhood cancer [30] reported that health care providers are increasingly integrating social media into their professional life, and that it provides several advantages for both patients and health care providers. For example, the presence of health care providers presents an opportunity for direct interaction with patients and the ability to provide them with reliable, data-based health information. In turn, the health care providers also learn about the experiences and concerns of teens with cancer in real time. Thus, we will consider inviting pediatric oncology health care providers to serve as additional moderators who can provide expert opinions in response to participant queries. This option will improve our study, promote better participant engagement, and enhance treatment fidelity.

Strengths and Limitations

This study had several strengths. We used rigorous qualitative analysis methods, including focus group discussions and user testing transcriptions. Additionally, the use of multiple methods provided a strong source of triangulation and enhanced the reliability of our results, which were informed by both qualitative and quantitative data. We described all development and usability evaluation processes conducted in 7 phases to provide guidance to researchers who want to use Facebook as an intervention platform. According to a recent systematic review of the use of Facebook, only 10 studies have investigated this platform for such interventions [51]. This could be partly because there is no guide on how to utilize this platform, and the task could seem daunting without guidance. Lastly, we used multiple usability evaluation methods. For example, we relied on expert reviews of our content by health care professionals, heuristics evaluation of the Facebook program by human-computer interaction experts, and user testing of the Facebook program by potential end users. Feedback comments from the different evaluation methods guided different aspects of usability issues and improved various facets of our program.

Our study also had several limitations. Regarding participant recruitment, focus group discussions and user testing employed convenience and snowball sampling methods. Sample sizes were relatively small and participants were from one metropolitan area. We did not apply strict inclusion criteria for user testing participants, instead including participants with any type of cancer and at any stage of treatment (ranging from newly diagnosed to completed chemotherapy). Additionally, the current health status of participants was biased toward teens who were in good health. These limitations could decrease the generalizability of the study findings. In addition, our findings may reflect a response bias, as participants who were already interested in Facebook may have been more likely to participate.

Acknowledgments

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MISP and MSIT; 2016R1C1B2013649 and 2020R1C1C1010602).

Authors' Contributions

BP was the lead investigator and oversaw implementation of the project. JK contributed to the content development and implementation of the study. VR contributed to the formulation of this manuscript. All authors reviewed and approved the manuscript for publication.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Quotes from focus group discussions and user testing sessions.
[PDF File (Adobe PDF File), 177 KB - jmir_v22i7e18779_app1.pdf]

Multimedia Appendix 2
Screenshots from the Healthy Teens for Soaam program.
[PDF File (Adobe PDF File), 441 KB - jmir_v22i7e18779_app2.pdf]

References

1. Cancer in Children and Adolescents. National Cancer Institute. 2018. URL: https://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet#r2 [accessed 2020-03-05]
2. National Cancer Institute. SEER Cancer Statistics Review (CSR) 1975-2016. 2020. 2020 Mar 05. URL: https://seer.cancer.gov/csr/1975_2016/ [accessed 2020-03-05]
3. Choi K. Vulnerabilities and Psychosocial Service Needs of Childhood Cancer Survivors and their Caregivers Based on the Cancer Trajectory. Health Soc Welf Rev 2018 Jun;38(2):417-451. [doi: 10.15709/hswr.2018.38.2.417]
4. Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers. Children's Oncology Group. 2018. URL: http://www.survivorshipguidelines.org/pdf/2018/COG_LTFU_Guidelines_v5.pdf [accessed 2020-03-05]
5. Bhakta N, Liu Q, Ness KK, Baassiri M, Eissa H, Yeo F, et al. The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). Lancet 2017 Dec;390(10112):2569-2582. [doi: 10.1016/s0140-6736(17)31610-0]
6. Chemaitilly W, Sklar C. Endocrine complications in long-term survivors of childhood cancers. Endocr Relat Cancer 2010 Sep;17(3):R141-R159. [doi: 10.1677/ERC-10-0002] [Medline: 20453080]
7. Phillips SM, Alfano CM, Perna FM, Glasgow RE. Accelerating translation of physical activity and cancer survivorship research into practice: recommendations for a more integrated and collaborative approach. Cancer Epidemiol Biomarkers Prev 2014 May;23(5):687-699 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-1355] [Medline: 24599577]
8. Printz C. Majority of survivors of childhood cancer experience comorbidities. Cancer 2015 Jul 15;121(14):2293-2293. [doi: 10.1002/cncr.29540] [Medline: 26132388]
9. Nathan P, Ford J, Henderson T, Hudson MM, Emmons KM, Casillas JN, et al. Health behaviors, medical care, and interventions to promote healthy living in the Childhood Cancer Survivor Study cohort. J Clin Oncol 2009 May 10;27(14):2363-2373 [FREE Full text] [doi: 10.1200/JCO.2008.21.1441] [Medline: 19255308]
10. Tyc V, Hadley W, Crockett G. Prediction of health behaviors in pediatric cancer survivors. Med Pediatr Oncol 2001 Jul;37(1):42-46. [doi: 10.1002/mpo.1161] [Medline: 11466722]
11. Warner EL, Nam GE, Zhang Y, McFadden M, Wright J, Spraker-Perlman H, et al. Health behaviors, quality of life, and psychosocial health among survivors of adolescent and young adult cancers. J Cancer Surviv 2016 Apr;10(2):280-290 [FREE Full text] [doi: 10.1007/s11764-015-0474-7] [Medline: 26248766]
12. Coccia PF, Altman J, Bhatia S, Borinstein SC, Flynn J, George S, et al. Adolescent and young adult oncology. Clinical practice guidelines in oncology. J Natl Compr Canc Netw 2012 Sep;10(9):1112-1150. [doi: 10.6004/jnccn.2012.0117] [Medline: 22956810]
13. Klosky JL, Howell CR, Li Z, Foster RH, Mertens AC, Robison LL, et al. Risky health behavior among adolescents in the childhood cancer survivor study cohort. J Pediatr Psychol 2012 Jul 16;37(6):634-646 [FREE Full text] [doi: 10.1093/jpepsy/jss046] [Medline: 22427699]
14. Brinkman TM, Li C, Vannatta K, Marchak JG, Lai J, Prasad PK, et al. Behavioral, Social, and Emotional Symptom Comorbidities and Profiles in Adolescent Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. J Clin Oncol 2016 Oct 01;34(28):3417-3425 [FREE Full text] [doi: 10.1200/JCO.2016.66.4789] [Medline: 27432919]
15. Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. J Clin Epidemiol 2001 Jul;54(7):661-674. [doi: 10.1016/s0895-4356(00)00363-2] [Medline: 11438406]
16. Torres-Espíndola LM, Demetrio-Ríos J, Carmona-Aparicio L, Galván-Díaz C, Pérez-García M, Chávez-Pacheco JL, et al. Comorbidity Index as a Predictor of Mortality in Pediatric Patients With Solid Tumors. Front Pediatr 2019;7:48. [doi: 10.3389/fped.2019.00048] [Medline: 30881949]
17. Landier W, Ahern J, Barakat LP, Bhatia S, Bingen KM, Bondurant PG, et al. Patient/Family Education for Newly Diagnosed Pediatric Oncology Patients. J Pediatr Oncol Nurs 2016;33(6):422-431 [FREE Full text] [doi: 10.1177/1043454216655983] [Medline: 27385664]
18. Beaulieu-Gagnon S, Bélanger V, Meloche C, Curnier D, Sultan S, Laverdière C, et al. Nutrition education and cooking workshops for families of children with cancer: a feasibility study. BMC Nutr 2019;5:52 [FREE Full text] [doi: 10.1186/s40795-019-0319-2] [Medline: 32153965]

XSL•FO
RenderX

19. Nova F, Allenidekania A, Agustini N. The effect of multimedia-based nutrition education on parents' knowledge and body weight change in leukemia children. Enferm Clin 2019 Sep;29(Suppl 2):229-233. [doi: 10.1016/j.enfcli.2019.04.027] [Medline: 31248731]
20. Ångström-Brännström C, Lindh V, Mullaney T, Nilsson K, Wickart-Johansson G, Svärd AM, et al. Parents' Experiences and Responses to an Intervention for Psychological Preparation of Children and Families During the Child's Radiotherapy. J Pediatr Oncol Nurs 2018;35(2):132-148. [doi: 10.1177/1043454217741876] [Medline: 29172925]
21. Anderson LM, Leonard S, Jonassaint J, Lunyera J, Bonner M, Shah N. Mobile health intervention for youth with sickle cell disease: Impact on adherence, disease knowledge, and quality of life. Pediatr Blood Cancer 2018 Aug;65(8):e27081. [doi: 10.1002/pbc.27081] [Medline: 29693797]
22. Howell CR, Krull KR, Partin RE, Kadan-Lottick NS, Robison LL, Hudson MM, et al. Randomized web-based physical activity intervention in adolescent survivors of childhood cancer. Pediatr Blood Cancer 2018 Aug;65(8):e27216 [FREE Full text] [doi: 10.1002/pbc.27216] [Medline: 29722481]
23. Merkle CJ, Chandler DE. Hyperosmolality inhibits exocytosis in sea urchin eggs by formation of a granule-free zone and arrest of pore widening. J Membr Biol 1989 Dec;112(3):223-232. [doi: 10.1007/BF01870953] [Medline: 2515283]
24. Jo M, Hong S, Park HR. Effects of Art Intervention Program for Siblings of Children With Cancer: A Pilot Study. J Pediatr Oncol Nurs 2018 May;35(3):178-187. [doi: 10.1177/1043454218762702] [Medline: 29577798]
25. Na JH, Yoo KY, Kim JY, Park SK, Kim SK, Choi EJ. Educational Interventions to Enhance Adherence to Prophylactic Treatment in Korean Hemophilia Patients. Clin Pediatr Hematol Oncol 2018 Apr 30;25(1):38-42. [doi: 10.15264/cpho.2018.25.1.38]
26. El Malla H, Ylitalo Helm N, Wilderäng U, El Sayed Elborai Y, Steineck G, Kreicbergs U. Adherence to medication: A nation-wide study from the Children's Cancer Hospital, Egypt. World J Psychiatry 2013 Jun 22;3(2):25-33 [FREE Full text] [doi: 10.5498/wjp.v3.i2.25] [Medline: 24175183]
27. Ford JS, Chou JF, Sklar CA. Attendance at a survivorship clinic: impact on knowledge and psychosocial adjustment. J Cancer Surviv 2013 Dec 22;7(4):535-543. [doi: 10.1007/s11764-013-0291-9] [Medline: 23793467]
28. Devine K, Viola A, Capucilli P, Sahler O, Andolina J. Factors Associated With Noncompliance With Long-term Follow-up Care Among Pediatric Cancer Survivors. J Pediatr Hematol Oncol 2017 Apr;39(3):167-173 [FREE Full text] [doi: 10.1097/MPH.0000000000000744] [Medline: 28060114]
29. Watson J. Social Media Use in Cancer Care. Semin Oncol Nurs 2018 May;34(2):126-131. [doi: 10.1016/j.soncn.2018.03.003] [Medline: 29622519]
30. Perales M, Drake E, Pemmaraju N, Wood W. Social Media and the Adolescent and Young Adult (AYA) Patient with Cancer. Curr Hematol Malig Rep 2016 Dec;11(6):449-455 [FREE Full text] [doi: 10.1007/s11899-016-0313-6] [Medline: 26893061]
31. Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surviv 2013 Sep;7(3):355-368 [FREE Full text] [doi: 10.1007/s11764-013-0279-5] [Medline: 23532799]
32. Branch R. Instructional design: The ADDIE approach. Boston, MA: Springer; 2009.
33. Krippendorff K. Content analysis: an introduction to its methodology. Thousand Oaks, CA: SAGE; 2014.
34. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
35. O'Mahony M, Smyth B. Using readability tests to predict helpful product reviews. 2010 Apr 28 Presented at: he 9th international conference on Adaptivity, Personalization and Fusion of Heterogeneous Information; 2010; Paris, France p. 28-30.
36. McClure GM. Readability formulas: Useful or useless? IEEE Trans Profess Commun 1987 Mar;PC-30(1):12-15. [doi: 10.1109/tpc.1987.6449109]
37. Korean Childhood Leukemia Foundation. URL: https://www.kclf.org/en/ [accessed 2020-03-05]
38. Park BK, Nahm E, Rogers VE. Development of a Teen-Friendly Health Education Program on Facebook: Lessons Learned. J Pediatr Health Care 2016;30(3):197-207. [doi: 10.1016/j.pedhc.2015.06.011] [Medline: 26254744]
39. Joyce A, Nielsen J. Teenager's UX: designing for teens. NN/g Nielsen Norman Group. 2013. URL: https://www.nngroup.com/articles/usability-of-websites-for-teenagers/ [accessed 2020-03-05]
40. What's the difference between a public and private Facebook group and how do I change it? 2020. Facebook Help Center. URL: https://www.facebook.com/help/286027304749263?helpref=search&sr=3&query=private&search_session_id=c03abc8ad049f3095413a657b8ef5c89 [accessed 2020-03-05]
41. What are the privacy options for Facebook groups? Facebook Help Center. URL: https://www.facebook.com/help/220336891328465?helpref=about_content [accessed 2020-03-05]
42. Nielsen J. Heuristic engineering. Orlando, FL: AP Professioinal; 1994.
43. Nielsen J. 10 Usability heuristics for user interface design. NN/g Nielsen Norman Group. 2005. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2020-03-05]

XSL•FO
RenderX

44. Alsumait AA, Al-Osaimi A. Usability Heuristics Evaluation for Child E-learning Applications. J Softw 2010 Jun 01;5(6):654-661. [doi: 10.4304/jsw.5.6.654-661]
45. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
46. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
47. Nahm E, Resnick B, Mills ME. Development and pilot-testing of the perceived health Web Site usability questionnaire (PHWSUQ) for older adults. Stud Health Technol Inform 2006;122:38-43. [Medline: 17102214]
48. Park BK. Development and usability evaluation of a Facebook-based program for pediatric cancer patients. 2019 Nov 18 Presented at: American Medical Informatics Association Annual Symposium; 2019; Washington, DC.
49. Park BK, Nahm E, Rogers VE, Choi M, Friedmann E, Wilson M, et al. A Facebook-Based Obesity Prevention Program for Korean American Adolescents: Usability Evaluation. J Pediatr Health Care 2017;31(1):57-66. [doi: 10.1016/j.pedhc.2016.02.002] [Medline: 26952300]
50. Valle C, Tate D. Engagement of young adult cancer survivors within a Facebook-based physical activity intervention. Transl Behav Med 2017 Dec;7(4):667-679 [FREE Full text] [doi: 10.1007/s13142-017-0483-3] [Medline: 28374211]
51. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

Abbreviations

ADDIE: Analysis, design, development, implementation, and evaluation
COG: Children's Oncology Group
eHEALS: eHealth Literacy Scale
eHealth: electronic health
eLearning: electronic learning
K-eHEALS: Korean version of the eHealth Literacy Scale
PHWSUQ: Perceived health website usability questionnaire

Edited by G Eysenbach; submitted 25.03.20; peer-reviewed by A Roundtree, J Karst, J Jonassaint; comments to author 14.04.20; revised version received 22.05.20; accepted 11.06.20; published 28.07.20.

Please cite as:
Park BK, Kim JY, Rogers VE
Development and Usability Evaluation of a Facebook-Based Intervention Program for Childhood Cancer Patients: Mixed Methods Study
J Med Internet Res 2020;22(7):e18779
URL: http://www.jmir.org/2020/7/e18779/
doi:10.2196/18779
PMID:32720897

©Bu Kyung Park, Ji Yoon Kim, Valerie E Rogers. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enabling Better Use of Person-Generated Health Data in Stroke Rehabilitation Systems: Systematic Development of Design Heuristics

Gerardo Luis Dimaguila[1,2], MIT (Health); Kathleen Gray[2], PhD; Mark Merolli[2], PhD

[1]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[2]Centre for Digital Transformation of Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Gerardo Luis Dimaguila, MIT (Health)
School of Computing and Information Systems
University of Melbourne
Parkville
Melbourne, 3010
Australia
Phone: 61 9035 5511
Email: dgl@student.unimelb.edu.au

Abstract

Background: An established and well-known method for usability assessment of various human-computer interaction technologies is called heuristic evaluation (HE). HE has been adopted for evaluations in a wide variety of specialized contexts and with objectives that go beyond usability. A set of heuristics to evaluate how health information technologies (HITs) incorporate features that enable effective patient use of person-generated health data (PGHD) is needed in an era where there is a growing demand and variety of PGHD-enabled technologies in health care and where a number of remote patient-monitoring technologies do not yet enable patient use of PGHD. Such a set of heuristics would improve the likelihood of positive effects from patients' use of PGHD and lower the risk of negative effects.

Objective: This study aims to describe the development of a set of heuristics for the design and evaluation of how well remote patient therapeutic technologies enable patients to use PGHD (PGHD enablement). We used the case of Kinect-based stroke rehabilitation systems (K-SRS) in this study.

Methods: The development of a set of heuristics to enable better use of PGHD was primarily guided by the R3C methodology. Closer inspection of the methodology reveals that neither its development nor its application to a case study were described in detail. Thus, where relevant, each step was grounded through best practice activities in the literature and by using Nielsen's heuristics as a basis for determining the new set of heuristics. As such, this study builds on the R3C methodology, and the implementation of a mixed process is intended to result in a robust and credible set of heuristics.

Results: A total of 8 new heuristics for PGHD enablement in K-SRS were created. A systematic and detailed process was applied in each step of heuristic development, which bridged the gaps described earlier. It is hoped that this would aid future developers of specialized heuristics, who could apply the detailed process of heuristic development for other domains of technology, and additionally for the case of PGHD enablement for other health conditions. The R3C methodology was also augmented through the use of qualitative studies with target users and domain experts, and it is intended to result in a robust and credible set of heuristics, before validation and refinement.

Conclusions: This study is the first to develop a new set of specialized heuristics to evaluate how HITs incorporate features that enable effective patient use of PGHD, with K-SRS as a key case study. In addition, it is the first to describe how the identification of initial HIT features and concepts to enable PGHD could lead to the development of a specialized set of heuristics.

(J Med Internet Res 2020;22(7):e17132) doi:10.2196/17132

KEYWORDS

person-generated health data; patient-reported outcome measures; heuristics; stroke rehabilitation; consumer health informatics; evidence-based practice; information technology

XSL•FO
RenderX

Introduction

Varying Effects of Person-Generated Health Data

Person-generated health data (PGHD) are produced from technologies that allow people to access and utilize health data that they themselves generate outside of a health clinic setting, and to share these data with health care providers and others, typically via the internet. These technologies are designed to generate data about a range of health conditions and in pursuit of a range of health outcomes for remote patient monitoring [1,2]. Example technologies include home-based, web-based, mobile, and wearable apps, which cater to a range of health conditions such as irritable bowel syndrome [3], diabetes [4,5], and bipolar disorder [6]; text reminders for asthma management [7]; and social networking sites for the mental well-being of cancer patients [8].

It is possible for PGHD utilization to have positive, negative, or nil effects on patients who use these technologies. PGHD use has been reported to provide patients with a range of benefits, such as increasing interest in their own health care processes [5,9,10], allowing them to monitor and manage their own health status [11], and motivating them to undergo positive behavior change [1,12]. However, PGHD utilization can also cause patients to feel frustrated and discouraged [1], and some patients may feel excluded from the benefits of PGHD [11]. This highlights the need to consider patients' perspective in the design and development of health technologies [13], particularly those that generate PGHD [14].

Measuring PGHD Effects

The varying effects of PGHD use on patients necessitates the development of a patient-reported outcome measure (PROM) of utilizing PGHD, or PROM-PGHD [15], to provide a standardized way of measuring PGHD outcomes and build empirical evidence about PGHD [16]. PROMs are used to assess the health status or experience of health services and interventions by patients themselves [17-19] and have been shown to improve the precision of evaluating health information technologies (HITs) [20]. Similarly, PROMs-PGHD would measure patients' health outcomes or status as a result of accessing and using PGHD. Moreover, they could complement other clinical health outcome indicators, similar to how PROMs are used alongside other biometric measurements of patient health [21].

The authors previously developed a PROM-PGHD that entailed establishing health outcomes that should be measured for a specific technology category and a health condition [15,22]. The process identified some features that a domain or type of PGHD-enabled technology, that is, stroke rehabilitation systems, should have to evaluate their effectiveness in producing positive effects on patients who use PGHD to self-monitor or manage a specific health condition. For example, simulated stroke rehabilitation technologies should provide patients with real time information on the remaining exercise repetitions they need to do for them to self-manage their limited energy [1].

Evaluation of Technologies Using Heuristics

Using a list of features based on predetermined criteria to assess technology categories is called heuristic evaluation (HE). First introduced by Nielsen and Molich [23], HE is an established and well-known method for the usability assessment of various human-computer interaction technologies [24]. Using a set of guidelines, called heuristics, expert evaluators are able to quickly and efficiently identify issues that could affect the usability of technology artifacts [25]. In practice, it could also be conducted during the design process to identify potential problems before launch [24,25].

HE has since been adopted for evaluations in a wide variety of specialized contexts and with objectives that go beyond usability. This has resulted in specialized heuristics, which could be used to direct evaluators to focus on assessing technologies according to specific concepts of interest [24]. Nielsen's original set of heuristics has been adopted or augmented to assess the quality of novice programing systems [26], identify common mistakes caused by novice and experienced users of a web-based health record system for nurses [27], and assess the safety of a HIT [28]. Such specialized heuristic evaluations have identified problems beyond the original scope of the original heuristics [29] and could therefore increase the effectiveness of evaluations in varied contexts [30] and improve the quality of insights gained [26].

Heuristics for PGHD Enablement

Developing a set of heuristics for use in the formative stages of development of patient and consumer health technologies could ensure more deliberate PGHD-enabling designs. This would improve the likelihood of positive effects from patients' use of PGHD and lower the risk of negative effects. Such heuristics would also offer health care providers a standard way to evaluate technologies as part of selecting and implementing PGHD programs with their patients [11]. HE would complement other in-depth PGHD evaluation methods [2,15] to explore PGHD outcomes [14,16], integration with clinical tools [2], and assessment of quality management [31]. A set of heuristics to evaluate how HITs incorporate features that enable effective patient use of PGHD is needed in an era where there is a growing demand and variety of PGHD-enabled technologies in health care and where a number of remote patient-monitoring technologies do not yet enable patient use of their PGHD. Furthermore, research into the outcomes and benefits of PGHD has not kept up [2,11,14]. Specifically, methods are only emerging for systematic measurement of patient-reported health outcomes from accessing and utilizing PGHD [11,15]. In contrast to the already strong standardization of PROMs in many aspects of health care [21], PGHD outcome evaluation methods are idiosyncratic and reported research is fragmented [11].

Objective

In this study, we describe the development of a set of heuristics for the design and evaluation of how well remote patient therapeutic technologies enable patients to use PGHD (PGHD enablement). We used the case of Kinect-based stroke

XSL•FO
RenderX

rehabilitation systems (K-SRS) to illustrate how the development of PROM-PGHD has led to the creation of these heuristics.

Methods

Case Study

An important use case for evaluating PGHD effects is in simulated rehabilitation technologies for stroke, in particular those using body-tracking technology Kinect (Microsoft) [15]. Key considerations within the context of stroke rehabilitation include the inherent complexity of care necessary to improve health outcomes [32,33], the requirement for stroke survivors to frequently conduct repetitive movement exercises [34,35], and the difficulty of accessing therapy accompanied by the high cost of care over an extended period [36].

K-SRS offer a more convenient and cost-effective option [14]. They allow stroke survivors to undertake prescribed movement exercises simulating activities of daily living (ADL) [37-40]. As survivors use these systems, they produce PGHD in the form of data indicating their therapeutic progress. These data, generally designed to be used by clinicians, have the potential to be used by stroke survivors themselves for health self-monitoring [14,38,39].

Methodology to Develop Specialized Heuristics

A number of heuristic development processes have been employed in the literature. This includes literature reviews, analysis of usability problems, mixed processes, use of guidelines, interviews, and analysis of theories related to a specific domain [41]. Of these, only one methodology known as the R3C [42] has been used extensively (14 times) [41,43,44]. The authors have since improved it [45]. Another widespread approach is to point out the limitations of existing heuristics, in particular [23], and build on these by exploring other aspects that need to be considered for the target domain. The original set of usability heuristics did not detail the steps taken to develop them [23,46]. Subsequently reported processes for developing heuristics for specific application domains have varied.

Procedure to Develop Heuristics for PGHD Enablement

The development of a set of heuristics for PGHD enablement was primarily guided by the R3C methodology most utilized and applied in the literature [45]. However, a closer inspection of the methodology reveals that neither its development [44] nor its application to a case study [47] were described in explicit detail, leaving room for flexibility and interpretation for future implementations. There are also overlaps between its steps and the other heuristic development processes described earlier [41]. For instance, step 1 explores and describes the target technology domain, followed by step 2, where the meaning of usability is reexamined within the context of the target technology domain. Thus, a literature review may be conducted for these steps [47]. Moreover, although those steps do not specify conducting qualitative activities such as interviews and focus groups, their goals may indeed be augmented by such activities.

Thus, the development of a specialized set of heuristics for K-SRS technologies was guided by the R3C [45]. Where relevant, each step was grounded through the best practice activities in the literature, that is, analysis of the context of use through a literature review, interviews, and focus groups, and by using the original heuristics [23] as a basis for determining the new set of heuristics. As such, this study builds on the R3C methodology, and the implementation of a mixed process is intended to result in a robust and credible set of heuristics. Moreover, the detailed implementation of each step is intended to aid future developers of specialized heuristics for PGHD enablement within various contexts of health conditions and technology domains.

The steps of R3C are described in Textbox 1. The objective of this study was to develop an initial set of heuristics for PGHD enablement; thus, only steps 1 to 4 are implemented in this study. Evaluation of the heuristic set (step 5) and further refinement (step 6) is an area for future research.

XSL•FO
RenderX

Textbox 1. Methodology to develop a specialized set of heuristics.

Steps of the R3C methodology to develop a specialized set of heuristics [44,45,47]

- Step 1: Exploratory stage
  - Examine literature related to the main topics or technology domain of the research.
- Step 2: Descriptive stage
  - Highlight key concepts from step 1. Connect the information, and assign weights.
- Step 3: Correlational stage
  - Identify the characteristics that the heuristics should have. Original heuristics [23] may be used as a basis. Clarify the need to consider the target technology domain.
- Step 4: Specification stage
  - Formally specify the set of proposed specialized heuristics, using a standard template. Prioritize heuristics, and report any missing elements that need to be added.
- Step 5: Validation or experimental stage
  - The new set of heuristics is tested on the target domain, and usability problems or issues found are compared with those that would be found using traditional heuristics.
- Step 6: Refinement stage
  - The new set of heuristics is refined or improved following analysis of the results of step 5. However, this occurs iteratively, that is, if no more changes are recommended after step 5 then the development ends there.

Results

Step 1: Exploratory Stage

A literature review was conducted as part of the PROM-PGHD development process to understand how K-SRS have been designed to enable PGHD and what benefits, if any, were recorded [14] using them. Similar to how this step was applied for this case [47], the review also analyzed features of this domain, such as the effectiveness of K-SRS. Moreover, the types of PGHD generated by K-SRS were reviewed as well as how people were able to access those PGHD, the intended users and uses of PGHD, and any effects on patients from using PGHD [14].

The literature review described the context under which PGHD is produced and used, relevant for step 4 when the heuristics are specified. As part of this step, factors that may influence the outcomes resulting from accessing and using PGHD were analyzed. The review has shown that PGHD was given to patients to guide them, for example, change movement behavior or actions to perform correct exercise movements. PGHD were also provided to help them achieve their short- and long-term goals [14]. However, although people were generally provided with some form of PGHD as feedback, they did not have access to all their PGHD, which clinicians have access to. Other uses of PGHD include clinicians analyzing them to tailor rehabilitation programs for patients; researchers analyzing the effects on patients when they use a K-SRS; and to assess a variety of K-SRS, for example, how effective and reliable they are, and even compare them with other domains of technologies [14]. The review also identified one study that reported that a patient's daily access to her PGHD caused her to remember them over time and be motivated to improve [48].

The review also highlighted the lack of patient-centered design in the development of K-SRS, given that the focus of providing data back to patients was for prescribing tasks and not to allow patients to access and make sense of their PGHD themselves. This is a missed opportunity to encourage patients to be more involved in their own health care [14], as patients who have direct access to their PGHD may become more engaged and thus improve their health outcomes [5,9,10,49]. Indeed, the lack of patient participation in the design of such PGHD systems may undermine patients' rehabilitation experience [50]. This is indicative of the need for developers of technology-based rehabilitation systems to consider PGHD in their design, implementation, and evaluation and allow patients to access them [14].

Step 2: Descriptive Stage

This step is described as having 3 tasks: highlighting, connecting, and assigning weights to the concepts found in step 1. Step 2 is applied for this study as follows. First, concepts found through the literature review that could enable the effective patient use of PGHD were identified. Second, the concepts were categorized based on their inherent meaning according to 5 different reported effects of PGHD use thematically derived from the literature. Articles from a significant journal special issue on PGHD were inductively analyzed as an efficient and targeted way to identify reported effects of PGHD utilization from a range of HITs for different health conditions [15]. This process identified 5 different effects that may result from patients' use of their PGHD. PGHD may influence health-related behavioral or attitude changes, influence

XSL•FO
RenderX

patients' management of their own care owing to changes in how they feel about their health status, influence interest in their health care processes, facilitate their personal care goals, and influence their relationship with care providers. Concepts that did not match any of these effects were assigned a different category.

Table 1 shows the categorized concepts highlighted from the literature review. More example quotes are shown in Multimedia Appendix 1. The *Weight* column is discussed in the following paragraphs.

In the third activity of this stage, the concepts were assigned weights through qualitative studies. As part of the PROM-PGHD development process, the input of target users (stroke survivors) and experts (physiotherapists) was elicited through focus groups and interviews [51]. On the basis of the health and technology case under consideration, stroke survivors with experience interacting with a leading K-SRS called Jintronix [14,52] were asked to respond to open-ended questions around the effects resulting from their access to and utilization of PGHD. Meanwhile, the clinicians were asked how they thought that access and utilization of PGHD would affect the stroke survivors they were working with. Participant recruitment was conducted in Australia at 3 different sites, with ethics approval granted by the Human Research Ethics Committees of Deakin University (2017-087), Austin Health (HREC/17/Austin/492), and the University of Melbourne (1852259.1). A total of 10 stroke survivors (7 females and 3 males) were recruited through Deakin University; 5 clinicians (4 females and 1 male) and 6 stroke survivors (1 female and 5 males) were recruited through the Austin Health hospital; and 1 stroke survivor (female) was recruited through Headway ABI in Queensland, Australia.

The qualitative studies enabled the authors to receive direct feedback from PGHD users to assign weights to the concepts found from step 1. In addition, the studies also bridged any relevant gaps between the concepts described in the literature, those perceived by clinicians, and those reported as important by stroke survivors themselves [1]. New PGHD outcomes were documented that were not previously identified through the review and analysis of the literature described earlier [14], leading to the identification of new features that should be incorporated in technologies to enable effective patient use of PGHD.

Concepts identified through the literature review but not described by either stroke survivors or clinicians were assigned a weight of 0, concepts described by clinicians only were assigned a weight of 1, concepts described by stroke survivors only were assigned a weight of 2, and concepts described by both stroke survivors and clinicians were assigned a weight of 3. Table 1 shows the key concepts identified through the literature review assigned with weights and example quotes describing the concept, if any. Table 2 shows new concepts identified through the qualitative studies, also assigned with weights and listing example quotes.

The categories were initially applied by GD and then independently reviewed by the coauthors. Any disagreements in the categories applied were discussed in a meeting until agreement was reached for each concept. More example quotes are shown in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Weighted key concepts identified from step 1: exploratory stage.

Categories	Key concepts	Weight
Stroke survivor health-related behaviors	PGHD[a] can guide stroke survivors to make appropriate movement behavior or action changes to perform an exercise correctly	• 2—stroke survivors: • "I automatically adjust what I was doing, if I couldn't burst every balloon I, I had to adjust… To work out what I was doing wrong" (FG1_STC2-3)
Stroke survivor personal care goals	PGHD can help stroke survivors achieve their short- and long-term goals	0—Not described
Functional effectiveness of therapy	PGHD can help inform clinicians, to analyze the functional effects of simulated rehabilitation therapy, and tailor programs for stroke survivors	0—Not described
Evaluation of the PGHD-enabled K-SRS[b]	PGHD can be used to assess the effectiveness and reliability of a K-SRS, compared with other types of simulated rehabilitation technologies	0—Not described

Stroke survivor interest in care processes	• PGHD may be remembered by stroke survivors over time and can provide them with more motivation to improve their therapy performance
	• PGHD needs to be accessible to the stroke survivors who produce them to allow them to be more involved in their own health care
		• 3—stroke survivors and clinicians:
		  • PGHD may provide "an extra percentage of motivation" (FG1_STC2-1) and an "incentive to do better" (FG1_STC2-3)
		  • Clinician: "something to keep striving" (FG3_AHC1_3)
		• 3—stroke survivors and clinicians:
		  • "it helped me understand my rehabilitation progress" (INT1_HWC1)
		  • "quite often you learn more when you've done something wrong" (FG2_AHP_2)
		  • Clinician: "[survivors] would understand better after a second session where they could compare their results...it's good for them to have a comparison, to see how they've changed from one session to another" (FG3_AHC1_5)

[a]PGHD: person-generated health data.

[b]K-SRS: Kinect-based stroke rehabilitation systems.

XSL•FO
RenderX

Table 2. Weighted new concepts identified from qualitative studies.

Categories	Key concepts	Weight

Stroke survivor health-related behaviors	• PGHD[a] can encourage stroke survivors to do more exercises related to their therapy
	• PGHD can discourage stroke survivors from doing more exercises if it is negative or low
		• 3—stroke survivors and clinicians:
		  • "if I can see the improvement I'm making then it would...encourage me to maybe have more of those sessions" (INT1_STC3)
		  • Clinician: "that might stimulate them to be more compliant" (FG3_AHC1_5)
		• 3—stroke survivors and clinicians:
		  • "probably lose faith in the system" (INT2_AHP)
		  • [unless there is] "some explanation [...] that you need to consider looking" (INT2_AHP)
		  • Clinician: "They might get over it, or might not be willing to participate" (FG3_AHC1_3)

Stroke survivor personal care goals	• PGHD can demotivate stroke survivors if it is negative	• 2—stroke survivors:
		  • "that can be a positive motivator, but can also (be) a negative one" (INT2_STC1)
		  • "I'm thinking well, are normal people at a hundred percent? And I'm only at 60?" (FG1_AHP_2)

Feelings about health status	• PGHD can cause positive or negative emotions, correlated with whether their PGHD is positive or negative
	• PGHD can make stroke survivors feel confused about their health progress
	• PGHD can make stroke survivors feel more self-aware about their health care
		• 3—stroke survivors and clinicians:
		  • "going backwards, that would be a little bit depressing" (INT1_AHP)
		  • seeing "yourself gradually making improvements, it just makes you feel so much better. Okay, I'm achieving something" (INT1_AHP)
		  • Clinician: "a score can motivate you or please you" (FG3_AHC1_5)
		• 3—stroke survivors and clinicians:
		  • "the percentages to me is more difficult to understand [...] are we trying to be a hundred percent at these things?" (FG1_AHP_2)
		  • Clinician: "scope to put in something about, I didn't really understand" (FG3_AHC1_4)
		• 2—stroke survivors:
		  • "to measure your improvement...or measure your, deterioration...You could see it in cold hard figures." (FG2_AHP_2)

Stroke survivor interest in care processes	• PGHD can interest stroke survivors in how their exercises are contributing to their activities of daily living
	• PGHD can help stroke survivors to self-manage their energy while undergoing therapy
	• PGHD can affect stroke survivors' perception about their therapy
		• 1—clinicians:
		  • "any relevance to a functional activity. You know like washing the dishes, or hanging the washing out or...climbing a flight of stairs" (FG3_AHC1_4)
		• 2—stroke survivors:
		  • "It was certainly something that I watched, to see where I was at. 'Cause you need to think about this, we have some sort of a budget of energy that you have to manage yourself, and you can't afford to get to empty" (FG1_STC2-3)
		• 3—stroke survivors and clinicians:
		  • Mismatch between PGHD and feeling of performance: "you probably think oh, the system's not doing its job" (INT2_AHP)
		  • "nothing going on in the background about anyone judging me...that there isn't anything that's being kept from me" (FG1_AHP_2)
		  • Clinician: if you just did the computer thing and they...just stopped without data they'd be like, why do I do it?" (FG3_AHC1_2)

XSL•FO
RenderX

Categories	Key concepts	Weight

Stroke survivor relationship with care provider(s)	• PGHD can prompt stroke survivors to contact their therapists about their therapy performance
	• PGHD can make stroke survivors be more conscious of the exercises prescribed by their clinicians
		• 3—stroke survivors and clinicians:
		  • "If they were always bad then I would need more assistance and even if they [were] good, they [are] not perfect, right so I would want to have more, more assistance to improve" (INT2_STC3)
		  • Clinician: how can I get a better score...why did I...not do very well" (FG3_AHC1_4)
		• 1—clinicians:
		  • "maybe if they're not being compliant, they might get a phone call from the therapist" (FG3_AHC1_5)

Relationship with family and carers	• PGHD can assist stroke survivors in communicating their rehabilitation progress with their loved ones	• 3—stroke survivors and clinicians:
		  • "you could show them something, it's easier for them to visualize" (INT1_AHP)
		  • Clinician: "share it with family...it's that...bragging power as well, perhaps" (FG3_AHC1_4)

[a]PGHD: person-generated health data.

Step 3: Correlational Stage

In this stage, the characteristics that the new set of heuristics should have are defined, which are later specified in step 4.

The key concepts identified through the first 2 steps are used as the basis to define the characteristics. As opposed to concepts, characteristics read as features that K-SRS should have to enable stroke survivors to use PGHD. These characteristics are also matched with the original heuristics [23] to identify any similarities. Some of the concepts were defined with similar characteristics and, therefore, combined. To further increase the credibility of the resulting heuristics, any characteristic matching an original heuristic is reworded using the original heuristic as a guide. Moreover, concepts with a weight of 0, that is, not described by stroke survivors or clinicians, were not defined as a characteristic and correspondingly not specified as a heuristic.

The characteristics were initially defined by GD and then independently reviewed by the coauthors. Any disagreements were discussed in a meeting until agreement was reached for each characteristic.

Table 3 lists the characteristics defined from the concepts identified previously that match an original heuristic [23], and as we have reworded them. Textbox 2 lists the characteristics that did not have a matching original heuristic [23]. The numbering continues from Table 3 through to Textbox 2, to indicate which characteristics were combined in the next step. To see which concepts were defined as which characteristics, please see Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Characteristics defined from the key concepts identified from steps 1 and 2—that match an original heuristic.

Number	Characteristics	Reworded using the original heuristic as a guide
1	• PGHD[a]-enabled systems should assist users in performing more exercises or actions, in a correct way. • Matched with original heuristic 9.	• Help stroke survivors in performing more exercises or actions and to recognize, understand, and recover from errors they make. Guidance or error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
2	• PGHD-enabled systems should provide PGHD to stroke survivors for increased understanding about their rehabilitation or therapy process. • Matched with original heuristic 1.	• PGHD-enabled systems should always provide PGHD to stroke survivors to keep them informed about what is going on with their health status, through appropriate feedback within reasonable time. • This would increase their understanding about their rehabilitation or therapy process.
3	• PGHD-enabled systems should avoid formatting PGHD through a scale that represents a completeness or an endpoint, for example, 100% as much as possible, as it would likely represent failure. Instead, PGHD should resemble the ongoing functional therapeutic progress of stroke survivors. • Matched with original heuristic 2.	• PGHD-enabled systems should present PGHD in a format that matches the real-world context, therapy progress, and goals of the stroke survivors.
4	• PGHD-enabled systems should ensure that PGHD is, or could be, presented in a way that is clearly understandable to a stroke survivor. • Matched with original heuristic 2.	• PGHD-enabled systems should ensure that PGHD is clearly understandable to stroke survivors. The system should speak their language, with words, phrases, and concepts familiar to them, rather than system-oriented terms. PGHD-enabled systems should follow real-world conventions, making information appear in a natural and logical order.
5	• PGHD-enabled systems should provide PGHD to stroke survivors for increased self-awareness about their health care. • Matched with original heuristic 1.	• PGHD-enabled systems should always provide PGHD to stroke survivors to keep them informed about what is going on with their health status, through appropriate feedback within reasonable time. • This would increase their self-awareness about their health care.
6	• PGHD-enabled systems should provide patients real-time PGHD that allow them to self-manage their energy while performing therapy exercises. • Matched with original heuristic 1.	• PGHD-enabled systems should always provide PGHD to stroke survivors to keep them informed about what is going on with their health status, through appropriate feedback within reasonable time. • This would allow them to self-manage their energy while performing therapy exercises.
7	• Notwithstanding the need for PGHD to be as accurate as possible, PGHD-enabled systems should inform the patients of its limitations or potential inaccuracies in the PGHD produced by stroke survivors. • PGHD-enabled systems should also provide PGHD to stroke survivors to foster an increased sense of trust about their rehabilitation or therapy process. • Matched with original heuristic 1.	• PGHD-enabled systems should always provide PGHD to stroke survivors to keep them informed about what is going on with their health status, through appropriate feedback within reasonable time. • This would foster an increased sense of trust about their rehabilitation or therapy process.
8	• PGHD-enabled systems should allow stroke survivors to contact their/a clinician about their PGHD or at least provide survivors with the option of viewing functional, action-based suggestions for them to improve their performance. • Matched with original heuristic 10.	• Even though it is better if the system can be used without additional help or documentation, it may be necessary to provide them. PGHD-enabled systems should provide stroke survivors the option to contact a clinician about their PGHD and vice versa or at least provide survivors with the option of viewing functional, action-based suggestions for them to improve their performance. Any such information should be easy to search, focused on the survivors' exercises, list concrete steps to be carried out, and not be too lengthy.

XSL•FO
RenderX

Number	Characteristics	Reworded using the original heuristic as a guide
9	• PGHD-enabled systems should provide stroke survivors the option of allowing their clinicians to contact them based on the progress of their PGHD. • Matched with original heuristic 10.	• Even though it is better if the system can be used without additional help or documentation, it may be necessary to provide them. PGHD-enabled systems should provide stroke survivors the option to contact a clinician about their PGHD and vice versa or at least provide survivors with the option of viewing functional, action-based suggestions for them to improve their performance. Any such information should be easy to search, focused on the survivors' exercises, list concrete steps to be carried out, and not be too lengthy.

[a]PGHD: person-generated health data.

Textbox 2. Characteristics defined from key concepts identified from steps 1 and 2, which do not have a matching original heuristic.

Characteristics

- 10
  - Person-generated health data (PGHD)–enabled systems should present PGHD that indicate negative or decreasing therapy progress carefully and in a form that elicits a stroke survivors' competitiveness with the self.
- 11
  - PGHD-enabled systems should highlight PGHD indicating positive or improving therapy progress more and providing them with more frequency than negative or decreasing progress.
- 12
  - PGHD-enabled systems should provide PGHD to stroke survivors for increased understanding about how their rehabilitation is contributing to their functional ability.
- 13
  - PGHD-enabled systems should provide stroke survivors with the option to share their PGHD with loved ones in a secure manner.

Step 4: Specification Stage

In this stage, the characteristics defined in step 3 are specified as a heuristic, following a structured format: ID, name, definition, explanation of how the heuristic was developed, example(s) of when a system being evaluated complies with or violates the heuristic, expected benefits if the system complies with the heuristic, and anticipated problems of heuristic misunderstanding [45].

The ID applied to the heuristics was structured as [Number]-PGHD-W[Weight]. The weights were indicated to provide implementers an idea of the process that went through developing the heuristics. The heuristic names were written succinctly, in a similar fashion to the original heuristics [29]. The characteristics were written as the heuristic definition. The explanation of each heuristic described how they were developed from one or more heuristic and provided example descriptions of those concepts by stroke survivors, clinicians, or both. The example and benefits of each heuristic were described based on the literature review [14] and qualitative studies [1]. The anticipated problems indicated where heuristic definitions may be quite close and may be misunderstood.

A number of characteristics were either similar or tightly complemented each other and were therefore combined into one heuristic. Characteristics 3 and 12 described the need for K-SRS to match the health and therapy context of the stroke survivors. Characteristics 5, 2, 6, and 7 described the need for K-SRS to always provide PGHD to stroke survivors. Finally, characteristics 8 and 9 both described the need for K-SRS to provide options for stroke survivors to seek more information, contact their clinicians, and allow their clinician to contact them.

A necessary augmentation is conducted in this step. In combining the characteristics, considerations had to be taken with the individual weights and the categories of those characteristics. Where the weights of the characteristics being combined were different, the highest weight was indicated for the resulting heuristic's ID. Meanwhile, the categories were added as tag(s) under each heuristic and reworded to indicate how the heuristic informs implementers to enable PGHD use. This preserves the categories identified in step 2 as metadata that could aid implementers in understanding the concepts underlying each heuristic. Interestingly, the R3C methodology did not describe how the categories were relevant to the resulting heuristics [41].

The heuristics were initially specified by GD and then independently reviewed by the coauthors. Any disagreements were discussed in a meeting until agreement was reached for each specified heuristic. Moreover, steps 5 and 6 are expected to refine the example, expected benefits, and anticipated problems of each heuristic. Therefore, the heuristics presented here may still be revised later on [45].

XSL•FO
RenderX

The next section shows the new, initial set of heuristics for PGHD enablement.

Design Heuristics to Enable Better Use of PGHD for Testing

Eight new heuristics for PGHD enablement in K-SRS were created. Heuristics with higher weights are presented first, to highlight heuristics that should be prioritized. The first 6 heuristics have a weight of 3, and the last 2 heuristics have a weight of 2. There are 6 tags used: improve health-related behaviors, increase positive feelings about health status, facilitate positive personal care goals, increase interest in care processes, improve relationships with care providers, and improve relationships with family and caregivers.

For brevity, only the ID, name, tag(s), definition, and explanation are presented in Textbox 3. This could also be the short form that implementers may use when evaluating HITs, similar to how brief definitions are presented for the original heuristics [29]. However, as with the original heuristics, implementers ought to understand the underlying concepts and anticipated problems of each heuristic and should, therefore, view the full specifications in Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 3. Short form of the specified heuristics for person-generated health data enablement.

1-PGHD-W3: Encouraging person-generated health data (PGHD)

- Tag(s): Improve health-related behaviors; increase positive feelings about health status
- Definition: The system should highlight PGHD indicating positive or improving therapy progress and provide them with more frequency than negative or decreasing progress.
- Explanation: This heuristic was formed from 3 concepts identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors and clinicians commented on how PGHD can cause positive or negative emotions, correlated with whether the PGHD is positive or negative, for example, when survivors frequently see their PGHD improve, it encourages them to do more of the exercises. A clinician described how seeing PGHD could help survivors be more compliant. On the other hand, if survivors see their PGHD decline, they might lose faith in the system and not perform their exercises.

2-PGHD-W3: Evoking competitiveness with self

- Tag(s): Facilitate positive personal care goals
- Definition: The system should present PGHD that indicates negative or decreasing therapy progress carefully and in a form that elicits a stroke survivors' competitiveness with the self.
- Explanation: This was a concept identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors described how seeing their PGHD can motivate them to try harder, sometimes even when their PGHD indicates that they are not doing very well with their therapy exercises, because it can feel like a competition with themselves. Clinicians agree that it can give survivors something to keep them striving to be better.

3-PGHD-W3: Understandable health data

- Tag(s): Increase positive feelings about health status
- Definition: PGHD-enabled systems should ensure that PGHD is clearly understandable to stroke survivors. The system should speak their language, with words, phrases, and concepts familiar to them, rather than system-oriented terms. They should follow real-world conventions, making information appear in a natural and logical order.
- Explanation: This was a concept identified through a literature review and confirmed by stroke survivors and clinicians. A stroke survivor found that simply having percentages as feedback for performing an exercise could be confusing, as it could be misunderstood as being compared with people who have not had stroke. Clinicians also agreed that PGHD may not be understood by survivors.

4-PGHD-W3: Visibility of health progress

- Tags: Increase positive feelings about health status; increase interest in care processes
- Definition: PGHD-enabled systems should always provide PGHD to stroke survivors to keep them informed about what is going on with their health status, through appropriate feedback within reasonable time.
- Explanation: This heuristic was formed from 4 concepts identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors described how PGHD could help them understand their rehabilitation progress, to determine whether their functional progress was headed in the right direction and learn when they have made a mistake. PGHD could also help them self-manage their energy as they perform their exercises and increase their trust in their therapy process.

5-PGHD-W3: Help and support

- Tag(s): Improve relationship with care providers
- Definition: Even though it is better if the system can be used without additional help or documentation, it may be necessary to provide them. PGHD-enabled systems should provide stroke survivors the option to contact a clinician about their PGHD; allow their clinicians to contact them; or at least provide survivors with the option of viewing functional, action-based suggestions for them to achieve their therapy goals. Any such information should be easy to search, focused on the survivors' exercises, list concrete steps to be carried out, and not be too lengthy.
- Explanation: This heuristic was formed from 2 concepts identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors described how they would want to contact their therapists, particularly if they have not been doing well to ask for assistance and ask for advice on what they could do better. Clinicians also described how it would be good if therapists could contact survivors when they are not being compliant. They also described how stroke survivors would likely ask for more explanation and for advice on how they can do things differently to improve their PGHD.

6-PGHD-W3: Communication of health data

- Tag(s): Improve relationship with family and carers
- Definition: PGHD-enabled systems should provide stroke survivors with the option to share their PGHD with concerned parties, for example, loved ones, in a secure manner.
-

XSL•FO
RenderX

Explanation: This was a concept identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors described how it could help them communicate their health status to people they would like to share it with, particularly when they are tired or when they "run out of words." Clinicians described how survivors might want to share their PGHD with family, particularly with the younger generation who are more technologically inclined.

7-PGHD-W2: Guide for correct exercise movement or actions

- Tag(s): Improve health-related behaviors
- Definition: The system should help stroke survivors in performing more exercises or actions and to recognize, understand, and recover from errors they make. Guidance or error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- Explanation: This was a concept identified through a literature review and confirmed by stroke survivors. A stroke survivor commented on how PGHD can help them work out what they were doing wrong and adjust accordingly.

8-PGHD-W2: Match between system PGHD and real-world context of stroke survivors

- Tag(s): Facilitate positive personal care goals; increase interest in care processes
- Definition: PGHD-enabled systems should present PGHD in a format that matches the real-world context, therapy progress, and goals of the stroke survivors. When presenting PGHD, for example, as a score, percentage, or a graph, they should be matched to the therapy goals of the survivors. These goals may change over the course of a survivors' therapy, from gaining functional progress in the first few months or years to maintaining function to perform ADL when their progress starts to plateau [1]. PGHD needs to be connected with the needs of the stroke survivors and where they are clinically.
- Explanation: This heuristic was formed from 2 concepts identified through a literature review and confirmed by stroke survivors and clinicians. Stroke survivors described how receiving a percent score that did not indicate a 100% could be a negative motivator. Achieving 100% given their condition could be very difficult or even impossible. Clinicians also described how PGHD should help survivors understand how their therapy is contributing to improved performance of ADL.

Discussion

Principal Findings

This study is the first to develop a new set of specialized heuristics to evaluate how HITs incorporate features that enable effective patient use of PGHD, with K-SRS as a key case study. In addition, it is the first to describe how the identification of initial HIT features and concepts to enable PGHD could lead to the development of a specialized set of heuristics. As such, it uncovers a valuable dimension to the PROM-PGHD development method, which considers the sociotechnical context of HITs [15,22]. This context is the complex nature of interactions between people's unique health conditions and behavior and that of the technologies and tools they use within their environment [53]. The prior development of a PROM-PGHD for the key case of K-SRS was guided by this context [15]. Although PROMs have long been known to support the evaluation of health interventions, which may be supported by HITs [20], they are developed to consider only the health condition of patients. Conversely, the nature of PGHD requires the development of a PROM-PGHD to consider the sociotechnical context of the patient experience, that is, both the health condition and technology category [15].

Subsequently, this study highlights the value of considering the sociotechnical context of HITs in their design and evaluation. Although unexpected consequences and even failures can occur from design flaws and technical limitations, they are also often the result of sociotechnical factors [54]. Thus, HIT usability testing and evaluation can benefit from frameworks, standards, and guidelines that consider the sociotechnical context of those technologies [53,55]. In particular, PROMs-PGHD [15], heuristics for PGHD enablement, and cognitive work analysis [56] ensure that understanding of the complexities of care is incorporated in HIT design and implementation [56].

The R3C methodology [45] guided the heuristic development process; however, it was not applied without some difficulty. The methodology's development, description, and application to a case study [44,47] were not described in explicit detail, which left too much room for flexibility. More specifically, it did not specify how the concepts identified from step 1 were going to be connected in step 2; how weights were going to be determined and applied to those concepts, and what influence those weights would have on the succeeding steps; how the characteristics were going to be identified in step 3 and how they differed from the concepts previously identified; and how the original heuristics were going to be "used" as a basis, after matching them with the defined characteristics. Moreover, although R3C presented a structured template for the specification of the heuristics in step 4, the resulting case heuristics only presented summaries or abbreviated versions [47,57,58]. Subsequently, implementations of R3C's steps to develop other specialized heuristics have varied [59-61].

A step-by-step implementation of the R3C methodology was presented. A systematic and detailed process was applied in each step of heuristic development, which bridged the gaps described earlier. It is hoped that this would aid future developers of specialized heuristics, who could apply the detailed process of heuristic development for other domains of technology, and additionally for the case of PGHD enablement for other health conditions. Throughout the process, it was observed that the concepts and characteristics may need to be combined. This study presents a way to retain the weights and categories applied to them as heuristic metadata to aid implementers in understanding their underlying concepts.

XSL•FO
RenderX

Although the last 2 heuristics have a weight of 2, implementers should still aim to conduct the evaluation using all 8 heuristics. However, in case of any resource constraint, the first 6 heuristics should be prioritized.

The R3C methodology [44] was also augmented through the use of qualitative studies with target users (stroke survivors) and domain experts (physiotherapists) to determine the weights, discover new concepts, and ensure the defined characteristics and resulting heuristics were as close to the experiences of the users as possible. This is intended to result in a robust and credible set of heuristics, before the validation stage (step 5), after which the specified heuristics may be improved through the refinement stage (step 6). This is especially valuable, as the R3C methodology's recommended iteration of the process leads from step 6 back to step 4 [45], implying that steps 1 to 3 need to be rigorously implemented. Subsequently, the initial set of heuristics presented here will be validated and refined following steps 5 and 6.

Limitations

Data gathered from the qualitative studies were used as an important augmentation in the heuristic development process. However, as those studies were primarily meant to support the development of a PROM-PGHD, it is possible that participants were not asked all the relevant questions around PGHD enablement features. In the event, many concepts found through the literature review were supported by qualitative studies, and indeed, most of the resulting heuristics were formed from new concepts found in those studies. Therefore, although it might appear to be a limiting factor, it actually indicates the robustness of the process followed and highlights the importance of this suggested augmentation. Moreover, the validation stage (step 5) still allows for possible additional heuristics to be identified.

The flexibility of the R3C methodology [44] meant that it is possible that the process followed in this study deviated from the original intent of the authors. However, we believe that the detailed, step-by-step implementation of R3C and the suggested augmentation have in effect modeled a way to build on and strengthen it.

Conclusions

The new set of heuristics for PGHD enablement, following a detailed, systematic development process augmented from best practice that we have presented, could serve as a guide for future developers of specialized heuristics in general and specifically for developers of heuristics for PGHD enablement of a variety of technology domains and health conditions. The new set of heuristics is needed in a period of rising demand for, supply of, and variety of PGHD-enabled technologies in health care. It offers health care providers a standard way to evaluate technologies as part of selecting and implementing PGHD programs with their patients [11], complementing other in-depth PGHD evaluation methods [2,15], and has a broader relevance for the design and implementation of HITs. In addition, an interesting dimension to the PROM-PGHD development process was discovered, and it highlights the value of considering the sociotechnical context of HITs in their design and evaluation.

Acknowledgments

GD would like to acknowledge the Melbourne School of Engineering through which his research scholarship is provided and his organizational sponsor—Newman College (University of Melbourne).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Steps 1-4, transforming concepts to heuristics.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v22i7e17132_app1.xlsx]

References

1. Dimaguila G, Gray K, Merolli M. Patient-reported outcomes of utilising person-generated health data in simulated rehabilitation technology: perceptions of stroke survivors. Stud Health Technol Inform 2019 Aug 21;264:993-997. [doi: 10.3233/SHTI190373] [Medline: 31438073]
2. Rosenbloom ST. Person-generated health and wellness data for health care. J Am Med Inform Assoc 2016 May;23(3):438-439. [doi: 10.1093/jamia/ocw059] [Medline: 27288293]
3. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. J Am Med Inform Assoc 2016 May;23(3):440-448 [FREE Full text] [doi: 10.1093/jamia/ocv150] [Medline: 26644399]
4. Mamykina L, Levine ME, Davidson PG, Smaldone AM, Elhadad N, Albers DJ. Data-driven health management: reasoning about personally generated data in diabetes with information technologies. J Am Med Inform Assoc 2016 May;23(3):526-531 [FREE Full text] [doi: 10.1093/jamia/ocv187] [Medline: 26984049]
5. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537 [FREE Full text] [doi: 10.1093/jamia/ocv206] [Medline: 27018263]

XSL•FO
RenderX

6. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
7. Johnson KB, Patterson BL, Ho Y, Chen Q, Nian H, Davison CL, et al. The feasibility of text reminders to improve medication adherence in adolescents with asthma. J Am Med Inform Assoc 2016 May;23(3):449-455 [FREE Full text] [doi: 10.1093/jamia/ocv158] [Medline: 26661717]
8. Erfani SS, Blount Y, Abedin B. The influence of health-specific social network site use on the psychological well-being of cancer-affected people. J Am Med Inform Assoc 2016 May;23(3):467-476. [doi: 10.1093/jamia/ocv170] [Medline: 26911816]
9. Sanger PC, Hartzler A, Lordon RJ, Armstrong CA, Lober WB, Evans HL, et al. A patient-centered system in a provider-centered world: challenges of incorporating post-discharge wound data into practice. J Am Med Inform Assoc 2016 May;23(3):514-525 [FREE Full text] [doi: 10.1093/jamia/ocv183] [Medline: 26977103]
10. Woods SS, Evans NC, Frisbee KL. Integrating patient voices into health information for self-care and patient-clinician partnerships: veterans affairs design recommendations for patient-generated data applications. J Am Med Inform Assoc 2016 May;23(3):491-495. [doi: 10.1093/jamia/ocv199] [Medline: 26911810]
11. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: a systematic review. Nurs Outlook 2019;67(4):311-330. [doi: 10.1016/j.outlook.2019.04.005] [Medline: 31277895]
12. Shaw RJ, Steinberg DM, Bonnet J, Modarai F, George A, Cunningham T, et al. Mobile health devices: will patients actually use them? J Am Med Inform Assoc 2016 May;23(3):462-466 [FREE Full text] [doi: 10.1093/jamia/ocv186] [Medline: 26911820]
13. Constand MK, MacDermid JC, dal Bello-Haas V, Law M. Scoping review of patient-centered care approaches in healthcare. BMC Health Serv Res 2014 Jun 19;14:271 [FREE Full text] [doi: 10.1186/1472-6963-14-271] [Medline: 24947822]
14. Dimaguila GL, Gray K, Merolli M. Person-generated health data in simulated rehabilitation using kinect for stroke: literature review. JMIR Rehabil Assist Technol 2018 May 8;5(1):e11 [FREE Full text] [doi: 10.2196/rehab.9123] [Medline: 29739739]
15. Dimaguila GL, Gray K, Merolli M. Measuring the outcomes of using person-generated health data: a case study of developing a PROM item bank. BMJ Health Care Inform 2019 Aug;26(1):e100070 [FREE Full text] [doi: 10.1136/bmjhci-2019-100070] [Medline: 31401587]
16. Cresswell KM, McKinstry B, Wolters M, Shah A, Sheikh A. Five key strategic priorities of integrating patient generated health data into United Kingdom electronic health records. J Innov Health Inform 2019 Jan 4;25(4):254-259 [FREE Full text] [doi: 10.14236/jhi.v25i4.1068] [Medline: 30672407]
17. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claimsF. 2009. URL: https://www.fda.gov/media/77832/download [accessed 2020-06-19]
18. PROMs background document. Canadian Institute for Health Information. 2015. URL: https://www.cihi.ca/sites/default/files/proms_background_may21_en-web_0.pdf [accessed 2020-06-19]
19. Walsh TR, Irwin DE, Meier A, Varni JW, DeWalt DA. The use of focus groups in the development of the PROMIS pediatrics item bank. Qual Life Res 2008 Jun;17(5):725-735 [FREE Full text] [doi: 10.1007/s11136-008-9338-1] [Medline: 18427951]
20. Basch E, Spertus J, Dudley RA, Wu A, Chuahan C, Cohen P, et al. Methods for developing patient-reported outcome-based performance measures (PRO-PMS). Value Health 2015 Jun;18(4):493-504 [FREE Full text] [doi: 10.1016/j.jval.2015.02.018] [Medline: 26091604]
21. Williams K. Patient-reported outcome measures: Literature review. Australian Health Services Research Institute, University of Wollongong.: -; 2016. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/PROMs-Literature-Review-December-2016.pdf [accessed 2020-06-19]
22. Dimaguila GL, Gray K, Merolli M. Patient-reported outcome measures of utilizing person-generated health data in the case of simulated stroke rehabilitation: development method. JMIR Res Protoc 2020 May 7;9(5):e16827 [FREE Full text] [doi: 10.2196/16827] [Medline: 32379052]
23. Nielsen J, Molich R. Heuristic Evaluation of User Interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI'90; April 21-26, 1990; Los Angeles, California, USA. [doi: 10.1145/97243.97281]
24. Kientz J, Choe EK, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. In: Proceedings of the 1st ACM International Health Informatics Symposium. -: -; 2010 Presented at: ACM International Health Informatics Symposium; November 2010; Arlington Virginia USA p. 555-564. [doi: 10.1145/1882992.1883084]
25. Liu L, Hayes G. Heuristic Evaluation of Personal Health Records Systems. 2010 Presented at: CHI 2010; April 2010; Atlanta, Georgia, USA.
26. Kölling M, McKay F. Heuristic evaluation for novice programming systems. ACM Trans Comput Educ 2016 Jun 27;16(3):1-30. [doi: 10.1145/2872521]
27. Scandurra I, Hägglund M, Moström D, Koch S. Heuristic evaluation extended by user analysis: a fast and efficient method to identify potential usability problems in health information systems. J Inform Technol in Healthc 2006;4(5):317-325 [FREE Full text]

XSL•FO
RenderX

28. Borycki E, Kushniruk A, Nohr C, Takeda H, Kuwata S, Carvalho C, et al. Usability methods for ensuring health information technology safety: evidence-based approaches. Contribution of the IMIA working group health informatics for patient safety. Yearb Med Inform 2013;8:20-27. [Medline: 23974544]
29. 10 Usability Heuristics for User Interface Design. Nielsen Norman Group: UX Training, Consulting, & Research. 1994. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2019-07-28]
30. Khowaja K, Salim SS. Correction: heuristics to evaluate interactive systems for children with autism spectrum disorder (ASD). PLoS One 2015;10(8):e0136977 [FREE Full text] [doi: 10.1371/journal.pone.0136977] [Medline: 26301960]
31. Abdolkhani R, Borda A, Gray K. Quality management of patient generated health data in remote patient monitoring using medical wearables - a systematic review. Stud Health Technol Inform 2018;252:1-7. [Medline: 30040674]
32. Mendis S. Stroke disability and rehabilitation of stroke: World Health Organization perspective. Int J Stroke 2013 Jan;8(1):3-4. [doi: 10.1111/j.1747-4949.2012.00969.x] [Medline: 23280261]
33. Cardiovascular diseases (CVDs). World Health Organization. 2016. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2020-06-19]
34. Albiol-Perez S, Palacios- Navarro G, Guerrón-Paredes N, Gil-Gómez JA, Quilis JA, Gil-Gómez H, et al. The Perfetti Method, a Novel Virtual Fine Motor Rehabilitation System for Chronic Acquired Brain Injury. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: PervasiveHealth'14; May 20-23, 2014; Oldenburg, Germany. [doi: 10.4108/icst.pervasivehealth.2014.255251]
35. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity. Ann Neurol 2015 Dec;78(6):848-859. [doi: 10.1002/ana.24472] [Medline: 26150318]
36. Mayo NE, MacKay-Lyons MJ, Scott SC, Moriello C, Brophy J. A randomized trial of two home-based exercise programmes to improve functional walking post-stroke. Clin Rehabil 2013 Jul;27(7):659-671. [doi: 10.1177/0269215513476312] [Medline: 23503738]
37. Lohse KR, Hilderman CG, Cheung KL, Tatla S, van der Loos HF. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
38. Darekar A, McFadyen BJ, Lamontagne A, Fung J. Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review. J Neuroeng Rehabil 2015 May 10;12:46 [FREE Full text] [doi: 10.1186/s12984-015-0035-3] [Medline: 25957577]
39. Moreira MC, de Amorim Lima AM, Ferraz KM, Benedetti Rodrigues MA. Use of virtual reality in gait recovery among post stroke patients--a systematic literature review. Disabil Rehabil Assist Technol 2013 Sep;8(5):357-362. [doi: 10.3109/17483107.2012.749428] [Medline: 23614694]
40. Ogourtsova T, Souza Silva W, Archambault PS, Lamontagne A. Virtual reality treatment and assessments for post-stroke unilateral spatial neglect: a systematic literature review. Neuropsychol Rehabil 2017 Apr;27(3):409-454. [doi: 10.1080/09602011.2015.1113187] [Medline: 26620135]
41. Quiñones D, Rusu C. How to develop usability heuristics: a systematic literature review. Comput Stand Inter 2017 Aug;53:89-122. [doi: 10.1016/j.csi.2017.03.009]
42. Figueroa I, Jiménez C, Allende-Cid H, Leger P. Developing usability heuristics with PROMETHEUS: a case study in virtual learning environments. Comput Stand Inter 2019 Jul;65:132-142. [doi: 10.1016/j.csi.2019.03.003]
43. Jiménez C, Rusu C, Roncagliolo S, Inostroza R, Rusu V. Evaluating a methodology to establish usability heuristics. -: -; 2012 Presented at: Conference of the Chilean Computer Science Society; November 2012; Valparaiso, Chile p. 51-59. [doi: 10.1109/SCCC.2012.14]
44. Rusu C, Roncagliolo S, Rusu V, Collazos C. A methodology to establish usability heuristics. ACHI 2011: The Fourth International Conference on Advances in Computer-Human Interactions. Wilmington, USA: IARIA; 2011. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.684.7653&rep=rep1&type=pdf [accessed 2020-06-19]
45. Quiñones D, Rusu C, Roncagliolo S, Rusu V, Collazos C. Developing usability heuristics for grid computing applications: lessons learned. Prog Adv Comput Intell Eng 2016:485-495. [doi: 10.1007/978-3-319-32467-8_43]
46. Nielsen J, Molich R. Teaching user interface design based on usability engineering. SIGCHI Bull 1989 Aug;21(1):45-48. [doi: 10.1145/67880.67885]
47. Rusu C, Roncagliolo S, Tapia G, Hayvar D, Rusu V, Gorgan D. Usability heuristics for grid computing applications. In: The Fourth International Conference on Advances in Computer-Human Interactions. 2011 Presented at: ACHI 2011: The 4th International Conferences on Advances in Computer-Human Interactions ACHI; 2011; France.
48. Pastor I, Hayes H, Bamberg S. A Feasibility Study of an Upper Limb Rehabilitation System Using Kinect and Computer Games. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: EMBC'12; August 28-September 1, 2012; San Diego, CA, USA. [doi: 10.1109/embc.2012.6346173]
49. Brennan P, Valdez R, Alexander G, Arora S, Bernstam EV, Edmunds M, et al. Patient-centered care, collaboration, communication, and coordination: a report from AMIA's 2013 Policy Meeting. J Am Med Inform Assoc 2015 Apr;22(e1):e2-e6. [doi: 10.1136/amiajnl-2014-003176] [Medline: 25359545]
50. Kristensen HK, Tistad M, Koch LV, Ytterberg C. The importance of patient involvement in stroke rehabilitation. PLoS One 2016;11(6):e0157149 [FREE Full text] [doi: 10.1371/journal.pone.0157149] [Medline: 27285997]

XSL•FO
RenderX

51. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007 May;45(5 Suppl 1):S12-S21 [FREE Full text] [doi: 10.1097/01.mlr.0000254567.79743.e2] [Medline: 17443114]
52. Johnson L, Bird M, Muthalib M, Teo W. Innovative STRoke Interactive Virtual thErapy (STRIVE) online platform for community-dwelling stroke survivors: a randomised controlled trial protocol. BMJ Open 2018 Jan 9;8(1):e018388 [FREE Full text] [doi: 10.1136/bmjopen-2017-018388] [Medline: 29317414]
53. Wozney LM, Baxter P, Fast H, Cleghorn L, Hundert AS, Newton AS. Sociotechnical human factors involved in remote online usability testing of two ehealth interventions. JMIR Hum Factors 2016 Feb 3;3(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.4602] [Medline: 27026291]
54. Percival J, McGregor C. An evaluation of understandability of patient journey models in mental health. JMIR Hum Factors 2016 Jul 28;3(2):e20 [FREE Full text] [doi: 10.2196/humanfactors.5640] [Medline: 27471006]
55. Asan O, Yang Y. Using eye trackers for usability evaluation of health information technology: a systematic literature review. JMIR Hum Factors 2015 Apr 14;2(1):e5 [FREE Full text] [doi: 10.2196/humanfactors.4062] [Medline: 27026079]
56. St-Maurice JD, Burns CM. Modeling patient treatment with medical records: an abstraction hierarchy to understand user competencies and needs. JMIR Hum Factors 2017 Jul 28;4(3):e16 [FREE Full text] [doi: 10.2196/humanfactors.6857] [Medline: 28754650]
57. Solano A, Rusu C, Collazos C, Roncagliolo S, Arciniegas JL, Rusu V. Usability Heuristics for Interactive Digital Television. 2011 Presented at: The Third International Conference on Advances in Future Internet (AFIN 2011); August 2011; Nice, France.
58. Sanz F, Galvez R, Rusu C, Roncagliolo S, Rusu V, Collazos C, et al. A set of usability heuristics and design recommendations for u-learning applications. Proc Int Conf Inf Technol New Gener 2016:983-993. [doi: 10.1007/978-3-319-32467-8_85]
59. Schefer R, Bezerra MS, Zaina LA. Supporting the Development of Social Networking Mobile Apps for Deaf Users: Guidelines Based on User Experience Issues. In: Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion. 2018 Presented at: DSAI'18; June 20-22, 2018; Thessaloniki, Greece. [doi: 10.1145/3218585.3218672]
60. Munoz R, Chalegre V. Defining Virtual Worlds Usability Heuristics. In: Ninth International Conference on Information Technology - New Generations. 2012 Presented at: ITNG'12; April 16-18, 2012; Las Vegas, NV, USA. [doi: 10.1109/itng.2012.138]
61. Díaz G, del Rio CM. A Proposal of Usability Heuristics Oriented to E-Banking Websites. In: International Conference of Design, User Experience, and Usability. 2018 Presented at: DUXU'18; July 15-18, 2018; Las Vegas, NV, USA. [doi: 10.1007/978-3-319-91797-9_23]

Abbreviations

ADL: activities of daily living
HE: heuristic evaluation
HIT: health information technology
K-SRS: Kinect-based stroke rehabilitation system
PGHD: person-generated health data
PROM: patient-reported outcome measure
PROM-PGHD: patient-reported outcome measure of utilizing person-generated health data

Edited by C Gibbons, G Eysenbach; submitted 20.11.19; peer-reviewed by PA Silva, L Novak; comments to author 30.03.20; revised version received 29.04.20; accepted 13.05.20; published 28.07.20.

Please cite as:
Dimaguila GL, Gray K, Merolli M
Enabling Better Use of Person-Generated Health Data in Stroke Rehabilitation Systems: Systematic Development of Design Heuristics
J Med Internet Res 2020;22(7):e17132
URL: http://www.jmir.org/2020/7/e17132/
doi:10.2196/17132
PMID:32720901

©Gerardo Luis Dimaguila, Kathleen Gray, Mark Merolli. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recruiting and Retaining Youth and Young Adults in the Policy and Communication Evaluation (PACE) Vermont Study: Randomized Controlled Trial of Participant Compensation

Andrea C Villanti[1], PhD, MPH; Christie P Vallencourt[2], MPA; Julia C West[1], BA; Catherine Peasley-Miklus[1], PhD; S Elisha LePine[1], BA; Caitlin McCluskey[1], BS; Elias Klemperer[1], PhD; Jeffrey S Priest[3], PhD; Alison Logan[4], MA; Bill Patton[4], MBA; Nancy Erickson[5]; Jennifer Hicks[6], MA; Kathleen Horton[5], BS; Shayla Livingston[7], MPH; Maria Roemhildt[6], PhD; Erin Singer[2], PhD; Megan Trutor[8], BA; Rhonda Williams[2], MES

[1]Department of Psychiatry, Vermont Center on Behavior and Health, University of Vermont Larner College of Medicine, Burlington, VT, United States
[2]Health Promotion & Disease Prevention, Vermont Department of Health, Burlington, VT, United States
[3]Biomedical Statistics Research Core, University of Vermont, Burlington, VT, United States
[4]Hark Inc, Burlington, VT, United States
[5]Communication, Vermont Department of Health, Burlington, VT, United States
[6]Health Surveillance, Vermont Department of Health, Burlington, VT, United States
[7]Public Health Policy, Vermont Department of Health, Burlington, VT, United States
[8]Alcohol & Drug Abuse Programs, Vermont Department of Health, Burlington, VT, United States

Corresponding Author:
Andrea C Villanti, PhD, MPH
Department of Psychiatry
Vermont Center on Behavior and Health
University of Vermont Larner College of Medicine
1 South Prospect Street, MS 482
Burlington, VT, 05401
United States
Phone: 1 8026561187
Email: avillant@uvm.edu

Abstract

Background: The standard approach for evaluating the effects of population-level substance use prevention efforts on youth and young adult perceptions and behaviors has been to compare outcomes across states using national surveillance data. Novel surveillance methods that follow individuals over shorter time intervals and capture awareness of substance use prevention policy and communication efforts may provide a stronger basis for their evaluation than annual cross-sectional studies.

Objective: This study aimed to identify a combination of strategies to recruit a sample of youth and young adults sufficiently representative of the Vermont population and determine how best to retain a web-based panel of youth and young adults over a 6-month period.

Methods: Eligible participants were Vermont residents aged 12 to 25 years who were willing to complete three 10 to 15-minute web-based surveys over a 6-month period. Recruitment was conducted via the following three main mechanisms: (1) web-based recruitment (paid and unpaid), (2) community-based recruitment through partners, and (3) participant referrals via a personalized link. Upon completion of the baseline survey, participants were randomly assigned to one of the following three retention incentive conditions: (1) guaranteed incentive (US $10), (2) lottery incentive (US $50 weekly lottery drawing), and (3) preferred method (guaranteed or lottery). Analyses examined cost per survey start by recruitment source, distribution of demographic characteristics across incentive conditions, and retention by study condition at 3-month and 6-month follow-ups.

Results: Over a 10-week period in 2019, we recruited 480 eligible youth (aged 12-17 years) and 1037 eligible young adults (aged 18-25 years) to the Policy and Communication Evaluation (PACE) Vermont Study. Facebook and Instagram advertising produced the greatest number of survey starts (n=2013), followed by posts to a state-wide web-based neighborhood forum (n=822) and Google advertisements (n=749). Retention was 78.11% (1185/1517) at 3 months and 72.18% (1095/1517) at 6 months. Retention was equivalent across all incentive study conditions at both waves, despite a strong stated preference among study

XSL•FO
RenderX

participants for the guaranteed payment at baseline. Youth had greater retention than young adults at both waves (wave 2: 395/480, 82.3% vs 790/1037, 76.18%; wave 3: 366/480, 76.3% vs 729/1037, 70.30%). Substance use prevalence in this cohort was similar to national and state-level surveillance estimates for young adults, but was lower than state-level surveillance estimates for youth. Most participants retained at wave 3 provided positive qualitative feedback on their experience.

Conclusions: Our study supports the feasibility of recruiting a web-based cohort of youth and young adults with representation across an entire state to evaluate substance use prevention efforts. Findings suggest that a guaranteed payment immediately upon survey completion coupled with a bonus for completing all survey waves and weekly survey reminders may facilitate retention in a cohort of youth and young adults.

(J Med Internet Res 2020;22(7):e18446) doi:10.2196/18446

KEYWORDS

recruitment; retention; adolescents; young adults; prevention

Introduction

Adolescence and young adulthood are defined by developmental processes that mark increased susceptibility to risk-taking behaviors, including substance use [1-4]. In tobacco control, prevention efforts have shifted from individual and group-level interventions to population-based approaches, including policy and mass media efforts to reduce the appeal and accessibility of tobacco products to young people [5]. Concurrently, state-level cannabis policies in the United States have aimed to liberalize the accessibility of cannabis products, though there have been few state-level prevention campaigns. Using national surveillance data across states has been the standard approach to evaluate the effects of these policies on youth and young adult perceptions and behaviors [6,7]. These evaluations, which use cross-sectional data over time, may mask more nuanced trends in individual-level changes in harm perceptions and behavior and have largely failed to address spillover effects on other substance use. Novel surveillance methods that follow individuals over time and capture awareness of substance use prevention policy and communication efforts may provide a stronger basis for their evaluation.

Vermont represents a unique test case for the evaluation of population-level interventions for substance use for three reasons. First, the prevalence of substance use in young people from Vermont is higher than national estimates [8-10]. Second, Vermont has implemented a number of new policies related to substance use in the past 5 years, including a state-wide opioid drug disposal program, stronger prescribing guidelines, and requirements in the Vermont Prescription Monitoring System (2016); legalized possession of cannabis for adults aged 21 years or above [11] (2018); a ban on the web-based sale of e-cigarettes [12] and a 92% tax on e-cigarettes [13] (2019); and an increase in the legal age of tobacco sale to 21 years [14] (2019). In addition to policy efforts, the Governor's 2018 Opioid Coordination Council Report recommended development and implementation of school-based primary prevention programs for opioid use and a comprehensive drug prevention messaging campaign [15]. Third, while Vermont has a relatively homogeneous population in terms of race/ethnicity, it is the second most rural state in the country [16], with between 60% and 100% of the area classified as rural depending on the definition [17]. Thus, policy and communication interventions may be implemented or experienced differently than in settings with greater population density.

In 2018, researchers and program staff at the University of Vermont and the Vermont Department of Health began discussing the development of a longitudinal cohort study of youth and young adults to evaluate responses to changes in tobacco, alcohol, and other substance use policies, communication, and interventions at the state level. The Policy and Communication Evaluation (PACE) Vermont Study was designed to complement existing evaluation efforts that rely on a combination of state-level surveillance and smaller convenience samples [18] The PACE Vermont Study uses web-based data collection in a large sample of youth and young adults, with surveys at shorter intervals to capture changes over time. The survey instrument was also designed to be flexible, allowing for assessment of emerging issues and communication outcomes (eg, knowledge, attitudes, and beliefs) not typically captured in state surveillance systems.

Web-based data collection was proposed to reduce barriers to participation in research among rural people from Vermont, given that as of 2018, 98% of young adults aged 18 to 29 years use the internet, as do 78% of adults who live in rural communities [19]. Similarly, 94% of young adults aged 18 to 29 years and 65% of adults in rural communities own a smartphone [20]. Adoption of web-based surveys for data collection is likely to appeal to young people, who grew up with computers and use them in virtually all aspects of their lives, while reducing transportation and other costs that serve as barriers to engaging in traditional clinical trials [21-24]. With respect to retaining young people in longitudinal studies, there was no clear recommendation on incentives to maximize retention, as previous studies had identified multiple means, including increasing participant payments, conducting sweepstakes, providing bonuses, and sending reminder postcards [25-28].

This study had the following two primary goals: (1) to identify the combination of recruitment strategies that would provide a sample of youth and young adults sufficiently representative of the Vermont population and (2) to determine how best to retain a web-based panel of youth and young adults to be able to attribute changes in knowledge, attitudes, beliefs, and behaviors to specific interventions. Specifically, this study experimentally compared the effects of a lottery payment, a guaranteed

XSL•FO
RenderX

payment, and participant preference for a particular completion incentive on retention at 3-month and 6-month follow-ups. Our a priori hypothesis was that retention would be higher in the participant preference incentive condition than in the lottery or guaranteed payment incentive condition, as providing choice for some study-related decisions has been described as a means to improve retention in studies involving young adults [29].

Methods

Study Overview

The study consisted of three web-based surveys conducted from March 2019 through October 2019 and was approved by the Institutional Review Boards of the University of Vermont and Vermont Department of Health. This research also received a Certificate of Confidentiality from the National Institutes of Health.

Recruitment and Enrollment

Eligible participants were Vermont residents aged 12 to 25 years who were willing to complete three 10- to 15-minute web-based surveys over a 6-month period. Youth participants aged 12 to 17 years also had to report being a US citizen or permanent resident. Recruitment was conducted by Hark, a Vermont-based digital design and marketing firm [30], over a 10-week period (March 26-June 4, 2019). Participants were recruited via the following three main mechanisms: (1) web-based recruitment including both paid and unpaid advertising, (2) community recruitment through partner organizations, and (3) participant referrals via a personalized link. Each recruitment type contained a unique link to the study website to be automatically tracked via Google Analytics. Web-based recruitment occurred through paid Facebook, Instagram, and Google display and Gmail advertisements; free Craigslist Vermont advertisements; posts on PACE Vermont's social media accounts (Facebook, Instagram, and Twitter); paid posts on a state-wide online neighborhood forum (Front Porch Forum) [31]; and paid placement of advertisements in local web-based and print newspapers with relevant news stories. Community recruitment occurred through outreach from engaged partners to their constituents and via news media on the PACE Vermont Study. Partner organizations received a recruitment toolkit, with a tailored newsletter blurb, flyers, sample email language, and sample social media posts for print and digital promotion of the study, as well as a timeline for distribution of recruitment material to their networks. Participant referrals were requested at the end of baseline survey completion via a thank you email with a personalized referral link, a request to share the survey link with friends on social networks, and a direct link to the PACE Vermont Facebook page. Parents of youth participants were able to opt in to be considered for a small study incentive (US $5) for referring another eligible family to the study using their unique link. The PACE research team met weekly to monitor representation by age, race, ethnicity, and county, and promotional strategies during this period were adjusted to focus on underrepresented areas of the state or sociodemographic groups.

Digital advertising channels, including Google display advertisements, Google Gmail advertisements, and Facebook display advertisements, delivered content to the following three Vermont-based segments: youth (aged 12-17 years), young adults (aged 18-25 years), and parents (aged ≥18 years). In addition to age and geography, Google and Facebook's interest and lifestyle-based targeting enabled specific targeting of parents. Prior to recruitment, the PACE team integrated Google and Facebook advertisements with the study landing page and its web analytics tool. Google and Facebook campaigns for the PACE Vermont Study were set to optimize their targeting algorithms and advertisement variations to achieve a maximum number of survey starts rather than clicks or page views. The initial structure of both the Facebook and Google advertisement campaigns mirrored the audience segments (youth, young adults, and parents). Advertisements for each segment featured visual assets (ie, photographs, illustrations, and graphics) and text in the form of headlines, posts, email body, and call-to-action buttons. Advertisements directed at youth and young adults used the following message themes: (1) earn cash rewards and buy the things you really want and (2) be a leader in your community and share your opinion on important topics. Advertisement visuals took the following two approaches: (1) photographs of youth and young adults smiling and holding a mobile phone or gift card and (2) eye-catching illustrations and graphics with action-oriented phrases like "We want to hear from you." Advertisements directed at parents were designed to motivate parents to encourage their children to join the study. Parent-focused messages used the following themes: (1) your family can help improve the health of Vermont's youth for years to come and (2) your teen's participation will help Vermont create substance-abuse resources for other families. Advertisement visuals took the following two approaches: (1) humorous photography of toddlers and young children making messes and (2) sentimental photography of parents embracing their babies and young children. A separate wave of Facebook and Google advertisements ran close to the end of the recruitment window and highlighted the urgency of participation with messages like "Don't miss out on PACE VT" and "Time is running out." Other promotional channels like Front Porch Forum and Craigslist primarily targeted parents and older young adults. Because these channels do not allow for advanced targeting beyond geography, promotional messages mirrored those of parents in our paid campaigns.

All study advertisements and links directed participants to the PACE Vermont website, where there was a brief study description and link to an open web-based screener. In addition to direct advertising to youth, parents of eligible youth were also targeted via promotional efforts. They were asked to review an information sheet and provide informed consent prior to youth providing assent to complete the screener. Youth who initiated the screening survey without parental consent were asked to provide parent contact information, which triggered an email to the parents. Upon parental consent, youth received a unique link to the screener to provide assent. Youth and young adults underwent an electronic informed consent process and received an email link to contact study personnel to ask questions about the study. Eligible consenting youth and young adults were automatically forwarded to the baseline survey, where they completed demographic information and questions on substance use knowledge, attitudes, beliefs, and behaviors.

XSL•FO
RenderX

These included measures of ever and past 30-day use of cigarettes, alcohol, and marijuana, which served as benchmarks for comparison to prevalence estimates from state and national surveillance on substance use data. Participants were then automatically forwarded to a web-based version of the University of Vermont participant payment form required for internal tracking of study payments. All surveys were voluntary, deployed via Qualtrics [31], and optimized for completion via a computer or mobile phone.

During baseline data collection, we revised our survey delivery to ensure the eligibility of participants and the validity of study responses in several ways as follows: (1) adding automatic screening within our survey platform, Qualtrics [32], to exclude participants with an IP address outside Vermont; (2) conducting consistency checks between age and date of birth, as well as state of residence and location of IP address; (3) adding a CAPTCHA item in the screener to ensure that respondents were human and not bots; (4) conducting additional screening of respondents with suspicious email addresses (eg, common e-mail format across surveys completed within minutes of each other and email addresses including names that did not correspond to contact information) and out-of-state phone numbers; and (5) using information from the screening and payment forms (eg, consistency of name across forms and location of participant address) to verify eligibility. Potentially fraudulent participants were flagged and received an email from the study team offering an opportunity to confirm their contact information and remain in the study. Respondents who did not confirm that they were valid participants were removed from the study.

Intervention and Retention

At the end of the baseline survey, participants were asked "Which of the following would you like to receive for completing other web-based surveys like this?" with the following two response choices: "Receive a $10 online gift card" or "Be entered into a lottery to receive $50." After responding to this question, participants were randomly assigned within the Qualtrics survey system to one of the following three study conditions: (1) guaranteed incentive (US $10), (2) lottery incentive (US $50 weekly lottery drawing), and (3) preferred method (guaranteed or lottery, based on the response to the question). Participants were informed of their study condition, and those in the guaranteed condition were automatically directed to a web interface [33] to confirm their email address and receive their US $10 electronic gift card (wave 1 study payment). Gift cards were emailed to wave 1 participants once they were confirmed as participants with valid completion. Lottery drawings were conducted among those who had completed the survey in a given week (n=3 youth and n=3 young adult winners selected each week during each survey wave), and winners were notified by email. Participants were told that they would receive a bonus payment for completing all three surveys (lottery payment group: US $50 bonus; guaranteed payment group: US $20 bonus in addition to US $10 per completed survey). Thus, all participants received US $50 upon completing all three surveys and lottery participant winners could receive more.

Follow-up surveys were launched approximately 3 months (June 27-July 31, 2019) and 6 months (September 17-October 15, 2019) after the baseline survey. Each survey was distributed initially via email or text message, based on participant preference, with a message that notified participants about the dates of data collection, including the 1-month window during which the follow-up survey would be open for completion. Weekly reminder messages were sent to youth and young adults who had not completed the survey via both email and text message throughout the 1-month window, with additional reminders in the last 2 days of each window via email, text message, and social media. Incentive payments at each follow-up were dictated by the study condition, with those in the guaranteed condition automatically linked to Rybbon to confirm their email address and receive their US $10 gift card immediately. As in the baseline survey, lottery drawings (US $50) were conducted each week during data collection for the two follow-up surveys and winners were notified by email. At the completion of the final follow-up survey, participants who had completed all three waves received their bonus payment immediately via Rybbon.

Participant Feedback

At the end of the final survey, participants were asked the following two questions about their experience in the PACE Vermont Study: "What was your favorite part of participating in the PACE Vermont Research Study?" and "What could we improve in the PACE Vermont Research Study to make it easier for you to participate?" Responses were open-ended. Two coders (SEL and CM) reviewed the responses and created in vivo inductive categories from themes that arose during the course of data analysis, as described by Miles and Huberman [34]. Responses for each question were then coded in NVivo software (QSR International) independently by each coder, and responses were allowed to fall into more than one category. Reliability of the coders for each category ranged from a kappa value of 0.47 to 0.97, representing moderate to almost perfect agreement. For the question "What was your favorite part of participating in the PACE Vermont Research Study?" the coding category with the lowest level of agreement was "other" (κ=0.47) and the category with the highest level of agreement was "compensation" (κ=0.97). For the question "What could we improve in the PACE Vermont Research Study to make it easier for you to participate?" the coding category with the lowest level of agreement was "learning" (κ=0.54) and the category with the highest level of agreement was "compensation" (κ=0.88).

Statistical Analysis

This study used the following two sources of data: advertising metrics, and enrollment and follow-up data from our surveys. First, we estimated the cost per survey start according to the recruitment source by dividing the total amount spent on each source by the number of survey starts. Second, we developed a CONSORT diagram to track participants from enrollment through the three survey waves and estimated differences in retention by the intervention condition at each stage using chi-square tests. We conducted additional analyses to test whether retention differed by concordance between preference

XSL•FO
RenderX

and the intervention condition (ie, lottery vs guaranteed compensation). We also examined the distribution of sociodemographic characteristics (ie, county, sex, race/ethnicity, and employment status) and substance use (ever and past 30-day cigarette, alcohol, and marijuana use) by age and the intervention condition using chi-square tests and *t* tests and the differences in these characteristics among those retained at all three waves versus those lost to follow-up. Survey weights were developed post-hoc from population estimates of females and males between the ages of 12 and 25 years (year by year) residing in each of Vermont's 14 counties in 2017 [35]. The goal of survey weighting was to determine how closely the convenience sample matched other state surveillance, as well as correct for the higher response by females and those residing in Chittenden County. Each cell of the table was divided by the total number of individuals between the ages of 12 and 25 years residing in the 14 counties (n=116,407) to generate population-based proportions. A comparable table of survey respondents was created, totaling the number of individuals who completed the baseline survey by sex, age, and county of residence. Each subtotal was divided by the total number of respondents (n=1517) to generate sample-based proportions. Survey weights were then calculated by dividing population proportions by sample proportions, again by sex, age, and county. We compared the weighted prevalence of ever and past 30-day use of cigarettes, alcohol, and marijuana in our sample to national and state-level surveillance estimates from the National Survey on Drug Use and Health (NSDUH) [10,36]. Finally, we assessed the major categories of responses to two participant feedback items from qualitative coding.

Results

Recruitment and Enrollment

Table 1 presents information on new visitors to the study website; survey starts for parents, youth, and young adults; and cost per survey start by recruitment source. Survey starts by recruitment source were tracked using Google Analytics on the PACE Vermont study website. Survey completions by recruitment source could not be captured as surveys were conducted within Qualtrics. Overall, there were 9975 new visitors to the study website, with Facebook and Instagram advertisements accounting for 54.55% (5441/9975) of web traffic, with 2013 survey starts. Google display and Gmail advertisements accounted for 35.10% (3501/9975) of web traffic and 749 survey starts. Three postings in an email digest with state-wide coverage (Front Porch Forum) generated 9.75% (973/9975) of web traffic and 939 survey starts. Partner referrals, newspaper print advertisements, and Craigslist advertisements produced smaller numbers of new users and survey starts. The cost per conversion to a survey start was US $382 considering all recruitment sources; when newspaper print advertisements were excluded, the cost per conversion to a survey start was US $11. Only four parents received the US $5 referral incentive.

The best performing Google advertisements were Gmail advertisements run close to the end of the recruitment window highlighting the urgency of participation for parents, youth, and young adults; the top Facebook posts for youth and young adults highlighted urgency and the importance of youth and young adult feedback ("We want to hear from you.") The top performing Facebook post for parents cited how findings from the PACE Vermont Study would be used to guide resources for substance use prevention in Vermont. While advertisements targeting youth ran for the duration of the recruitment period, advertisements targeting parents were more effective at recruiting youth than youth-specific advertisements.

In addition to the 2723 youth and young adult survey starts from the study advertisements, participants referred others to the study, resulting in 2861 completed screenings (Figure 1). Of those assessed for eligibility, 1008 were deemed ineligible (eg, not aged 12-25 years and not Vermont residents) and 336 were excluded for other reasons, including being flagged as potentially fraudulent.

Overall, there were 480 youth respondents and 1037 young adult respondents included in the PACE Vermont Study sample (Table 2). The recruitment cost per eligible enrolled participant was US $29 (US $44,111/1517 participants). Participants represented each of the 14 counties in the state, with the distribution by county generally reflecting 2017 population estimates for Vermont youth and young adults [35]. The majority of the sample was female (1071/1517, 70.60%), was white (1318/1517, 86.88%), and reported working either part-time or full-time (898/1517, 59.20%).

Table 3 presents the number of responses and weighted prevalences of cigarette, alcohol, and marijuana use by age group in the PACE Vermont Study compared with estimates for the same measures in the NSDUH. Ever use estimates for each substance were provided in the 2018 NSDUH national report [36] and past 30-day measures were drawn from the 2016-2017 state-level report [10]. Prevalence of ever cigarette, alcohol, and marijuana use in PACE Vermont Study youth participants was generally similar to national estimates from the NSDUH, although past 30-day use estimates tended to be lower in PACE Vermont youth compared with state-level estimates from the NSDUH. Ever and past 30-day use in young adult PACE Vermont participants tracked closely with NSDUH estimates, with the following three exceptions: past 30-day cigarette use was lower in the PACE Vermont sample compared with the NSDUH estimate (18.78% vs 33.38%), and the PACE Vermont sample showed higher ever alcohol use (PACE vs NSDUH: 89.52% vs 79.70%) and ever marijuana use (70.57% vs 51.50%).

XSL•FO
RenderX

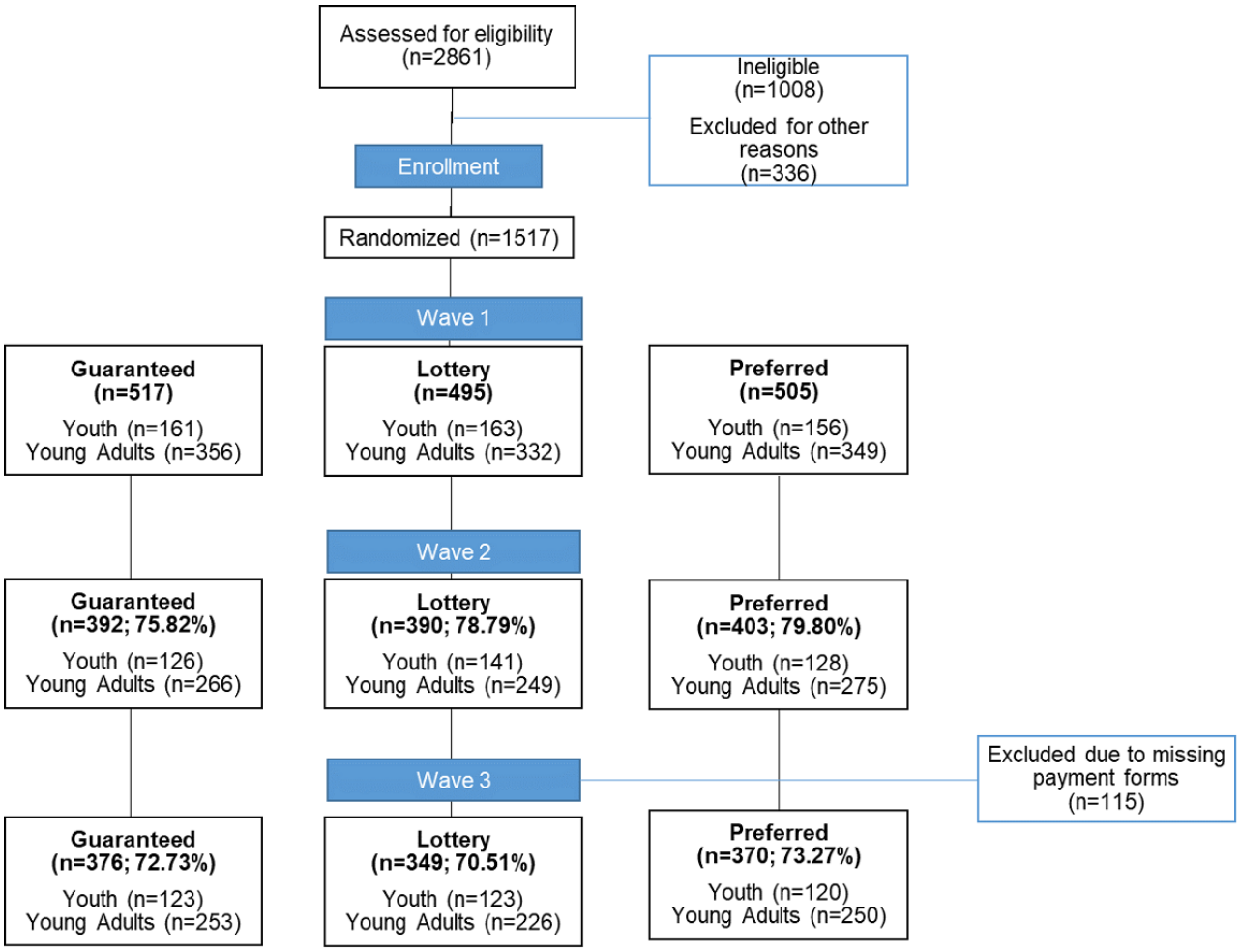
Table 1. Recruitment sources, costs, and conversions in the PACE Vermont Study 2019.

Source	Cost (total: US $44,110.71)	New users (N=10,250)	Conversions, n				Cost per conversion[a]
			Parent survey starts (N=1026)	Young adult survey starts (N=1772)	Youth survey starts (N=951)	Total survey starts (N=3749)	
Facebook & Instagram advertising	US $23,676.42	5441	443	1247	323	2013	US $11.76
Google display & Gmail advertising	US $9213.78	3501	331	233	185	749	US $12.30
Front Porch Forum (paid post, partial state coverage)	US $4950.00	223	37	26	54	117	US $42.31
Front Porch Forum (two sponsored posts, state wide[b])	US $3666.00	1025	199	248	375	822	US $4.46
Newspaper print advertisements	US $2604.51	1	1	0	0	1	US $2604.51
Craigslist	US $0.00	8	1	4	0	5	US $0.00
Partner sources	US $0.00	51	14	14	14	42	US $0.00

[a]Average cost per conversion was US $381.55, and average cost per conversion, excluding print advertisements, was US $11.06.

[b]Sponsored posts available to the Vermont Department of Health for outreach activities.

Figure 1. CONSORT flow diagram of participants in the Policy and Communication Evaluation (PACE) Vermont Study, 2019.



XSL•FO
RenderX

Table 2. Participant characteristics by incentive condition in the PACE Vermont Study 2019.

Characteristic[a]	Incentive condition			Total (N=1517), n (%)	*P* value
	Guaranteed (N=517), n (%)	Lottery (N=495), n (%)	Preference (N=505), n (%)		
Age group (years)					.75
12-17	161 (31.14)	163 (32.93)	156 (30.89)	480 (31.64)	
18-25	356 (68.86)	332 (67.07)	349 (69.11)	1037 (68.36)	
County of residence					.42
Addison	41 (7.93)	33 (6.67)	40 (7.92)	114 (7.51)	
Bennington	20 (3.87)	18 (3.64)	16 (3.17)	54 (3.56)	
Caledonia	16 (3.09)	23 (4.65)	21 (4.16)	60 (3.96)	
Chittenden	228 (44.10)	225 (45.45)	203 (40.20)	656 (43.24)	
Essex	__[b]	—	8 (1.58)	14 (0.92)	
Franklin	33 (6.38)	25 (5.05)	25 (5.95)	83 (5.47)	
Grand Isle	6 (1.16)	5 (1.01)	—	15 (0.99)	
Lamoille	10 (1.93)	19 (3.84)	25 (4.95)	54 (3.56)	
Not sure	9 (1.74)	10 (2.02)	12 (2.38)	31 (2.04)	
Orange	14 (2.71)	20 (4.04)	12 (2.38)	46 (3.03)	
Orleans	10 (1.93)	8 (1.62)	11 (2.18)	29 (1.91)	
Rutland	22 (4.25)	24 (4.85)	37 (7.33)	83 (5.47)	
Washington	68 (13.15)	51 (10.30)	62 (12.28)	181 (11.93)	
Windham	18 (3.48)	18 (3.64)	18 (3.56)	54 (3.56)	
Windsor	19 (3.68)	12 (2.42)	11 (2.18)	42 (2.77)	
Sex					.16
Female	356 (68.86)	361 (72.93)	354 (70.10)	1071 (70.60)	
Male	161 (31.14)	132 (26.67)	151 (29.90)	444 (29.27)	
Race/ethnicity					.21
White	442 (85.49)	444 (89.70)	432 (85.54)	1318 (86.88)	
Nonwhite/other/multiple race	45 (8.70)	34 (6.87)	45 (8.91)	124 (8.17)	
Hispanic	30 (5.80)	16 (3.23)	28 (5.54)	74 (4.88)	
Employment status					.33
Do not currently work for pay	222 (42.94)	207 (41.82)	189 (37.43)	618 (40.74)	
Work part-time (<15 hours/week)	101 (19.54)	80 (16.16)	106 (20.99)	287 (18.92)	
Work part-time (15-34 hours/week)	74 (14.31)	83 (16.77)	85 (16.83)	242 (15.95)	
Work full-time (35 hours/week or more)	120 (23.21)	124 (25.05)	125 (24.75)	369 (24.32)	

[a]There were missing data for county (n=1), sex (n=2), race/ethnicity (n=1), and employment status (n=1).

[b]Suppressed due to unweighted numerator <5 or unweighted denominator <50.

XSL•FO
RenderX

Table 3. Comparison of substance use prevalence by age group in the PACE Vermont sample and the National Survey on Drug Use and Health.

Characteristic	Youth (aged 12-17 years)			Young adults (aged 18-25 years)		
	Value, n	Weighted % (95% CI)	NSDUH[a] estimate	Value, n	Weighted % (95% CI)	NSDUH estimate
Cigarette use						
Ever	42	9.1% (6.3-12.9)	9.6% [36][b]	487	47.4% (43.4-51.4)	45.9% [36][b]
Past 30 days	11	2.2% (1.1-4.4)	5.8% [10] (4.5-7.3)	178	18.8% (15.9-22.1)	33.4% [10] (28.9-38.2)
Alcohol use						
Ever	141	29.4% (24.5-34.7)	26.3% [36][b]	935	89.5% (86.9-91.7)	79.7% [36][b]
Past 30 days	43	9.3% (6.6-12.9)	13.6% [10] (11.3-16.3)	743	70.8% (66.9-74.3)	70.9% [10] (66.7-74.8)
Binge alcohol use, past 30 days	13	3.1% (1.6-5.8)	7.2% [10] (5.7-9.0)	484	48.3% (44.3-52.3)	49.3% [10] (44.8-53.7)
Marijuana use						
Ever	80	16.3% (12.8-20.7)	15.4% [36][b]	742	70.6% (66.7-74.1)	51.3% [36][b]
Past 30 days	47	8.7% (6.3-12.1)	10.8% [10] (8.7-13.2)	412	41.3% (37.4-45.3)	38.8% [10] (34.2-43.7)

[a]NSDUH: National Survey on Drug Use and Health.

[b]95% CI not provided for NSDUH national estimates of ever use.

Intervention and Retention

Randomization of the 1517 eligible participants was generally equal across study conditions, with 34.08% (517/1517) allocated to the "guaranteed" incentive, 32.63% (495/1517) to the "lottery" incentive, and 33.29% (505/1517) to the "preferred" incentive. Proportions were similar when looking at youth and young adult subgroups separately. There were no differences in the distribution of participants to the study condition by age, county of residence, sex, race, ethnicity, or employment status (Table 2). Response to the question about incentives for completing similar web-based surveys indicated a strong preference for the guaranteed incentive (1304/1517, 85.96%), and 67.96% (1031/1517) of participants were assigned to an incentive condition concordant with their preference.

At wave 2, 78.11% (1185/1517) of the full sample completed the survey. Youth retention at wave 2 was 82.3% (395/480), while young adult retention was 76.18% (790/1037; *P*=.007). At wave 3, 72.18% (1095/1517) of the full sample completed the survey; again, youth had higher retention (366/480, 76.3%) than young adults (729/1037, 70.30%; *P*=.02). Overall, 70.20% (1,065/1517) completed all three waves of the study (74.0% [355/480] of youth and 68.47% [710/1037] of young adults; *P*=.03).

Retention did not differ by incentive condition at wave 2 (*P*=.28) or wave 3 (*P*=.59). Retention was also similar when examining concordance between preference and the intervention condition, with 77.98% (804/1031) retention at wave 2 among those with concordant preference and an incentive condition compared with 78.4% (380/485) in the nonconcordant group (*P*=.87). Proportions were also similar at wave 3 (concordant: 751/1031, 72.84%; nonconcordant: 343/485, 70.7%; *P*=.39). Retention across all three waves did not differ by county, sex, race/ethnicity, or employment status. Retention was lower, however, among participants who had ever used a cigarette (*P*=.005) and those who reported past 30-day use of marijuana (*P*=.001; data available upon request).

Participant Feedback

Of the 1095 respondents in the wave 3 survey, 86.58% (948/1095) provided a response to the question regarding their favorite part of the survey and 81.55% (893/1095) provided a response to the question regarding potential improvements to the survey. Table 4 presents coded responses to each item. With some responses categorized into more than one code, there were 1049 responses to the item on the favorite part of participating in the study, with 28.60% (300/1049) noting financial compensation as their favorite part, followed by learning something new (192/1049, 18.30%) and making a meaningful contribution to science or to the community (184/1049, 17.54%). There were 910 responses to the item on potential improvements to the study, with the majority noting no changes needed (496/910, 54.5%), followed by specific recommendations related to ease of use (302/910, 33.2%). Recommendations included suggestions for improvements to survey structure, survey timing, and survey wording.

XSL•FO
RenderX

Table 4. Participant feedback at the end of the PACE Vermont Study.

Question, code, and subcode	Value, n (%)
What was your favorite part of participating in the PACE[a] Vermont Research Study? (1049 responses)	
Survey task (κ=0.50)	162 (15.4)
Favorite part was the survey task itself.	
Learning (κ=0.57)	192 (18.3)
Favorite part was learning something new.	
Learning: self (κ=0.91)	85 (8.1)
Enjoyed learning something about themselves or having an opportunity to self-reflect.	
Learning: other (κ=0.72)	89 (8.5)
Enjoyed learning new information about policies, organizations, or substances.	
Ease of use (κ=0.80)	110 (10.5)
Favorite part was the ease of use and accessibility of the surveys.	
Contribution (κ=0.83)	184 (17.5)
Favorite part was making a meaningful contribution to something (eg, science and the community).	
Compensation (κ=0.97)	300 (28.6)
Favorite part was the financial compensation.	
Other (κ=0.47)	80 (7.6)
Response not otherwise categorized.	
None (κ=0.71)	21 (2.0)
Did not generate a response (eg, "I did not have a favorite part").	
What could we improve in the PACE Vermont Research Study to make it easier for you to participate? (910 responses)	
Learning (κ=0.54)	16 (1.8)
Requested more resources or opportunities to learn.	
Ease of use (κ=0.88)	302 (33.2)
Suggested improving survey's ease of use.	
Survey design (κ=0.81)	119 (13.1)
Suggested improvement to survey structure or design.	
Reminders & timing (κ=0.87)	43 (4.7)
Suggested improvement to timing of surveys or the reminder system.	
Question improvement (κ=0.84)	140 (15.4)
Suggested improvement to wording or content of survey questions.	
Compensation (κ=0.88)	53 (5.8)
Suggested improvement to the compensation system.	
Other (κ=0.58)	43 (4.7)
Not otherwise categorized.	
None (κ=0.78)	496 (54.5)
No response generated (eg, "nothing").	
None: positive (κ=0.83)	244 (26.8)
Response generated was wholly positive.	

[a]PACE: Policy and Communication Evaluation.

Discussion

Principal Findings

This study identified successful recruitment strategies for a web-based cohort study of youth and young adults to inform and evaluate state-level substance use prevention efforts. It also tested the effect of three incentive conditions (guaranteed, lottery, and preferred) on retention over a 6-month period. Over a 10-week period in 2019, we were able to recruit 480 eligible youth and 1037 eligible young adults to the PACE Vermont Study. Findings from this study indicated that Facebook and Instagram advertising produced the greatest number of survey starts, followed by posts to a state-wide online neighborhood

XSL•FO
RenderX

forum in Vermont (Front Porch Forum) and Google advertisements (display and Gmail). The integration of Google and Facebook advertisements with the study landing page and its web analytics tool was critical to evaluating survey starts as the advertising metric of interest in support of PACE's recruitment efforts. Validation of study responses was achieved through multiple methods, in line with other web-based studies [37]. Data collected on county supported the distribution of responses across the state, in line with the distribution of youth and young adults in the population. The success of the local online neighborhood forum in driving traffic to the study website highlights the potential importance of these venues in recruitment. While community partner sources did not drive the same level of traffic to the site, advertisement of the study via these community organizations and through the local web-based digest may have lent credibility to the study and increased awareness that improved recruitment. Substance use prevalence estimates in PACE Vermont youth and young adults generally tracked national estimates from the NSDUH, although youth reported lower prevalence of current cigarette, alcohol, and marijuana use than estimates from NSDUH's state-level surveillance.

Retention was 78.11% (1185/1517) at 3 months and 72.18% (1095/1517) at 6 months. Contrary to our hypothesis, retention was equivalent across all incentive study conditions. This may have been due to a strong stated preference among study participants for the guaranteed payment and assignment of approximately two-thirds of participants to an incentive condition that was concordant with their preference. Youth participants at both waves had greater retention than young adult participants. Participants retained at all three waves were less likely to be ever cigarette users or past 30-day marijuana users compared with those lost to follow-up. The majority of participants retained at wave 3 provided feedback on their experience of the study, with largely positive comments about compensation, learning something new, and making a difference. Participants also provided specific feedback to improve future surveys, such as requests for more resources or opportunities to learn and suggestions to improve timing and reminders for surveys.

Limitations

The PACE Vermont Study was limited to a small largely rural state, and thus, successful recruitment strategies may not be generalizable to other study contexts. Additionally, while the sample was generally aligned with population distribution by county, there were imbalances by sex in the study sample, with the majority of the sample being female (1071/1517, 70.60%) in contrast to 48% of the state population of youth and young adults. There was also one county (Chittenden) with a higher response than expected according to population distribution, but this county is both the most populous and home to the University of Vermont with the largest population of undergraduates in the state who would have been eligible for the study. Lower prevalence of past 30-day cigarette, alcohol, and marijuana use in the PACE Vermont youth compared with state estimates suggests that youth enrolled in this study may represent a lower risk sample. This may be due to the recruitment process for youth that required parental consent to participate, thus attracting a lower risk pool of youth, or the smaller sample size of youth than young adults. Additional attrition by ever cigarette users and past 30-day marijuana users may have produced a sample with fewer risk behaviors at follow-up. Given the randomized nature of the incentive condition and balance in demographic characteristics across conditions, findings related to retention are likely to be generalizable to other web-based survey studies of youth and young adults.

Comparison With Prior Work

Similar to other web-based studies [28,38], Facebook and Instagram advertisements provided the greatest reach and lowest cost per survey start when considering parent, youth, and young adult screener surveys; these were followed by Google advertisements via display and Gmail. The success of these strategies may have been related to consistency with recommendations for Facebook recruitment, including having an attractive website [18] and existing social media accounts (@pace_vt) that identified the partner organizations involved in the study (ie, University of Vermont and Vermont Department of Health) and supported the credibility of the study [34].

Retention in our cohort of youth and young adults was higher at 6 months (72.18%) compared with a national cohort of young adults aged 15 to 21 years who also completed web-based surveys (63%) [27]. Our randomized experiment regarding incentive conditions showed, similar to previous studies [37-40], that multiple means of compensation produce equal retention when combined with providing completion bonuses and sending multiple reminders. However, our baseline item regarding incentive preference suggests that the majority of youth and young adults prefer a small guaranteed payment for responding to a web-based survey. Thus, providing guaranteed compensation could help to improve recruitment of youth and young adults in future studies. Our ongoing retention efforts draw on expertise gathered from in-person cohort studies of youth and young adults [25,29]; we send birthday postcards to participants at the start of each month and continue to post on our social media accounts to retain awareness of and engagement in the study.

Conclusions

Findings from the PACE Vermont Study demonstrated the feasibility of using traditional web-based advertising strategies (eg, Facebook and Google), in addition to web-based outreach through local community forums and organizations, to recruit a cohort of 1517 Vermont youth and young adults for a web-based study to evaluate state-level substance use prevention efforts. Participants were well distributed by age and county according to state population estimates and reported substance use prevalence comparable to national estimates in these age groups. Youth participants generally reported lower prevalence of risk behaviors compared with state-level estimates [10]. The higher proportion of female participants enrolled in the PACE Vermont Study is consistent with other studies, documenting higher recruitment of women to health studies via web-based advertising [38,40,41]. Retention in this web-based cohort study was over 70% at a 6-month follow-up and did not differ by the incentive condition, as seen in other studies of young adults

XSL•FO
RenderX

[27]. Participant feedback on the study experience was positive. Results from our study suggest that providing a guaranteed payment immediately upon survey completion coupled with a bonus for completing all survey waves and weekly survey reminders may facilitate retention in a cohort of youth and young adults. Future work in our cohort will assess the impact of other means of retaining participants, including ongoing cohort engagement via regular contact (eg, birthday postcards and social media posts), ongoing community engagement (eg, reporting to community partners on PACE Vermont Study outcomes), and nonfinancial incentives (eg, lottery for PACE-branded items).

Acknowledgments

The PACE Vermont Study was funded by the Vermont Department of Health, the University of Vermont Cancer Center, and the University of Vermont Larner College of Medicine. The effort of the author ACV was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number P20GM103644. Although this work was funded in whole or in part with monies provided by or through the State of Vermont, the State does not necessarily endorse the researchers' findings and/or conclusions. The findings and/or conclusions may be inconsistent with the State's policies, programs, and objectives. Additionally, the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

This randomized study was not registered as the authors state that it is a pilot study to determine best practices in recruitment and retention for other trials, which have been registered. The editor granted an exception. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

References

1. Erikson E. Identity: Youth and crisis. New York, NY: WW Norton & Company; 1968.
2. Oetting ER, Donnermeyer JF. Primary socialization theory: the etiology of drug use and deviance. I. Subst Use Misuse 1998 Mar;33(4):995-1026. [doi: 10.3109/10826089809056252] [Medline: 9548633]
3. Sussman S, Arnett JJ. Emerging Adulthood: Developmental Period Facilitative of the Addictions. Eval Health Prof 2014 Jun;37(2):147-155. [doi: 10.1177/0163278714521812] [Medline: 24492245]
4. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
5. Aldrich MC, Hidalgo B, Widome R, Briss P, Brownson RC, Teutsch SM. The role of epidemiology in evidence-based policy making: a case study of tobacco use in youth. Ann Epidemiol 2015 May;25(5):360-365 [FREE Full text] [doi: 10.1016/j.annepidem.2014.03.005] [Medline: 24875267]
6. Johnson RM, Guttmannova K. Marijuana Use among Adolescents and Emerging Adults in the Midst of Policy Change: Introduction to the Special Issue. Prev Sci 2019 Feb;20(2):179-184. [doi: 10.1007/s11121-019-0989-7] [Medline: 30701394]
7. Friedman AS, Wu RJ. Do Local Tobacco-21 Laws Reduce Smoking among 18 to 20 Year-Olds? Nicotine Tob Res 2020 Jun 12;22(7):1195-1201. [doi: 10.1093/ntr/ntz123] [Medline: 31348515]
8. Moss JL, Liu B, Zhu L. State Prevalence and Ranks of Adolescent Substance Use: Implications for Cancer Prevention. Prev Chronic Dis 2018 May 31;15:E69 [FREE Full text] [doi: 10.5888/pcd15.170345] [Medline: 29862962]
9. 2015-2016 National Survey on Drug Use and Health National Maps of Prevalence Estimates, by State. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. URL: https://www.samhsa.gov/data/sites/default/files/NSDUHsaeMaps2016/NSDUHsaeMaps2016.pdf [accessed 2018-10-14]
10. 2016-2017 NSDUH State Prevalence Estimates. Substance Abuse and Mental Health Services Administration. 2018 Nov 28. URL: https://www.samhsa.gov/data/report/2016-2017-nsduh-state-prevalence-estimates [accessed 2019-09-27]
11. Zezima K. Vermont is the first state to legalize marijuana through legislature. Washington Post. 2018. URL: https://www.washingtonpost.com/news/post-nation/wp/2018/01/23/vermont-is-the-first-state-to-legalize-marijuana-through-legislature/ [accessed 2019-09-24]
12. H.26 (Act 22): An act relating to restricting retail and Internet sales of electronic cigarettes, liquid nicotine, and tobacco paraphernalia in Vermont. Vermont State House of Representatives. URL: https://legislature.vermont.gov/bill/status/2020/H.26 [accessed 2019-09-27]
13. H.47 (Act 28): An act relating to the taxation of electronic cigarettes. Vermont State House of Representatives. URL: https://legislature.vermont.gov/bill/status/2020/H.47 [accessed 2019-09-27]
14. S.86: An act relating to increasing the legal age for buying and using cigarettes, electronic cigarettes, and other tobacco products from 18 to 21 years of age. Vermont State Senate. 2019. URL: https://legislature.vermont.gov/bill/status/2020/S.86 [accessed 2019-05-08]

XSL•FO
RenderX

15. Summary of Initial Report of Recommended Strategies to Governor Phil Scott. Vermont Opioid Coordination Council. 2018 Jan. URL: http://www.healthvermont.gov/sites/default/files/documents/pdf/OCC%20Report%20Summary%202-page.pdf [accessed 2018-04-26]
16. Lisa A. States with the biggest rural populations. Stacker. 2019 Apr 08. URL: https://thestacker.com/stories/2779/states-biggest-rural-populations#50 [accessed 2020-01-12]
17. Vermont - Rural Definitions: State-Level Maps. United States Department of Agriculture, Economic Research Service. 2018. URL: https://www.ers.usda.gov/webdocs/DataFiles/53180/25600_VT.pdf?v=0 [accessed 2020-01-12]
18. Policy and Communication Evaluation (PACE Vermont). URL: https://www.pacevt.org [accessed 2019-09-27]
19. Internet/Broadband Fact Sheet. Pew Research Center. 2018 Feb 05. URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2018-10-15]
20. Mobile Fact Sheet. Pew Research Center. 2018 Feb 05. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-10-15]
21. Young L, Montgomery M, Barnason S, Schmidt C, Do V. A Conceptual Framework for Barriers to the Recruitment and Retention of Rural CVD Participants in Behavior Intervention Trials. GSTF J Nurs Health Care 2015 Aug;2(2):19-24 [FREE Full text] [doi: 10.5176/2010-4804_2.2.71] [Medline: 28580381]
22. Young L, Barnason S, Do V. Conducting Behavioral Intervention Research in Rural Communities: Barriers and Strategies to Recruiting and Retaining Heart Failure Patients in Studies. Nurs Health Care (Winfield) 2016;1(1):101 [FREE Full text] [Medline: 29215663]
23. Young L, Barnason S. Review Strategies to Recruit and Retain Rural Patient Participating Self-management Behavioral Trials. Online J Rural Res Policy 2015;10(2):1-12 [FREE Full text] [doi: 10.4148/1936-0487.1070] [Medline: 28580049]
24. Tanner A, Kim S, Friedman DB, Foster C, Bergeron CD. Barriers to medical research participation as perceived by clinical trial investigators: communicating with rural and african american communities. J Health Commun 2015;20(1):88-96. [doi: 10.1080/10810730.2014.908985] [Medline: 25204763]
25. Diviak KR, Wahl SK, O'Keefe JJ, Mermelstein RJ, Flay BR. Recruitment and retention of adolescents in a smoking trajectory study: who participates and lessons learned. Subst Use Misuse 2006;41(2):175-182. [doi: 10.1080/10826080500391704] [Medline: 16393741]
26. Cantrell J, Hair EC, Smith A, Bennett M, Rath JM, Thomas RK, et al. Recruiting and retaining youth and young adults: challenges and opportunities in survey research for tobacco control. Tob Control 2018 Mar;27(2):147-154. [doi: 10.1136/tobaccocontrol-2016-053504] [Medline: 28432211]
27. Cantrell J, Bennett M, Thomas RK, Rath J, Hair E, Vallone D. It's Getting Late: Improving Completion Rates in a Hard-to-Reach Sample. Surv Pract 2017 Dec 28;11(2). [doi: 10.29115/sp-2018-0019]
28. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
29. Hanna KM, Scott LL, Schmidt KK. Retention strategies in longitudinal studies with emerging adults. Clin Nurse Spec 2014;28(1):41-45 [FREE Full text] [doi: 10.1097/NUR.0000000000000020] [Medline: 24309576]
30. Hark, Inc. URL: https://www.hark.bz [accessed 2019-09-27]
31. Front Porch Forum, Inc. URL: https://frontporchforum.com/ [accessed 2019-09-27]
32. Qualtrics International, Inc. URL: https://www.qualtrics.com [accessed 2019-09-27]
33. Rybbon, Inc. URL: https://www.rybbon.net [accessed 2019-09-27]
34. Miles M, Huberman A. Qualitative data analysis: An expanded sourcebook. Thousand Oaks, CA: Sage Publications; 1994.
35. Population of Vermont Counties by Single Year of Age and Sex, 2017. Vermont Department of Health. 2018 Dec 17. URL: https://www.healthvermont.gov/sites/default/files/documents/xls/HS_STAT_Population_of_Vermont_Counties_by_single_year_of_age_and_sex_2017.xls [accessed 2020-01-12]
36. 2018 NSDUH Detailed Tables. Substance Abuse and Mental Health Services Administration. 2019 Aug 20. URL: https://www.samhsa.gov/data/report/2018-nsduh-detailed-tables [accessed 2020-01-20]
37. Godinho A, Schell C, Cunningham JA. Out damn bot, out: Recruiting real people into substance use studies on the internet. Subst Abus 2020;41(1):3-5. [doi: 10.1080/08897077.2019.1691131] [Medline: 31821108]
38. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
39. Akers L, Gordon JS. Using Facebook for Large-Scale Online Randomized Clinical Trial Recruitment: Effective Advertising Strategies. J Med Internet Res 2018 Nov 08;20(11):e290 [FREE Full text] [doi: 10.2196/jmir.9372] [Medline: 30409765]
40. Byaruhanga J, Tzelepis F, Paul C, Wiggers J, Byrnes E, Lecathelinais C. Cost Per Participant Recruited From Rural and Remote Areas Into a Smoking Cessation Trial Via Online or Traditional Strategies: Observational Study. J Med Internet Res 2019 Nov 12;21(11):e14911 [FREE Full text] [doi: 10.2196/14911] [Medline: 31714253]
41. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

XSL•FO
RenderX

Abbreviations

NSDUH: National Survey on Drug Use and Health
PACE: Policy and Communication Evaluation

Edited by G Eysenbach; submitted 26.02.20; peer-reviewed by A Godinho, S Stevelink; comments to author 23.03.20; revised version received 18.04.20; accepted 14.05.20; published 20.07.20.

Please cite as:
Villanti AC, Vallencourt CP, West JC, Peasley-Miklus C, LePine SE, McCluskey C, Klemperer E, Priest JS, Logan A, Patton B, Erickson N, Hicks J, Horton K, Livingston S, Roemhildt M, Singer E, Trutor M, Williams R
Recruiting and Retaining Youth and Young Adults in the Policy and Communication Evaluation (PACE) Vermont Study: Randomized Controlled Trial of Participant Compensation
J Med Internet Res 2020;22(7):e18446
URL: http://www.jmir.org/2020/7/e18446/
doi:10.2196/18446
PMID:32706681

©Andrea C Villanti, Christie P Vallencourt, Julia C West, Catherine Peasley-Miklus, S Elisha LePine, Caitlin McCluskey, Elias Klemperer, Jeffrey S Priest, Alison Logan, Bill Patton, Nancy Erickson, Jennifer Hicks, Kathleen Horton, Shayla Livingston, Maria Roemhildt, Erin Singer, Megan Trutor, Rhonda Williams. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Social Media Approaches to Recruiting Young Adult Cigarillo Smokers: Cross-Sectional Study

David Cavallo[1], PhD; Rock Lim[2], BS; Karen Ishler[2], PhD; Maria Pagano[3], PhD; Rachel Perovsek[2], MPH; Elizabeth Albert[2], PhD; Sarah Koopman Gonzalez[4], PhD; Erika Trapl[4], PhD; Susan Flocke[5], PhD

[1]Department of Nutrition, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

[2]Center for Community Health Integration, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

[3]Division of Child Psychiatry, Department of Psychiatry, Case Western Reserve University, Cleveland, OH, United States

[4]Prevention Research Center for Healthy Neighborhoods, Departments of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, United States

[5]Department of Family Medicine, Oregon Health & Science University, Portland, OR, United States

Corresponding Author:
David Cavallo, PhD
Department of Nutrition
School of Medicine
Case Western Reserve University
10900 Euclid Avenue
Cleveland, OH, 44106-4954
United States
Phone: 1 216 368 6627
Email: david.cavallo@case.edu

Abstract

Background: The prevalence of social media use among youth and young adults suggests it is an appropriate platform for study recruitment from this population. Previous studies have examined the use of social media for recruitment, but few have compared platforms, and none, to our knowledge, have attempted to recruit cigarillo users.

Objective: The purpose of this study was to examine the effectiveness of different social media platforms and advertisement images for recruiting cigarillo users aged 14-28 years to complete a cigarillo use survey.

Methods: We obtained objective data for advertisement impressions for a 39-week social media recruitment campaign. Advertisements were targeted to users based on their age, geography, and interests. Effectiveness was defined as the percentage of approved surveys per advertising impression. Chi-square tests were performed to compare the effectiveness of different advertisement images and platforms.

Results: Valid survey completers (n=1089) were predominately older (25-28 years old, n=839, 77%). Of the 1089 survey completers, 568 (52%) identified as male, 335 (31%) as African American, and 196 (18%) as Hispanic. Advertisements delivered via Facebook/Instagram were more effective than Twitter; 311/1,027,738 (0.03%) vs 661/2,998,715 (0.02%); χ^2_1=21.45, N=4,026,453); *P*<.001. Across platforms, ads featuring exclusively an image of cigarillos were more effective (397/682,994, 0.06%) than ads with images of individuals smoking (254/1,308,675, 0.02%), individuals not smoking (239/1,393,134, .02%), and groups not smoking (82/641,650, 0.01%); $\chi^2_3$133.73, N=4,026,453; *P*<.001.

Conclusions: The campaign was effective in recruiting a diverse sample representative of relevant racial/ethnic categories. Advertisements on Facebook were more effective than Twitter. Advertisements that featured an image of a cigarillo were consistently the most effective and should be considered by others recruiting cigarillo users via social media.

(J Med Internet Res 2020;22(7):e12619) doi:10.2196/12619

KEYWORDS

adolescent, young adult; tobacco products; social media; research subject recruitment

XSL•FO
RenderX

Introduction

Recent data on smoking prevalence indicates that a substantial proportion of youth in the US smoke cigar products, including cigarillos [1]. Like other tobacco products, cigarillo use increases the risk of cancer, cardiovascular disease, and mortality, and some early evidence suggests that cigarillos may be more harmful to organ systems than cigarettes [2]. Cigarillo use is also concentrated among young, urban, minority populations that have traditionally experienced health disparities related to tobacco use and may be more vulnerable to the long-term harmful effects of smoking [3]. Considering the public health risk presented by cigarillo use, understanding patterns of use and the addictive properties of these products via survey research could serve as the first step in reversing current trends in their adoption and use by youth and young adults.

Recruiting cigarillo users for survey research presents unique challenges using traditional methods. Youth and young adults are less likely to respond to survey recruitment via telephone and change their residency often, making mail-based recruitment less effective [4]. Youth and young adults may also feel more comfortable completing surveys addressing behaviors such as tobacco use, which may be illegal or stigmatized among these participants, online rather than in a school setting or other public venue [5]. Web-based recruitment may, therefore, be more appropriate given the increasing comfort and amount of time spent with digital platforms in this population and the relative anonymity of accessing and completing surveys online [6].

Web-based recruitment using social media may be particularly appropriate in reaching youth and young adult cigarillo users [7]. In addition to being delivered on the web, social media reaches a diverse group of users, including significant proportions of low-socioeconomic status, minority populations, especially among younger age groups [8]. Advertising on social media also allows for the specific targeting of these groups using demographic characteristics as well as specific behaviors associated with tobacco use [9]. There is also a seamless user recruitment experience as social media advertisements link participants directly to online screeners and surveys.

A growing body of research has examined the use of social media advertisements for tobacco research recruitment. Studies have found social media advertising recruitment to be more efficient for recruiting smokers when compared to specialized mailings [10], research databases, course-based recruitment [11], and online panels [12]. Conversely, school-based referrals, bus ads, participant referrals, and fliers were found to be more cost-effective than social media when recruiting adolescent smokers [13]. Other studies that did not compare social media to alternative recruitment methods describe social media as effective in recruiting young adult smokers for survey and intervention research [14,15].

Most studies examining social media advertising recruitment have used Facebook [7]. The few direct comparisons of Twitter and Facebook for study recruitment have yielded mixed results, primarily used interpersonal modes of social media recruitment vs paid advertising, and assessed effectiveness with nonmonetary outcomes [16,17]. Given the significant risk of cigarillo use among youth and young adults, the growing evidence of the efficacy of social media in recruiting young adult smokers, and a lack of evidence supporting effective social media advertising strategies, the purpose of the current study was to examine the relative effectiveness of using Facebook, Instagram, and Twitter to recruit cigarillo users to complete an online nicotine dependence survey and provide guidance for future efforts using social media to study cigarillo use behavior.

Methods

Recruitment

Participants between 14 and 28 years old who smoked at least two cigarillos per week were considered eligible. Participants also had to read and understand English and have access to an internet-connected device. Study procedures were approved by the Case Western University Institutional Review Board, and participants provided informed consent. Our study did not collect individually identifiable data from social media platforms, reducing risks related to social media data privacy and anonymity recently raised in the research community [18]. Six social media advertisements were developed to run simultaneously on Facebook, Instagram, and Twitter platforms. Because advertisements were run together on a combined Facebook/Instagram campaign with the same URL, we were unable to separate the results for Facebook and Instagram at the survey completion level. We will refer to these combined platforms as “Facebook” when describing analyses for survey initiation and valid completion. To partially address this limitation, we analyzed advertising effectiveness between Facebook and Instagram at the level of advertisement clicks, which is available from the advertising management platform. Each advertisement contained identical wording that encouraged viewers to participate and featured a direct link to a brief eligibility screener administered via Qualtrics (Figure 1).

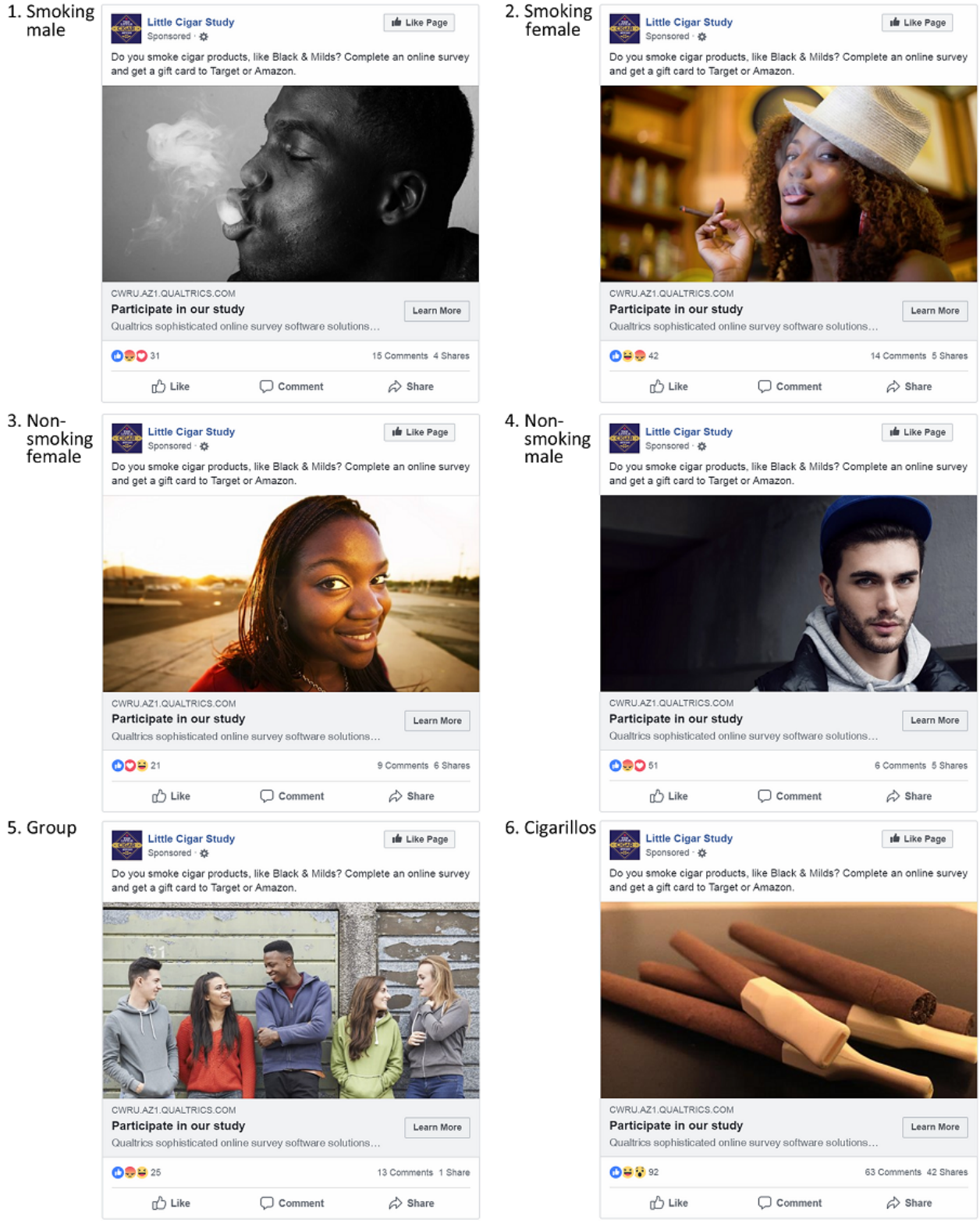
Advertisements featured planned images representing characteristics we intended to evaluate for advertising effectiveness (a cigar product, a smoking individual, a group of people, and gender). Advertisements were targeted to geographic areas reporting a high prevalence of cigarillo use (Cuyahoga County, OH; Baltimore, MD; Broward County, FL; Detroit, MI; DeKalb County, GA; Houston, TX; Philadelphia, PA; Washington, DC, Duval County, FL; and Fort Worth, TX) [1]. Advertisements were delivered daily for approximately 4 weeks in each location over 39 weeks beginning in June of 2017 and ending in March 2018. Advertisements were targeted using the age range most similar to our inclusion criteria that were an available option. We targeted Facebook advertisements to users 14-28 years old during the entire campaign and Twitter advertisements to users 13-34 years old, starting in November, once that feature became available. Advertisements were also targeted on relevant keywords such as “cigarillo” and “blunt” that were considered relevant to the target population. We did not encourage the sharing of social media advertisements through advertising messages or any other mechanism.

A budget of $240/day was set, and advertisements were scheduled to run with identical budgets so that advertisements would be delivered with similar frequency by the advertising

XSL•FO
RenderX

platform algorithm. This method was essential to allow a fair comparison of advertisements as less effective advertisements may run for a brief time when using social media advertising optimization algorithms. We used the cost/click method of payment traditionally used by advertisers wishing to direct users to a website whereby costs are based on an online auction for individual clicks on the advertisement.

Figure 1. Facebook advertisement examples.



XSL•FO
RenderX

Survey

Individuals who completed the screener and met the eligibility criteria were invited to take the survey. Invitations were delivered via a brief email or text message, based on participant preference, and included a link to a survey administered via Qualtrics. Reminders were sent via the preferred method at 2, 5, and 8 days after initially sending the survey link. Respondents who submitted valid surveys were remunerated with a $15 gift card.

The survey included questions assessing several dimensions relevant to cigarillo consumption, including demographic information, nicotine dependence, brand preference, group smoking behavior, psychosocial constructs, and the sequence of product initiation. Several sections of the survey used branching logic such that for each tobacco product endorsed, additional questions were assessed. Therefore, the number of survey items varied based on participants' reported product use.

Because nomenclature for tobacco products varies across audiences and geographically [19], several features were incorporated into the questions about tobacco products to ensure that items about tobacco products would be consistently understandable across participants with varying demographic and geographic characteristics. The questions referring to tobacco products used: the name of the product type (eg, "tipped cigarillo"), a familiar product brand (eg, "Black & Mild"), and a photograph of the product. Instructions and survey items were written at a 6th-grade literacy level. On average, the survey took about 30 minutes to complete.

Strategies to Ensure the Quality of the Survey Data

Multiple strategies were used to ensure the quality of online survey data, which can be adversely affected by respondents who falsely complete surveys for financial gain [20]. First, for the screener survey, we analyzed contact information items (email address and phone number) for validity and consistency and to identify possible repeat respondents [21]. Second, the screener survey included an open-answer item asking for a brief description of how participants got started smoking little cigars or cigarillos. This item required a multiple word response and was used to screen out responses generated by computer programs. This question was manually evaluated before a link to the survey was sent to the eligible participant. Third, a question about how the participant heard about the survey was used to ensure that participants came across the survey legitimately. If they claimed to hear about the survey through a channel that was not used by us, their responses were flagged and examined more carefully.

Completed surveys were manually assessed by study staff for validity before analysis. Validation was performed through the use of "trap" or "red herring" questions [22], checking for surveys completed abnormally fast or with poor response consistency [23], checking for unusually repetitive responses ("straight-lining") [24], and checking response sets with similar email addresses or IP addresses. Surveys that exhibited these characteristics were reviewed more carefully. Suspected repeat responders (based on emails or IP addresses) were rejected if there was sufficient evidence (based on the reviewer's judgment) to do so. For example, if respondents with the same IP address gave similar answers, the surveys were completed one immediately after the other, or the respondents had emails sufficiently similar to suggest they were, in fact, the same person. Suspected satisficing cases were reviewed and either outright rejected or placed in a questionable category for further review. Multiple members of the study staff then reviewed these cases to make a final determination to accept or reject the survey.

Measures

Data on advertising effectiveness were obtained objectively from the Facebook/Instagram and Twitter advertising analytics platforms. Data included advertisement impressions (number of individuals to whom the advertisement was delivered), cost, number of times an advertisement was clicked, and number of times an advertisement was shared on the social media platform. The advertisement that generated each survey was identified by linking the response to a unique URL. Survey data included demographics and metadata related to survey and screener completion.

Statistical Analysis

Statistical analyses were conducted with SAS (Version 9.2) and SPSS (Version 25). Descriptive statistics were used to assess participant characteristics and sharing rates. Advertising effectiveness by platform and content is expressed as ratios. These included the number of screeners started per impression, the number of valid surveys per impression, and the cost per started screener and valid survey. We chose started screeners and valid surveys as the indicators of recruitment effectiveness except for the analyses of Facebook vs Instagram, where clicks were used as a metric of effectiveness. We also calculated the time in hours from when a participant started a screener to when they completed a survey. Differences between advertisement effectiveness were examined using Pearson chi-square tests. Analysis of which participant characteristics predicted recruitment advertising platform was performed using logistic regression models. Differences in survey time to completion were examined using *t* tests.

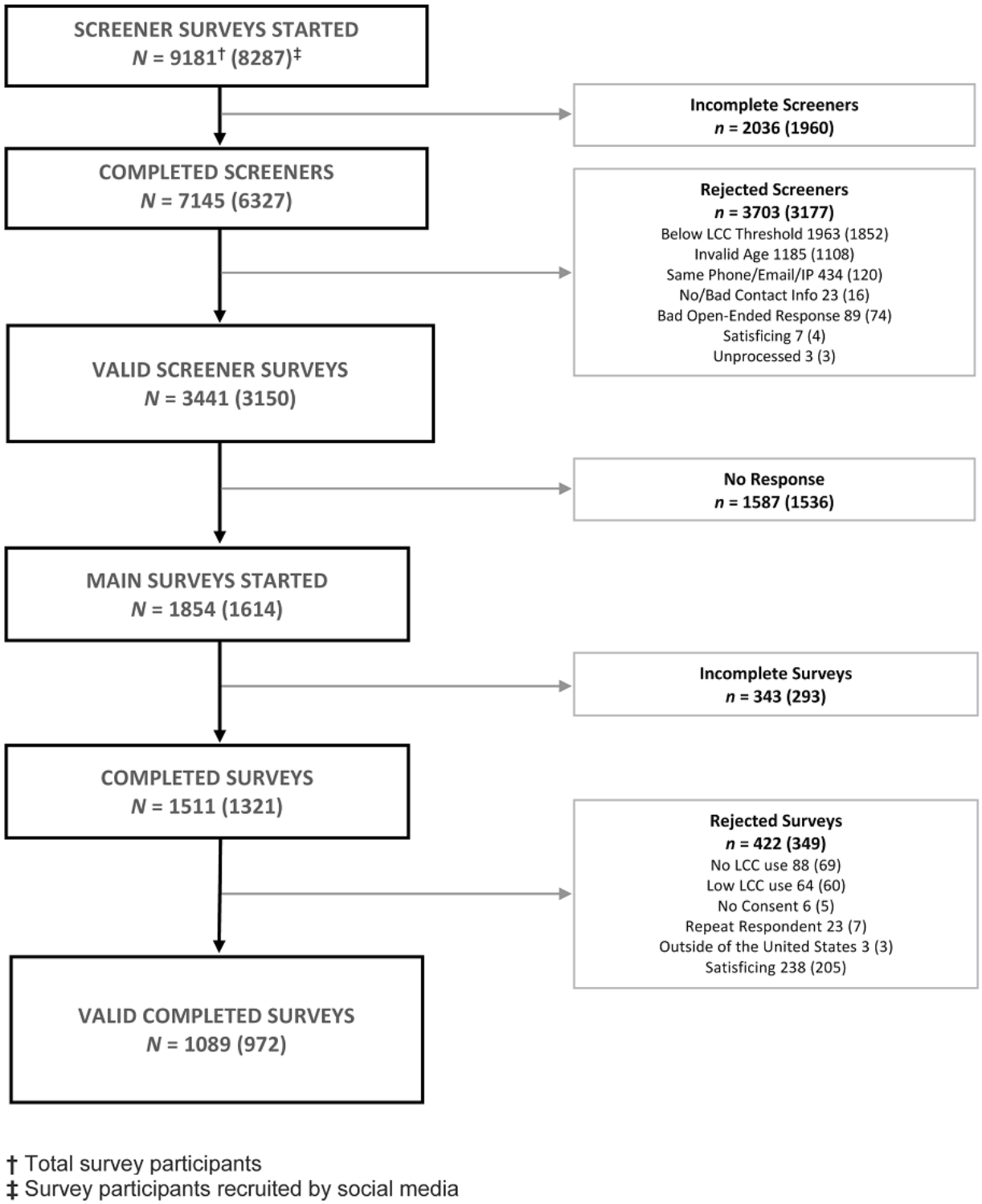
Results

The combined social media advertising campaign generated 4,026,453 impressions, and 8287 started screeners. Of 6327 completed screeners, 3150 (50%) were identified as eligible and were sent a link to the survey (Figure 2). Slightly more than half of the eligible participants started the survey (1614/3150, 51%), and 1321 of those completed it. Of the completed surveys, 349 were rejected due to a low level of cigarillo use, satisficing, or not meeting other review criteria. Table 1 shows the demographic characteristics of the total valid survey completers (n=1089) and valid survey completers who were recruited using social media (n=972). We found no significant differences in gender, age, race/ethnicity, and education between those who were recruited via Twitter or Facebook and subjects recruited by other methods (eg, palm cards, referral by another participant, n=117). Of note, 509 (47%) of the total sample were female, 839 (77%) were aged 25-28, 335 (31%) were African American, and 196 (18%) were Hispanic. The median number of cigarillos

XSL•FO
RenderX

per day for all valid survey completers was 0.93. In addition, 585 (53.7%) also smoked cigarettes, and 866 (79.5%) reported using at least one other combustible tobacco product in the past 30 days.

Figure 2. Screener to survey completion flowchart.



Tables 2-3 describe advertisement effectiveness and cost comparisons. There was no significant difference between the percent of started screeners by impressions between Facebook and Twitter. Advertisements featuring a female image were more successful in generating started screeners than those with male images, and advertisements with a cigarillo image were more effective than those without. Facebook generated a greater percentage of valid surveys per impression than Twitter. There was no significant difference by the gender of the advertising image, and advertisements with a cigarillo image again were markedly more effective. Facebook was less cost-effective for generating started screeners but more cost-effective for producing valid surveys. Facebook was significantly more effective in generating clicks on the study advertisement per impression than Instagram (n=8507, 1.76% vs n=1198, 1.16%,

XSL•FO
RenderX

χ^2_1=188.24, *P*<.001, N=585,648). Instagram was also more expensive than Facebook per click obtained ($1.32 US vs $0.81 US). Facebook and Twitter significantly differed in the interval between respondents starting a screener and completing a survey (mean 66.63 h, SD 69.60 vs mean 80.03 h, SD 79.14; $t=2.55_{967}$, *P*=.01). Across platforms, advertisements were shared 310 times (79 on Facebook, 231 on Twitter).

Analysis of the predictors of social media recruitment platform is detailed in Table 4.

Gender and race/ethnicity of survey completers did not predict through which social media platform a participant was recruited. The odds of being recruited by Facebook (n=310) versus Twitter (n=659) were 1.5-fold higher for those aged ≥21 years when compared to adolescents. Those recruited via Facebook were also more likely to report lower levels of education.

Table 1. Demographic characteristics and tobacco use of valid survey completers.

Characteristic	Total valid survey completers (N=1089), n (%)	Valid survey completers recruited via social media (N=972), n (%)
Age		
14-18	19 (1.7)	15 (1.54)
19-24	231 (21.2)	202 (20.78)
25-28	839 (77.0)	755 (77.67)
Gender		
Female	509 (46.7)	448 (46.1)
Male	568 (52.2)	514 (52.9)
Other	12 (1.1)	10 (1.0)
Race/ethnicity		
Black or African American	335 (30.8)	309 (31.9)
Hispanic	196 (18.0)	170 (17.5)
White	436 (40.1)	385 (39.7)
Other	120 (11.0)	106 (10.9)
Education		
<GED or high school	109 (10.0)	98 (10.1)
GED or high school	235 (21.6)	199 (20.5)
Some college	432 (39.7)	396 (40.8)
Associate	93 (8.5)	83 (8.6)
BA+	219 (20.1)	195 (20.1)
Tobacco use		
Number of cigarillos smoked/day (median)	0.93	0.90
Cigarette smoker	585 (53.7)	520 (53.5)
Smoked at least one other combustible tobacco product	866 (79.5)	772 (79.4)

XSL•FO
RenderX

Table 2. Advertising effectiveness by platform, image type, and cost (N=4,026,453).

Advertisement type	Impressions, n (%)	Started screeners	%	Chi-square (*df*)	*P* value
All	4,026,453 (100)	8,287			
Platform					
Twitter	2,998,715 (74)	6,203	.21	0.62 (1)	.44
Facebook	1,027,738 (26)	2,084	.20		
Image gender					
Female	1,303,549 (32)	2,249	.17	15.55 (1)	<.001
Male	1,398,260 (35)	2,142	.15		
Image type					
Cigarillo	682,994 (17)	3,196	.47	941.75 (3)	<.001
Smoking	1,308,675 (32)	2,174	.17		
Nonsmoking	1,393,134 (35)	2,217	.16		
Group	641,650 (16)	700	.11		

Table 3. Advertising effectiveness continued.

Advertisement type	Valid surveys	%	Chi-square (*df*)	*P* value	Cost/started screener	Cost/valid survey
All	972				$4.73	$40.34
Platform						
Twitter	661	0.02	21.45 (1)	<.001	$4.63	$43.41
Facebook	311	0.03			$5.05	$33.82
Image gender						
Female	246	0.02	0.54 (1)	.38	$5.65	$51.69
Male	247	0.02			$6.17	$53.54
Image type						
Cigarillo	397	0.06	133.73 (3)	<.001	$2.19	$17.62
Smoking	254	0.02			$5.99	$51.26
Nonsmoking	239	0.02			$5.83	$54.06
Group	82	0.01			$8.97	$76.54

Table 4. Odds of recruitment via Facebook vs Twitter (N=969).

Demographic	Coefficient	SE	*P* value	OR[a] (95% CI)
Male gender (vs female and other)	–0.282	.145	.06	0.755 (0.568-1.002)
Age 21-28 (vs 14-20)	0.372	.180	.04	1.451 (1.019-2.065)
Race/ethnicity				
Black or African American	–0.247	.176	.16	0.781 (0.553-1.104)
Hispanic	0.104	.198	.60	1.110 (0.752-1.637)
Other	0.113	.239	.64	1.120 (0.701-1.788)
Education				
<GED or high school	1.422	.304	<.001	4.147 (2.286-7.521)
GED or high school	1.447	.250	<.001	4.250 (2.603-6.939)
Some college	0.983	.228	<.001	2.672 (1.710-4.173)
Associate's degree	1.019	.306	<.001	2.770 (1.519-5.050)

[a]OR: odds ratio.

XSL•FO
RenderX

Discussion

This study was successful in recruiting a large sample of youth and young adult cigarillo users, supporting the use of social media-based advertising to reach this group, which traditionally presents a challenge to survey and study recruitment efforts. The recruited sample was diverse and representative of individuals most likely to use cigarillos, with over a third of our sample consisting of African Americans, a demographic group with the highest reported cigarillo use among adults [3].

Advertisements on Facebook were more effective than Twitter in producing valid surveys. Facebook was also a more cost-effective recruitment method. Differences observed in the cost per approved survey (approximately $10 US) could represent a significant increase in the overall costs of recruitment in extensive studies. A limited number of previous studies have compared Facebook and Twitter research recruitment. A feasibility study examining the use of QR codes, Facebook, and Twitter for recruiting adolescents to take a health-related survey found Twitter to be marginally more cost-effective than Facebook [16]. That study, however, used student seeds to deliver Twitter recruitment as opposed to advertising. In another study that recruited focus groups about vaccination, Facebook was found to be more efficient than Twitter when comparing staff time spent per questionnaire received, but this study also used individual Facebook and Twitter accounts, not advertisements [17].

Our cost per valid survey of $40 US falls within a broad range of previous findings related to the cost-effectiveness of social media recruitment of tobacco users. Ramo et al reported substantially lower Facebook costs of $8.80 US per eligible, consented participant in a tobacco cessation intervention and $4.28 per completed survey in a separate study, both targeting young adults [14,15]. Other recent studies recruiting tobacco-related survey participants reported a cost per completed survey of $1.86 US (Facebook) [11], $21.73 US (Twitter) [12], and a cost per enrolled smoker for randomized controlled trials ranging from $41 to $62 US (Facebook) [25,26]. Many factors affect social media advertising costs, including seasonality, industry type, and the specificity of the audience being targeted [27]. It might be the case that a more specific audience (eg, individuals who smoke cigarillos) reduces the cost-effectiveness of the advertisements. In addition, social media advertising costs have risen in recent years as more advertisers use the platform and as a result of recent changes to the advertising algorithm, which may account for some of the increase from previous studies [28]. When comparing social media costs in the current study to other modalities for recruiting tobacco users, Facebook and Twitter advertisements fall in the middle of the cost-effectiveness range. For instance, Brodar et al reported a cost per enrolled smoker in an experimental trial ranging from $7 US using Craigslist to $375 US for newspaper advertisements [25].

Our results also demonstrate that differences in effectiveness are a function of the metric used. Twitter was as effective in generating started screeners as Facebook but was less effective in generating approved surveys. Investigators should use the measure of effectiveness most closely related to their recruitment goal as opposed to more distal measures such as advertising clicks that are more accessible via social media advertising platforms. Participants recruited via Facebook also completed the process of producing a valid survey more rapidly than those recruited via Twitter. For large studies or those on a short timeline, this measure of performance should be considered to ensure that the study goals are met. We observed minimal naturally occurring sharing of advertisements across platforms when compared to the reach of paid advertising. The viral nature of social media platforms, however, lends itself to this mode of recruitment, and researchers are beginning to use this strategy for a variety of recruitment and communications tasks [29,30]. Given the fact that sharing is a cost-free method of disseminating social media advertisements, future research should test methods for increasing the effectiveness of viral social media survey recruitment, such as nonmonetary rewards or small monetary incentives.

Regardless of the platform, the inclusion of a cigarillo image was consistently the most effective advertisement design. A similar result was found in a comparison of Facebook advertisement images used to recruit young smokers, where the advertisement with an image of a cigarillo had the lowest cost per unique click of 4 images used [15]. This strategy was also used in another successful Twitter recruitment campaign where all advertisements featured a prominent image of a tobacco product [12]. The impact of the image may be the result of having a strong visual cue similar to existing cigarillo marketing or the simplicity of the image given the limited space available for social media advertisements.

This study has several strengths. We had a large sample of advertising impressions on which to base our analysis. We created individual URLs for each advertisement on each platform linked to individual participants' screeners and surveys, allowing us to examine more relevant forms of effectiveness and assess demographic factors associated with social media platform advertising effectiveness. Limitations include our inability to examine Facebook and Instagram separately at the same level of effectiveness as our other analyses. The demographic profile of these social media platforms differs on several key demographic factors such as age and race, which may create differences in their effectiveness as suggested by our results comparing clicks per impression. In addition, over the past five years, Instagram use has consistently increased while Facebook has leveled off [31]. Understanding the effectiveness of Instagram should increase in importance if this trend continues. This study is not experimental, meaning that individuals were not randomized to the social media platform exposure. This limitation creates the possibility that participants may have been exposed to advertisements from Facebook and Twitter and responded only after being exposed to a specific dose. Additionally, the targeting features of social media platforms, such as geo-targeting, are inferred based on data collected in part through self-report, which may reduce their accuracy.

Social media platforms should be considered for tobacco-related research, especially in studies attempting to recruit hard-to-reach populations, such as youth, young adults, and minorities. This

XSL•FO
RenderX

method is cost-effective when compared to other modalities and provides a convenient means of digital recruitment on a nationwide basis. Although in the current study, Facebook was a more cost-effective platform, given the fact that we observed some differences in the likelihood of recruitment platform between age and education levels and our inability to determine how many individuals in the sample use both, it may be that using multiple platforms remains the optimal recruitment strategy. Using multiple social media platforms may be less cost-effective, but by increasing the potential audience size, researchers may improve the ability to reach subgroups and recruit rapidly, if required.

Acknowledgments

This project was funded by a grant from NIH, the National Cancer Institute, and the Center for Tobacco Prevention, R01CA109130.

Conflicts of Interest

None declared.

References

1. Kann L, Kinchen S, Shanklin S, Flint K, Hawkins J, Harris W. Youth risk behavior surveillance - United States, 2013. Morbidity and Mortality Weekly Report Surveillance Summaries 2014;63(4):1-168. [doi: 10.1037/e307132005-001]
2. Ghosh A, Abdelwahab SH, Reeber SL, Reidel B, Marklew AJ, Garrison AJ, et al. Little Cigars are More Toxic than Cigarettes and Uniquely Change the Airway Gene and Protein Expression. Sci Rep 2017 Dec 27;7:46239 [FREE Full text] [doi: 10.1038/srep46239] [Medline: 28447619]
3. Hu SS, Neff L, Agaku IT, Cox S, Day HR, Holden-Hayes E, et al. Tobacco product use among adults - United States, 2013 - 2014. MMWR Morbidity and mortality weekly report 2016;65(27):685-691.
4. Stern MJ, Bilgen I, Dillman DA. The State of Survey Methodology. Field Methods 2014 Feb 10;26(3):284-301. [doi: 10.1177/1525822X13519561]
5. Kreuter F, Presser S, Tourangeau R. Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. Public Opinion Quarterly 2009 Jan 26;72(5):847-865. [doi: 10.1093/poq/nfn063]
6. Cantrell J, Hair EC, Smith A, Bennett M, Rath JM, Thomas RK, et al. Recruiting and retaining youth and young adults: challenges and opportunities in survey research for tobacco control. Tob Control 2017 Apr 21;27(2):147-154. [doi: 10.1136/tobaccocontrol-2016-053504]
7. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
8. Perrin A. Social Media Usage: 2005-2015. Pew Research Center. Washington, DC: Pew Research Center; 2015. URL: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/ [accessed 2017-03-02] [WebCite Cache ID 6ofC9xmEt]
9. Akers L, Gordon JS. Using Facebook for Large-Scale Online Randomized Clinical Trial Recruitment: Effective Advertising Strategies. J Med Internet Res 2018 Nov 08;20(11):e290 [FREE Full text] [doi: 10.2196/jmir.9372] [Medline: 30409765]
10. Derrick JL, Eliseo-Arras RK, Hanny C, Britton M, Haddad S. Comparison of internet and mailing methods to recruit couples into research on unaided smoking cessation. Addict Behav 2017 Dec;75:12-16. [doi: 10.1016/j.addbeh.2017.06.012] [Medline: 28662435]
11. Thornton LK, Harris K, Baker A, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. Drug Alcohol Rev 2015 Jul 14;35(4):494-502. [doi: 10.1111/dar.12305] [Medline: 26179224]
12. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing Twitter and Online Panels for Survey Recruitment of E-Cigarette Users and Smokers. J Med Internet Res 2016 Nov 15;18(11):e288 [FREE Full text] [doi: 10.2196/jmir.6326] [Medline: 27847353]
13. Rait MA, Prochaska JJ, Rubinstein ML. Recruitment of adolescents for a smoking study: use of traditional strategies and social media. Transl Behav Med 2015 Sep;5(3):254-259 [FREE Full text] [doi: 10.1007/s13142-015-0312-5] [Medline: 26327930]
14. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
15. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64. [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
16. Gu L, Skierkowski D, Florin P, Friend K, Ye Y. Facebook, Twitter, & Qr codes: An exploratory trial examining the feasibility of social media mechanisms for sample recruitment. Computers in Human Behavior 2016;60:86-96 [FREE Full text] [doi: 10.1016/j.chb.2016.02.006]

XSL•FO
RenderX

17. Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, et al. The good, bad, and ugly of online recruitment of parents for health-related focus groups: lessons learned. J Med Internet Res 2013 Jan;15(11):e250 [FREE Full text] [doi: 10.2196/jmir.2829] [Medline: 24231040]
18. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical Issues in Social Media Research for Public Health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
19. Dickinson DM, Johnson SE, Coleman BN, Tworek C, Tessman GK, Alexander J. The Language of Cigar Use: Focus Group Findings on Cigar Product Terminology. NICTOB 2016 Jan 29;18(5):850-856. [doi: 10.1093/ntr/ntv285]
20. Mustanski BS. Getting wired: Exploiting the internet for the collection of valid sexuality data. Journal of Sex Research 2001 Nov;38(4):292-301. [doi: 10.1080/00224490109552100]
21. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying Multiple Submissions in Internet Research: Preserving Data Integrity. AIDS Behav 2008 Feb 1;12(6):964-973. [doi: 10.1007/s10461-007-9352-2]
22. Miller J, Baker-Prewitt J. Beyond 'trapping'the undesirable panelist: The use of red herrings to reduce satisficing. 2009 Presented at: CASRO Panel Quality Conference; 2/3/2009; New Orleans, LA. [doi: 10.1016/b978-0-443-06914-7.00010-2]
23. Jones MS, House LA, Gao Z. Respondent Screening and Revealed Preference Axioms: Testing Quarantining Methods for Enhanced Data Quality in Web Panel Surveys. PUBOPQ 2015 Jun 02;79(3):687-709. [doi: 10.1093/poq/nfv015]
24. Rogers F, Richarme M. The honesty of online survey respondents: Lessons learned and prescriptive remedies. Decision Analyst, Inc White Papers 2009:1-5. [doi: 10.1002/9780470688335.ch9]
25. Brodar KE, Hall MG, Butler EN, Parada H, Stein-Seroussi A, Hanley S, et al. Recruiting Diverse Smokers: Enrollment Yields and Cost. Int J Environ Res Public Health 2016 Dec 16;13(12):1251 [FREE Full text] [doi: 10.3390/ijerph13121251] [Medline: 27999280]
26. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
27. Irvine M. Facebook Ad Benchmarks for YOUR Industry [New Data]. WordStream. 2018. URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2018-04-09]
28. AdStage's Q4 2017 Paid Search and Paid Social Benchmark Report. AdStage.: AdStage; 2018. URL: https://www.adstage.io/resources/q4-ppc-benchmark-report/ [accessed 2020-02-15]
29. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
30. Cobb NK, Jacobs MA, Wileyto P, Valente T, Graham AL. Diffusion of an Evidence-Based Smoking Cessation Intervention Through Facebook: A Randomized Controlled Trial. Am J Public Health 2016 Jun;106(6):1130-1135. [doi: 10.2105/AJPH.2016.303106] [Medline: 27077358]
31. Social Media Fact Sheet. Pew Research Center. Washington, DC: Pew Research Center; 2019 May 12. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-02-15]

Edited by G Eysenbach; submitted 31.10.18; peer-reviewed by A Gough, W Tang; comments to author 05.04.19; revised version received 20.05.19; accepted 25.05.19; published 22.07.20.

Please cite as:
Cavallo D, Lim R, Ishler K, Pagano M, Perovsek R, Albert E, Koopman Gonzalez S, Trapl E, Flocke S
Effectiveness of Social Media Approaches to Recruiting Young Adult Cigarillo Smokers: Cross-Sectional Study
J Med Internet Res 2020;22(7):e12619
URL: https://www.jmir.org/2020/7/e12619
doi:10.2196/12619
PMID:32459629

©David Cavallo, Rock Lim, Karen Ishler, Maria Pagano, Rachel Perovsek, Elizabeth Albert, Sarah Koopman Gonzalez, Erika Trapl, Susan Flocke. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Privacy-Preserving Properties of Word Embeddings: Algorithmic Validation Study

Mohamed Abdalla[1,2,3], BSc, MSc; Moustafa Abdalla[4,5,6], BSc, DPhil; Graeme Hirst[1,2], BSc, PhD; Frank Rudzicz[1,2,7,8], BSc, MSc, PhD

[1]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[2]The Vector Institute for Artificial Intelligence, Toronto, ON, Canada

[3]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

[4]Deptartment of Statistics, Computational Statistics & Machine Learning Group, University of Oxford, Oxford, United Kingdom

[5]Wellcome Centre for Human Genetics, Nuffield Dept of Medicine, University of Oxford, Oxford, United Kingdom

[6]Harvard Medical School, Boston, MA, United States

[7]International Centre for Surgical Safety, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada

[8]Surgical Safety Technologies Inc, Toronto, ON, Canada

Corresponding Author:
Mohamed Abdalla, BSc, MSc
Department of Computer Science
University of Toronto
Bahen Centre for Information Technology
40 St. George Street, Room 4283
Toronto, ON, M5S 2E4
Canada
Phone: 1 4169787816
Email: mohamed.abdalla@mail.utoronto.ca

Abstract

Background: Word embeddings are dense numeric vectors used to represent language in neural networks. Until recently, there had been no publicly released embeddings trained on clinical data. Our work is the first to study the privacy implications of releasing these models.

Objective: This paper aims to demonstrate that traditional word embeddings created on clinical corpora that have been deidentified by removing personal health information (PHI) can nonetheless be exploited to reveal sensitive patient information.

Methods: We used embeddings created from 400,000 doctor-written consultation notes and experimented with 3 common word embedding methods to explore the privacy-preserving properties of each.

Results: We found that if publicly released embeddings are trained from a corpus anonymized by PHI removal, it is possible to reconstruct up to 68.5% (n=411/600) of the full names that remain in the deidentified corpus and associated sensitive information to specific patients in the corpus from which the embeddings were created. We also found that the distance between the word vector representation of a patient's name and a diagnostic billing code is informative and differs significantly from the distance between the name and a code not billed for that patient.

Conclusions: Special care must be taken when sharing word embeddings created from clinical texts, as current approaches may compromise patient privacy. If PHI removal is used for anonymization before traditional word embeddings are trained, it is possible to attribute sensitive information to patients who have not been fully deidentified by the (necessarily imperfect) removal algorithms. A promising alternative (ie, anonymization by PHI replacement) may avoid these flaws. Our results are timely and critical, as an increasing number of researchers are pushing for publicly available health data.

(J Med Internet Res 2020;22(7):e18055) doi:10.2196/18055

KEYWORDS

privacy; data anonymization; natural language processing; personal health records

XSL•FO
RenderX

Introduction

Motivation

Natural language processing (NLP) is increasingly used to assist medical practitioners with various tasks, ranging from patient phenotyping to unplanned hospital readmission prediction [1-3]. Although a diverse range of approaches are used, a large number of NLP applications use algorithms, such as Continuous Bag of Words (CBOW), Skipgram, and Global Vectors (GloVe) [4,5], which represent tokens as dense numeric vectors termed as *word embeddings*. Most of these representations are computed from large corpora of text, such as clinical notes or narratives from health records, made available by health care providers (HCPs). Usually, before these data are provided to researchers, the HCPs apply anonymization algorithms to deidentify the personal health information (PHI) in the data. In this work, we adopted the terminology of the US Health Insurance Portability and Accountability Act (HIPAA), where *PHI* refers to *individually identifiable health information*, which includes personal identifiers ranging from names and phone numbers to fingerprints.

There is a wide variety of techniques to locate and deidentify PHI in clinical text, ranging from dictionaries [6] to recurrent neural networks [7]. Once the sensitive information is located within a record, anonymization can employ either *removal* or *replacement*, that is, the sensitive information is either simply deleted, changed to a data-type identification tag such as **NAME**, or replaced with another randomly chosen PHI of the same type. Many publicly available resources use PHI removal, for example, the Multiparameter Intelligence Monitoring in Intensive Care (MIMIC-III) dataset [8] used informative deidentification tags. However, in this paper, we showed that as no perfect PHI search algorithm exists, data *secured* this way can be exploited because traces of identities remain in the text and are detectable even in embeddings that are generated from it.

Specifically, we discussed the privacy concerns that arise from publicly releasing word embeddings that have been trained on clinical notes secured using the PHI removal paradigm. At first glance, it may seem that releasing word embeddings has low risk because of the unordered nature of these models; all that is released is a list of words, arbitrarily ordered, with dense numeric vectors associated with each word. However, through our experiments with three of the most popular embedding techniques, we showed that they can be leveraged to learn information presumed to be removed.

Our work relies on the assumption that some name tokens will inevitably be missed by the deidentification process. This is a realistic assumption as, to date, there is no deidentification algorithm that has perfect recall (ie, captures all PHI). This necessarily means that the word list of the embedding model will contain names that are not properly protected. We also assume that malicious actors will be able to successfully identify these tokens from a very large wordlist. Given these two assumptions, a publicly released traditional word embedding model then presents a small, but nontrivial, risk of patient identities being attacked. This risk is relative to the number of patients in the data set and the particular deidentification and embedding algorithms used. Up to 0.6% of all patients may be at risk of having their full names detected in a data set (built from individual name tokens), and as many as 0.02% to 0.15% may have their full name associated with a diagnosis. Although these risks appear small, with the growing number of publicly available embeddings trained on clinical data, we aimed to draw attention to the possible critical mass of potential privacy exposure.

Specifically, we showed that (1) it is possible to associate name tokens together to form *true* name pairs, (2) there is a significant difference between the distances of diagnoses that have been associated with a patient and those of diagnoses not associated, and this is true both at the population level and at the patient level, and (3) it is possible for a malicious actor to determine diagnoses assigned to multiple patients, using only precomputed embeddings. In this work, we will refer to diagnostic codes and diagnoses interchangeably, although this is not, of course, a general equivalence. Here, we take the diagnostic code simply as an indication of the condition that the patient is suspected of having, which is sensitive information that must be protected. Finally, we replicate these results and perform further experiments with a synthetic data set that we make publicly available.

Our work is the first to study the privacy implications of releasing word embeddings. This demonstrates how anonymizing clinical notes using PHI removal is likely to leave sensitive patient information vulnerable. By methodically exploring a variety of algorithms and hyperparameters, we showed that our observation holds in the general case. Furthermore, we demonstrated that it is easier to reassociate sensitive information with rare names compared with common ones. Finally, we argue that, given our results, data holders and providers should explore whether other paradigms, such as PHI replacement, are more successful in securing sensitive information when compared with PHI removal.

Background

Clinical Word Embeddings

Word embeddings (ie, *word vectors* or *distributed representations*) are dense numeric vectors used to represent words. Many word embedding techniques fall into one of two categories: low-rank approximations of co-occurrence matrices [4,9] and those created using shallow neural networks using contextual information [10]. There is also a recent and growing body of embedding models employing deeper neural networks to create contextual word embeddings, which vary depending on the surrounding context [11,12].

Inspired by the distributional hypothesis [13], word embeddings trained on health care data are strongly correlated with human-annotated word similarity metrics for medical terms [2], although their performance on clinical classification tasks is strongly dependent on the quality, size, and type of data from which they are created [14]. In fact, embeddings created from clinically related data (eg, clinical notes and biomedical text, such as a collection of all PubMed Central articles and PubMed

XSL•FO
RenderX

abstracts), often performed better than, and never performed worse than, unspecialized corpora [2].

Until recently, there had been no publicly released embeddings trained on clinical data [15-18]. However, some newly released embeddings [15-18] are trained using contextual word embedding models on MIMIC, which itself uses PHI removal to abide by HIPAA regulations. Our work demonstrates how, if no additional security measures are taken, then traditional (noncontextual) models may be compromised. More work is required to assess whether our findings hold for the four new models as well (ie, contextual word embeddings) [15-18].

Privacy of Clinical Notes

There are 3 main approaches to protect the privacy of patients: dictionary-based, statistical, and hybrid approaches. Dictionary-based methods often use large wordlists or predefined regular expressions to locate private information in the text [6,19]. Statistical methods, often more robust than dictionary-based approaches [20], use models such as recurrent neural networks [7] to automatically detect private information. Hybrid methods combine the two approaches to compensate for their respective weaknesses [21]. No matter the method used to detect PHI, once it is detected, there are two ways to secure the data: PHI removal and PHI replacement.

Personal Health Information Removal

In PHI removal, sensitive information is located in text via a specialized search algorithm and then is either deleted or replaced with an informative deidentification tag (eg, all names are replaced with *[*NAME*]*). Although simple and common, this approach is not secure and can be easily exploited. Given that no PHI search algorithm is perfect, as a data set increases in size, it becomes increasingly certain that some PHI will be missed. Thus, if clinical notes are shared in a text format after this technique is used for deidentification, a malicious party can uncover names missed by the algorithm by manually inspecting the data. We demonstrate later that word embeddings created from such data are also vulnerable to similar exploits.

Personal Health Information Replacement

In PHI replacement, sensitive information, once located within the text by the search algorithm, is replaced with other information of the same type; for example, names can be randomly replaced with other names. This approach is more secure than PHI removal as it obscures instances where the PHI detection algorithm has failed and thus provides the data-curator with plausible deniability for any specific record.

We advocate that HCPs and data providers employ this paradigm because, if done correctly, it is much harder to exploit and thus reduces the risk to patient privacy. It is also a simple and effective way to protect against the exploitation of word embeddings that we demonstrated in this work.

Methods

Data

In our experiments, we used consultation notes. In Multimedia Appendix 1, we demonstrate how these findings are reproducible with an experiment performed with a selected subset of Wikipedia pages. We made the latter publicly available alongside the code. For all texts, we removed all punctuation and numeric characters, and we lowercased all text but performed no lemmatization, tokenization, or any other preprocessing.

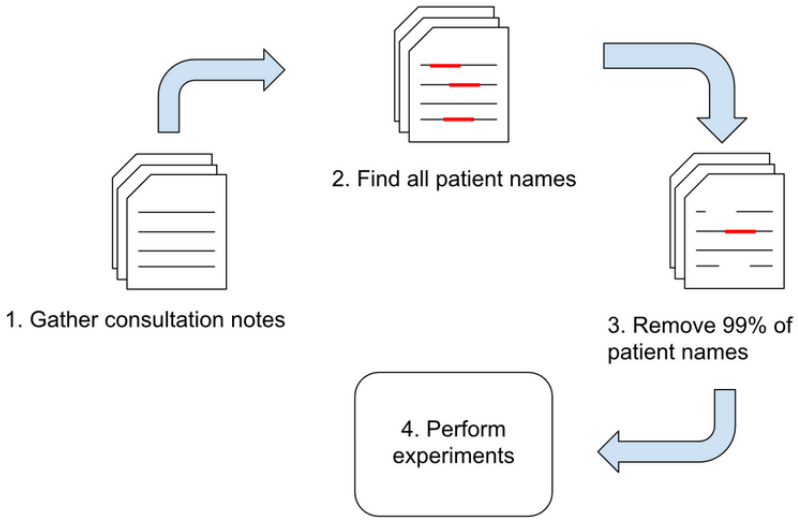
We used consultation notes provided to the authors by ICES (formerly known as the Institute for Clinical Evaluative Sciences) under data sharing agreements with physicians for the purposes of evaluation and research. Consultation notes are written by specialist physicians and other health care consultants to a patient's family physician. They describe the tests performed, results observed, and other details that the specialist physician or health care consultant considers relevant. We compiled patients' consultation notes and all their prescribed diagnostic codes that are indicative of suspected diagnoses and ordered tests, and are therefore sensitive health information that must not be connected to patient identities. The billing codes table includes text fields describing each code in 1 to 3 words, for example, *colon screening*. These data sets are linked using unique encoded identifiers and analyzed using ICES.

Although this work is conducted at ICES, ICES does not grant its research affiliates (including the authors of this paper) access to *true* patient names, but replaces them in the manner described earlier (PHI replacement), using a semimanual, dictionary-based masking process to consistently replace each true name with a randomly chosen fake name. We used heuristics to detect names in the notes. More concretely, we looked for semistructured notes that have *Name: str1, …, strN* (representing a series of alphabetical tokens separated by commas followed by a semicolon) to indicate the presence of a name. The heuristic is not 100% accurate, which is why, in Multimedia Appendix 2, we can provide only an estimate of how many true names exist by manually analyzing a randomly sampled set.

We perform our experiments on clinical consultation notes for which we can locate the associated fake patient name. For our experiments, we treat the fake names as if they were the *true* names and removed 99% of them, thus emulating current PHI removal algorithms [7]. This protected data set is then used as the first step of our experiments, as shown in Figure 1. Detailed information regarding the data is provided in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. Process flow for gathering and preparing the clinical notes for embedding generation and experimentation.



Experiments

Experimental Hypothesis

The intuition behind reidentifying patient information solely from word embeddings stems from the *distributional hypothesis* [13]—that words appearing in similar contexts tend to have similar meanings and therefore have closer vector representations than other words. Knowing this, we expect differences between both:

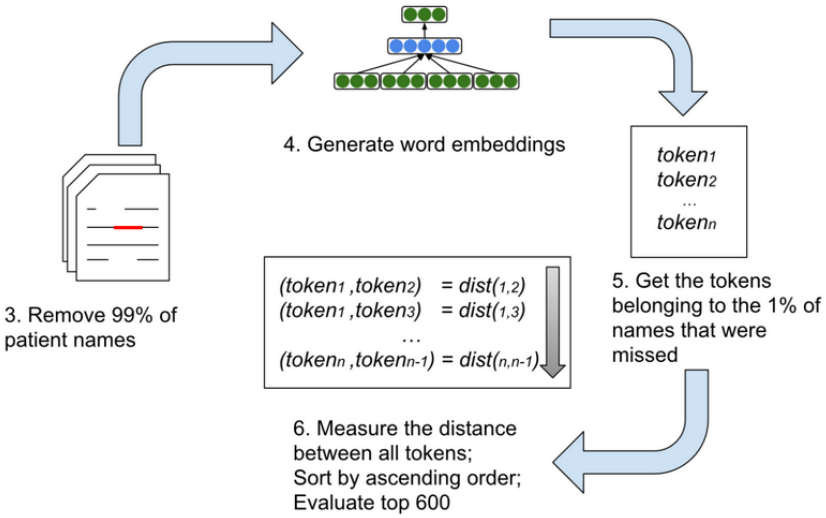
1. The average distance between the tokens that make up a person's name, compared with tokens from different names.
2. The average distance in vector space between a person's name and their diagnoses (referred to as the in-group), compared with the average distance between their name and those diagnoses with which they are not associated (referred to as the out-group.

If there is a large enough distance between a person's in-group and out-group, then this observation could be used to extract sensitive information thought to have been hidden by the unordered nature of embeddings. In the following sections, we validated this hypothesis empirically.

Experiment 1: Name Reconstruction Experiment

In the first experiment, we tested whether it is possible to reconstruct true name pairs simply from a list of individual name tokens. Figure 2 presents the steps of this experiment, picking up from the last step of Figure 1.

Figure 2. Process flow for generating word embeddings and performing the name reconstruction experiment.



A list of individual name tokens, corresponding to the fifth step in Figure 2, is easily generated by manual exploration of the words. However, as we left 1% of the names, to emulate the imperfect deidentification algorithms, we knew all the tokens (ie, the 1% of name tokens purposefully left in place).

We performed this experiment on our consultation notes data set, where over 99% of names were removed to emulate a PHI removal approach and only 1054 unique name tokens (from 650 full names) remained in the text.

We performed our experiment with 3 commonly used traditional word embedding algorithms (CBOW, Skipgram, and GloVe) for clinical prediction and modeling tasks. For each, we tested a variety of hyperparameters. Where a specific hyperparameter is not explicitly mentioned, we used the default hyperparameter

XSL•FO
RenderX

of the training model, which can be found in Multimedia Appendix 2.

However, for the sixth step, an attacker would not know how many full names were in the data set. If we assume that each name is composed of 2 tokens and none of the names share any name tokens; we would expect the number of complete names to be half the number of name tokens (ie, 1052/2 complete names). Relaxing both assumptions increases the expected names. Given name tokens *A* and *B*, we considered a name to exist if either ⟨*A*,*B*⟩ or ⟨*B*,*A*⟩ exist as names (ie, ignoring ordering). On this data set, we created many word embedding models (n=88) with a wide set of hyperparameters (ie, model specifications) that included variations in the distance metric (cosine or cityblock) and context window size.

Experiment 2: Name-Diagnostic Code Association Experiment

In this section, we explored the second part of our hypothesis: is there a difference between the average distance in vector space between a person's name and their diagnoses (their *in-group*) compared with the average distance between their name and those diagnoses with which they are *not* associated (their *out-group*)?

For this experiment, we used the same data and tested the properties of the same word embedding algorithms for various hyperparameters, as in the last experiment. We first define a patient's name vector as the average of the vectors of its components (ie, first, last, and possibly middle names). Here, *numtoken* is the number of space-separated tokens in a string and is the vector representation of the *i*th token of the name:



Second, we defined the in-group d_{in} as the set of diagnoses for *name* and the out-group d_{out}, as all other diagnoses, with d_i representing any individual diagnosis. The average distance for each of these groups from their respective names are referred to as *in_group* and *out_group*, respectively:



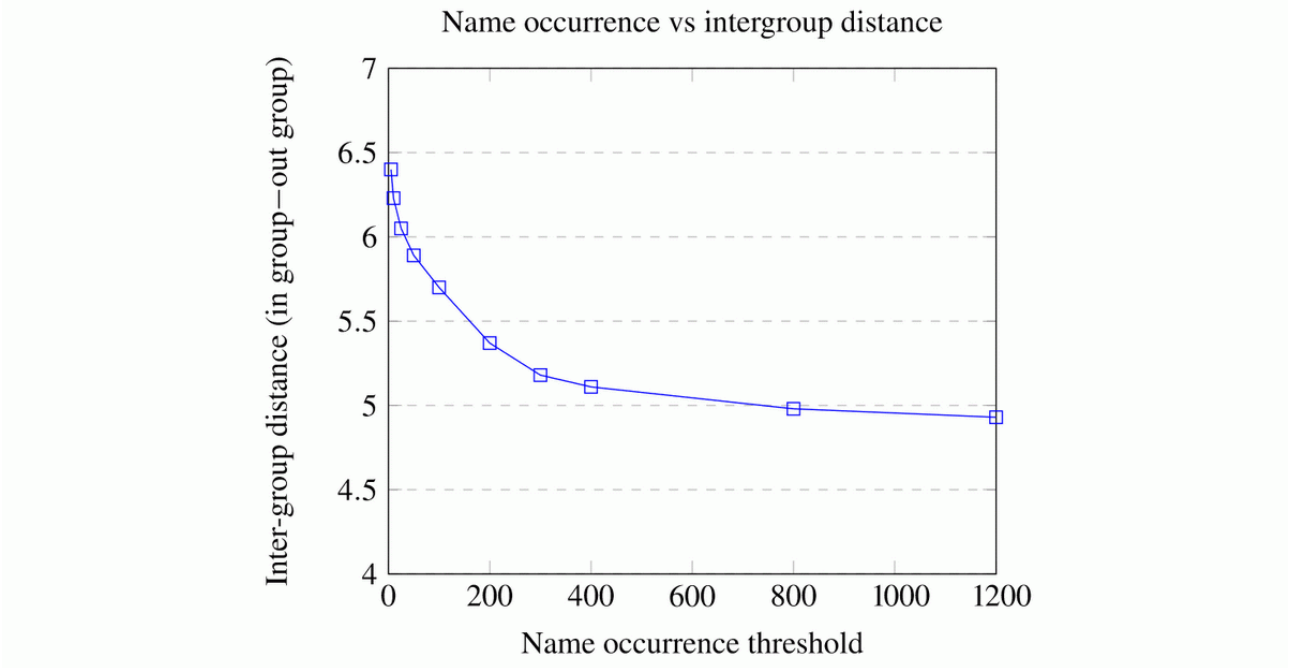


We presented the results using the *cityblock* distance (ie, the Manhattan distance) instead of the cosine distance because it performs better at this task (by uncovering more information), and past work has shown that the vector magnitude (ie, the sum of all dimensions) is affected by the number of times that the word occurs in the corpus [22]. However, our experiments were performed using the cosine distance metric as well, and complete results can be found in Multimedia Appendix 2.

Initially, we explored the raw data (ie, without any deidentification algorithm) by plotting the difference between the in- and out-groups for names that occur below different frequency thresholds. A name is below the threshold if the average counts of its components are below that threshold. For example, if “James” occurs 201 times in the corpus and “Qwerty” appears twice, then “James Qwerty” is below an arbitrary threshold of 200 (101.5<200).

Figure 3 shows that the more frequently a name occurs, the smaller the difference between the in-groups and out-groups. Nonetheless, the difference is still pronounced when all names are considered, with the lowest value being just under 5. Surprisingly, against our intuition, the in-group is larger than the out-group. We saw this result consistently throughout our testing described in the following sections.

Figure 3. Relationship between frequency of name occurrence and the average difference between the in-group and out-group for patients. This graph is generated from an experiment run on a GloVe model with a dimension of 100, window of 10, learning rate of 0.05, minimum occurrence of 1, and alpha of .75.

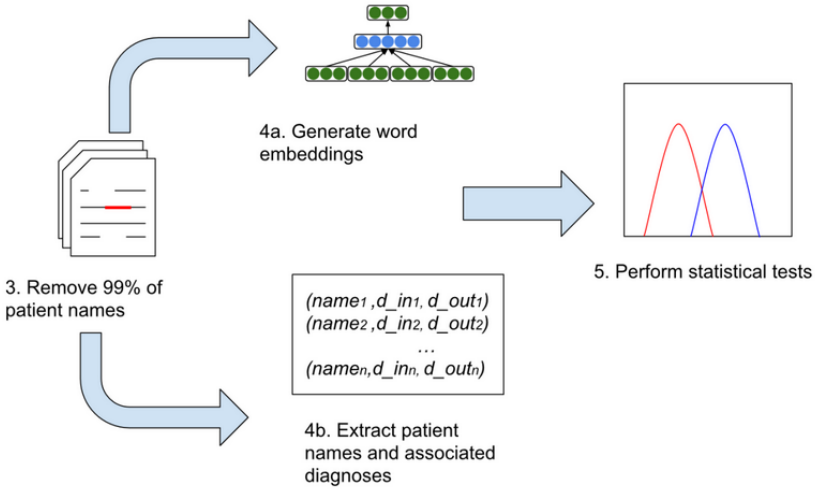


XSL•FO
RenderX

Statistical Testing

Given our initial observation that, on raw data, there is a difference between in- and out-groups on the population level on raw data, we now examine if the observed differences are statistically significant at both the population and patient levels for various embedding algorithms and hyperparameters on the deidentified data set (ie, 99% of names have been removed to emulate an optimum real-life data sharing scenario). A diagram of the experimental process is shown in Figure 4.

Figure 4. Process flow for generating word embeddings and performing statistical testing. For population-level statistical testing, we performed a Wilcoxon signed-rank test, and for patient-level statistical testing, we calculated empirical *P* values using 1000 randomly generated permutations.



Experiment 2a: Population-Level Statistical Testing

In this experiment, we aimed to determine whether the difference between the in- and out-groups on the population level is statistically significant.

Here, as with all the clinical text experiments, the embedding model is trained using all consultation notes after 99% of the names have been removed. Using the same setup as in the previous section to obtain distances between in- and out-groups, we used the Wilcoxon signed-rank test to compare the pairings of in- and out-groups for each name on the population level. The Wilcoxon signed-rank test is nonparametric and, unlike the paired Student two-tailed *t* test, makes no assumptions regarding normality.

This experiment is performed for various embedding algorithms, distance metrics, and hyperparameter ranges.

Experiment 2b: Patient-Level Statistical Testing

Here, we explored whether there is a statistically significant difference between the in- and out-groups for each patient, which would indicate that an individual patient is at risk of having their diagnostic code uncovered.

In this experiment, we compared the average difference between a patient's in-group and the out-group. Although each comparison will result in a *P* value for each patient, for brevity and privacy, we do not report the per patient analysis of the ICES data, but instead report the number of patients for which the difference is significant after correcting for multiple comparisons. To determine statistical significance at the patient level, we calculated empirical *P* values by randomly sampling in- and out-groups generated using 1000 permutations of the same size from the same data set.

We experimented with various embedding algorithms, distance metrics, and hyperparameter ranges.

Experiment 3: Scenario Simulation

In this experiment, we performed a hypothetical attack to examine whether the results of the previous 2 experiments demonstrate an actionable level of risk. Assuming the role of an attacker who has access only to released embeddings built from doctor-patient consultation notes that have been secured by using PHI removal, we showed how we are able to associate name tokens that were missed by PHI removal to arrive at a list of complete patient names and that we are able to associate these names with some target diagnoses.

For this hypothetical scenario, we used the same data and tested the properties of the same word embedding algorithms for various hyperparameters as in the last experiment.

The attack is as follows:

1. Identify a list of target diagnoses that we wish to attribute to patients. As an example, we considered the following set of diagnoses: constipation, diarrhea, vaginitis, sexual dysfunction, urinary infection, herpes genitalis, dementia, anorexia, alcoholism, threatened abortion, and AIDS.
2. For each name, calculate the 5 diagnoses that are farthest from the name.
3. Using these 5 diagnoses as the basis for prediction, we calculated Top-1 (A@1) and Top-5 (A@5) accuracy.

To ensure that our results are not an artifact of the selected diagnoses, we repeated the above experiment 1000 times for each tested hyperparameter, randomly selecting 30 target diagnoses. To be as stringent as possible, we chose from diagnoses that appeared at least 10 times in the data (which likely will result in a pessimistic bias, as demonstrated in Multimedia Appendix 2).

XSL•FO
RenderX

Results

Experiment 1: Name Reconstruction Experiment

The results of this experiment demonstrate that it is possible to reconstruct true name pairs simply from a list of individual name tokens and their respective embeddings.

In this section, we present the results for various context window sizes, an expected name list of size 600, and a cosine distance metric. We observed that up to 68.5% (411/600) of the paired tokens come from true names, as shown in Table 1 and Figure 5. As there are over 170,000 name-pair combinations, these embeddings clearly carry patient information that can be identified, thus affirming our hypothesis. The complete results for other hyperparameters, the number of names expected, and the cityblock distance metric are presented in Multimedia Appendix 2.

Table 1. The number and percentage of paired tokens that are part of true names as a function of context window size, using the cosine distance metric of the first 600 paired tokens sorted in ascending order.

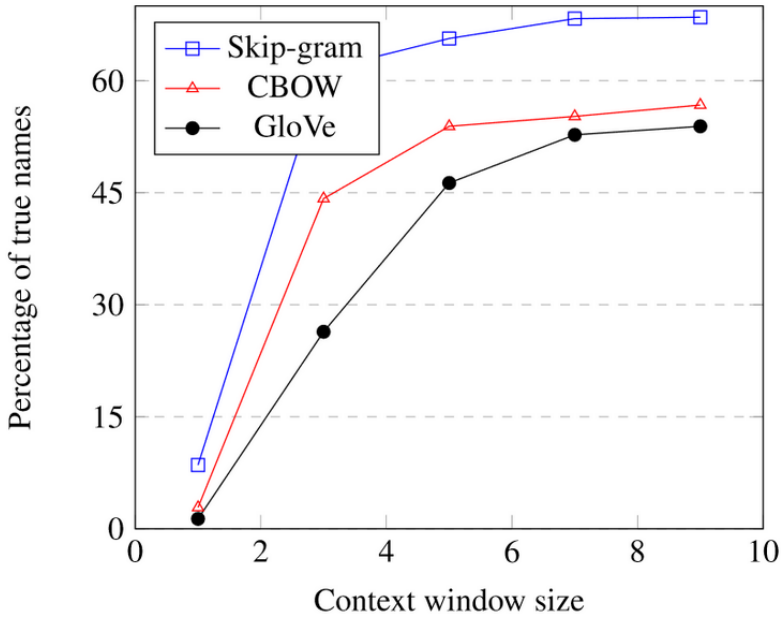
Context window size	Skipgram names, n (%)	CBOW[a] names, n (%)	GLoVe[b] names, n (%)
1	51 (8.5)	17 (2.8)	8 (1.3)[c]
3	369 (61.5)	265 (44.2)	158 (26.3)
5	393 (65.6)	323 (53.8)	278 (46.3)
7	410 (68.3)	331 (55.2)	317 (52.8)
9	411 (68.5)	340 (56.7)	323 (53.8)

[a]CBOW: Continuous Bag of Words.

[b]GLoVe: Global Vectors.

[c]Result not significant after correcting for multiple comparisons using the Holm-Bonferroni correction.

Figure 5. Visual representation of the percentage of paired names belonging to true names from the first 600 paired tokens when sorted in ascending order.



Experiment 2: Name-Diagnostic Code Association Experiment

Experiment 2a: Population-Level Statistical Testing

The results of this experiment indicate that, at the population level, the average difference between the in- and out-groups per patient is statistically significant. Table 2 and Figure 6 show the results for various embedding algorithms, varying context window sizes, and a cityblock distance metric. The complete results for other hyperparameters, other distance measures, and absolute distances are shown in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Difference between the in-group and out-group as a function of context window size for various word embedding algorithms using the cityblock distance metric. The differences are relative distances between word embedding vectors in an n-dimensional space.

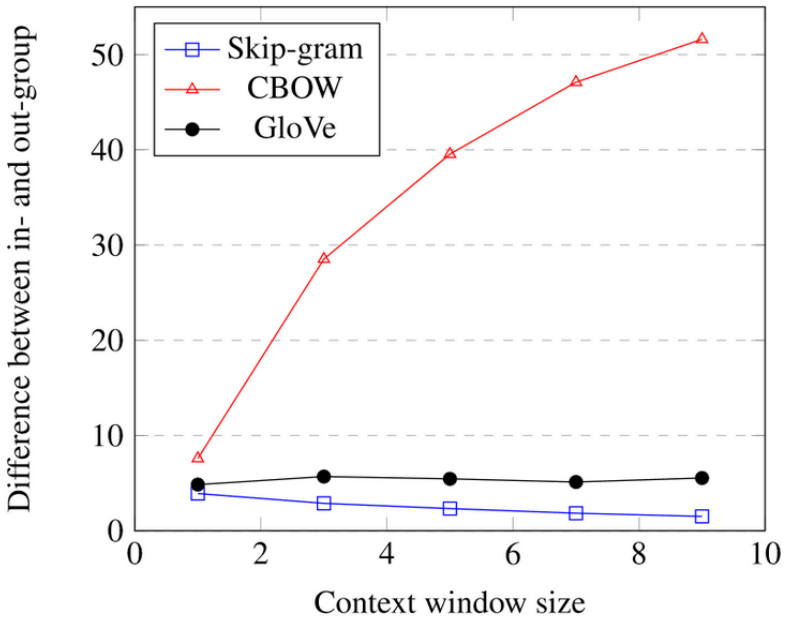
Context window size[a]	Skipgram difference	CBOW[b] difference	GLoVe[c] difference
1	3.91	7.59	4.85
3	2.88	28.53	5.69
5	2.33	39.55	5.45
7	1.84	47.10	5.12
9	1.51	51.61	5.54

[a]All differences were statistically significant after correcting for multiple comparisons.

[b]CBOW: Continuous Bag of Words.

[c]GLoVe: Global Vectors.

Figure 6. Visualization of the difference between the in-group and the out-group as a function of context window size for various word embedding algorithms using the cityblock distance metric.



Given our selected hyperparameters, we observed that for all sizes tested and for all embedding techniques, the difference between the in- and out-groups on the population level was statistically significant with $P<.001$ calculated using the Wilcoxon test, after correcting for multiple comparisons using the Holm-Bonferroni correction [23]. The Holm-Bonferroni correction is a sequentially rejective procedure for correcting multiple comparisons that keeps the family-wise type I error bounded. Figure 6 shows that the difference between the in-group and out-group decreases for embeddings created with the Skipgram algorithm as the context window increases. Conversely, the difference grows for CBOW, while it remains relatively stable for all GloVe models.

Experiment 2b: Patient-Level Statistical Testing

Building on our previous observations, the results of this experiment indicate that, at the patient level, for a percentage of examined patients (up to 449/638, 70.4%), the average difference between in- and out-groups per patient is statistically significant.

Table 3 and Figure 7 show the results for various embedding algorithms, varying context window sizes, and a cityblock distance metric. The complete results for other hyperparameters, other distance measures, and absolute distances are shown in Multimedia Appendix 2.

XSL•FO
RenderX

Table 3. The percentage of patients whose diagnoses are identifiable due to a statistically significant difference between the in-group and out-group as a function of context window size for various word embedding algorithms using the cityblock distance metric.

Size	Skipgram patients, %	CBOW[a] patients, %	GLoVe[b] patients, %
1	49 (7.7)	77 (12.1)	400 (62.7)
3	41 (6.4)	149 (23.4)	401 (62.8)
5	33 (5.2)	152 (23.8)	403 (63.2)
7	16 (2.5)	153 (24.0)	380 (59.6)
9	12 (1.9)	153 (24.0)	449 (70.4)

[a]CBOW: Continuous Bag of Words.

[b]GLoVe: Global Vectors.

Figure 7. Visualization of the percentage of patients who have a significant difference between their in- and out-groups as a function of context window size for multiple word embedding algorithms using the cityblock distance metric.

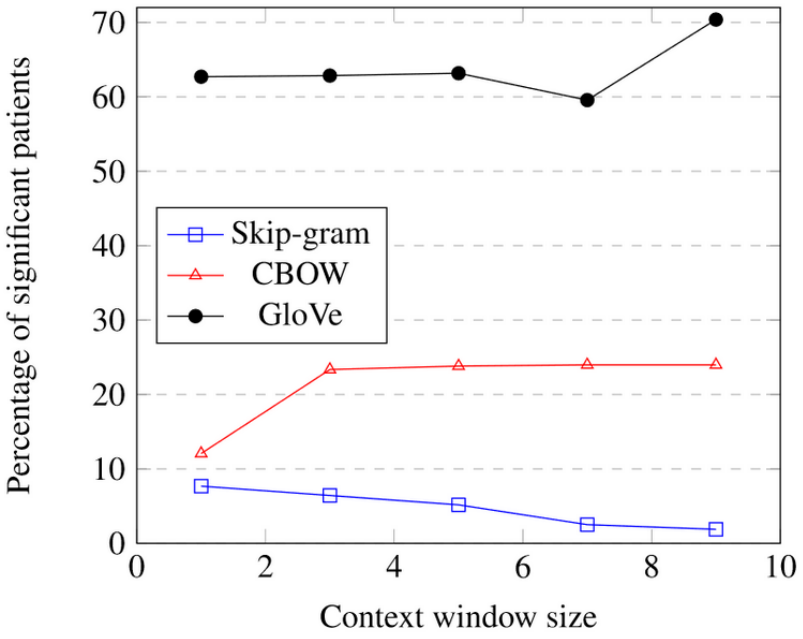


Table 3 presents the patient-level analysis for different context window sizes. As shown in Figure 7, using the CBOW algorithm, an increasing window size initially correlates positively with the number of vulnerable patients, defined as having a significant difference between the in-group and out-group. The opposite trend is observed for the Skipgram model. Context window size does not appear to have an effect on word embeddings created using GloVe, as the number of patients remains relatively stable.

Experiment 3: Scenario Simulation

Having demonstrated that the difference between in- and out-groups is statistically significant, in this section, we showed that our hypothetical attack results in an actionable level of risk, that is, an attacker who has access only to released embeddings built from doctor-patient consultation notes that have been secured by using PHI removal may be able to arrive at a list of complete patient names, and associate these names with target diagnoses.

We observed that for our chosen target diagnoses (ie, constipation, diarrhea, vaginitis, sexual dysfunction, urinary infection, herpes genitalis, dementia, anorexia, alcoholism, threatened abortion, and AIDS) our approach out performs the majority baseline for both top 1 (A@1) and top 5 (A@5) accuracy of 0.00 and 0.70, respectively, Multimedia Appendix 2 (top *n* rate is the fraction of examples for which the correct label is among the *n* labels considered most probable by the model). The complete results for all hyperparameters as well as both distance metrics are presented in Multimedia Appendix 2.

We observed similar results when the above experiment was repeated 1000 times for each tested hyperparameter, randomly selecting 30 target diagnoses. Table 4 shows how often our attacker's approach surpasses the baseline of choosing the majority diagnoses for both top-1 and top-5 accuracies. We show that we can consistently beat strong baselines, although the highest top-1 and top-5 accuracies are modest at 0.08 and 0.15, respectively. The complete results for all hyperparameters as well as both distance metrics are presented in Multimedia Appendix 2.

XSL•FO
RenderX

Table 4. The percentage of times using a word embedding–based attack beats the majority baseline for A@1 and A@5 for various context window sizes over 1000 random diagnosis selections.

Context window size[a]	Skipgram A@1, A@5	CBOW[b] A@1, A@5	GLoVe[c] A@1, A@5
1	55.8, 56.7	61.8, 61.8	55.4, 56.9
3	55.6, 53.1	51.2, 52.6	60.5, 59.5
5	57.4, 55.6	53.6, 54.5	59.4, 57.2
7	57.4, 53.5	54.6, 53.9	55.9, 54.0
9	57.2, 53.2	53.7, 51.2	60.6, 56.7

[a]We observed that the majority baseline is surpassed consistently and up to 60% of the time.

[b]CBOW: Continuous Bag of Words.

[c]GLoVe: Global Vectors.

Discussion

Principal Findings

In this work, we have shown the following:

- There is a statistically significant difference between the distance of patients' in- and out-groups at the population level.
- For many individual patients, the difference between their personal in-group and out-group is also statistically significant.
- A malicious actor working only with word embeddings may identify full names occurring in the training corpus of the embeddings as well as sensitive attributes associated with these names.

Limitations

We explored the induced privacy (or lack of privacy) of embeddings created from medical notes. We empirically highlighted the security risks of sharing clinically sourced word embeddings. Although their nature does serve to obfuscate information, we have shown that it is still possible to connect PHI to names from word embeddings secured using PHI removal. There is much variation in the risks observed in this work, which are dependent on imperfect deidentification algorithms and very skilled attackers. The actual risk to patient information, while nonzero, remains small and dependent on many variables such as the attack strategy, deidentification method, and embedding algorithm. We therefore advocate for more research to see whether the adoption of PHI replacement would better secure released embeddings. In addition to deidentification methods (where more research needs to be done), appropriate controls on who can access the anonymized data and oversight of these data are also recommended.

Conclusion

We have focused on the reidentification of names and their association with diagnostic codes, although other sensitive PHIs may also be vulnerable. We demonstrated how sharing word embeddings trained on clinical notes that have been protected using only PHI removal is not safe, as any PHI missed by the algorithm will remain in its original context. The risk of obtaining sensitive information from embeddings can be diminished by applying the anonymization methods of PHI replacement on the clinical notes before training the embedding, that is, when all known PHIs have been randomly shuffled, it becomes much more difficult (but not impossible) to determine which names in the data set belong to true patients, as the names that are shuffled together will behave in a manner similar to true names that have been missed. Such embeddings can theoretically still be at risk if an attacker is able to determine how to differentiate between fake and true names. However, this would mitigate the methods of attack described in this work, thereby making the created embeddings more secure. Alternatively, noise can be added to the generated embedding model to induce privacy and reduce risk. This risk reduction is, naturally, relative to the amount of noise added, and determining the exact amount of noise without distorting the signal or degrading performance is the subject of future research.

Regarding reassociating name tokens or associating names with diagnostic codes, Skipgram is least effective at preserving privacy, followed by CBOW and GloVe. However, when examining the number of statistically significant differences, we observe the opposite ranking. Although many sentence- and text-classification tasks observe little difference in downstream performance between these 3 algorithms, past work [5] has demonstrated differences in the ability of these algorithms to represent words. For example, Skipgram can perform better than CBOW for more frequent words [5], possibly explaining the difference in modeling names (which are infrequent in our vocabulary).

As expected, tokens from the same complete name have closer vector representations. However, despite our intuition, we find that the in-group is surprisingly larger than the out-group. That is, the average distance between a name and a diagnosis is larger if the person with that name has a diagnosis. This was consistent among all parameters tested, and among all 3 embedding models. This was also observed in our novel data set. This was perplexing because our expectation of word embeddings informs us that words that occur in similar contexts should be closer together, and in-group diagnoses are often in the same note as the name, while out-group diagnoses are not in the note at all. Even though the name and diagnosis tokens may not co-occur directly, as they would gravitate to words that co-occur with both, this would result in the names and in-group diagnoses being closer. Although a deeper theoretical investigation remains to be conducted, we hypothesize that this may be due to interaction effects within the contexts; names are quite tightly

XSL•FO
RenderX

clustered together, and they rarely occur in the same context window of the diagnosis with which they are associated. It may be that these other names draw the *common* diagnoses closer, as they occur with more names, in turn leaving the less common, but relevant, diagnoses further from the *name* cluster. This requires further research.

Acknowledgments

The authors thank the helpful staff at ICES (Dr Liisa Jaakkimainen, Dr Therese Stukel, Elisa Candido, Daniella Barron) for granting us access to data and for their insight into our manuscript. This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care. The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada to GH and FR, a Vanier Canada Graduate Scholarship to Moh A. FR is supported by a Canadian Institute for Advanced Research Chair in Artificial Intelligence.

Authors' Contributions

Moh A and Mou A designed the statistical experiments. Moh A programed the experiments. Moh A, Mou A, GH, and FR wrote the paper. Moh A and FR formulated the original problem. GH provided direction and guidance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wikipedia Appendix.
[DOCX File , 14 KB - jmir_v22i7e18055_app1.docx]

Multimedia Appendix 2
ICES Appendix.
[DOCX File , 59 KB - jmir_v22i7e18055_app2.docx]

References

1. Gehrmann S, Dernoncourt F, Li Y, Carlson ET, Wu JT, Welt J, et al. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. PLoS One 2018;13(2):e0192360 [FREE Full text] [doi: 10.1371/journal.pone.0192360] [Medline: 29447188]
2. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Nov;87:12-20 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.008] [Medline: 30217670]
3. Craig E, Arias C, Gillman D. Predicting readmission risk from doctors' notes. arXiv 2017:- epub ahead of print (1711.10663) [FREE Full text]
4. Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
5. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 2013:2265-2273 [FREE Full text]
6. Thomas SM, Mamlin B, Schadow G, McDonald C. A successful technique for removing names in pathology reports using an augmented search and replace method. Proc AMIA Symp 2002:777-781 [FREE Full text] [Medline: 12463930]
7. Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc 2017 May 1;24(3):596-606 [FREE Full text] [doi: 10.1093/jamia/ocw156] [Medline: 28040687]
8. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
9. Sahlgren M. An Introduction to Random Indexing. In: 7th International Conference on Terminology and Knowledge Engineering. 2005 Presented at: TKE'05; August 17-18, 2005; Denmark URL: http://soda.swedish-ict.se/221/1/RI_intro.pdf
10. Bengio Y, Ducharme R, Vincent P. A neural probabilistic language model. J Mach Learn Res 2003;3:1137-1155 [FREE Full text] [doi: 10.5555/944919.944966]
11. Peters M, Neumann M, Iyyer M. Deep Contextualized Word Representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,. 2018 Presented at: NAACL-HLT'18; June 1-6, 2018; New Orleans, Louisiana, USA. [doi: 10.18653/v1/n18-1202]

XSL•FO
RenderX

12. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL'19; February 15, 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/N19-1423]
13. Sahlgren M. The distributional hypothesis. Ital J Linguist 2008;20:31-51 [FREE Full text]
14. Lai S, Liu K, He S, Zhao J. How to generate a good word embedding. IEEE Intell Syst 2016 Nov;31(6):5-14. [doi: 10.1109/mis.2016.45]
15. Alsentzer E, Murphy J, Boag W, Weng WH, Jin D, Naumann T, et al. Publicly Available Clinical BERT Embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. 2019 Presented at: NLP'19; June 7, 2019; Minneapolis, Minnesota, USA p. 72-78. [doi: 10.18653/v1/w19-1909]
16. Huang K, Jaan A, Rajesh R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. arXiv 1904:- epub ahead of print(1904.05342) [FREE Full text]
17. Peng Y, Shankai Y, Zhiyong L. Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets. In: Proceedings of the 18th BioNLP Workshop and Shared Task. 2019 Presented at: BioNLP'19; August 1, 2019; Florence, Italy p. 58-65. [doi: 10.18653/v1/w19-5006]
18. Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc 2019 Nov 1;26(11):1297-1304. [doi: 10.1093/jamia/ocz096] [Medline: 31265066]
19. Miller R, Boitnott JK, Moore GW. Web-based free-text query system for surgical pathology reports with automatic case deidentification. Arch Pathol Lab Med 2001;125:1011 [FREE Full text]
20. Kushida CA, Nichols DA, Jadrnicek R, Miller R, Walsh JK, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012 Jul;50(Suppl):S82-101 [FREE Full text] [doi: 10.1097/MLR.0b013e3182585355] [Medline: 22692265]
21. Neamatullah I, Douglass MM, Lehman LH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008 Jul 24;8:32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]
22. Schakel A, Wilson B. Measuring word significance using distributed representations of words. arXiv 2015:- epub ahead of print (1508.02297) [FREE Full text]
23. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979:65-70 [FREE Full text]

Abbreviations

CBOW: Continuous Bag of Words
GLoVe: Global Vectors
HCP: health care provider
HIPAA: Health Insurance Portability and Accountability Act
MIMIC: Multiparameter Intelligence Monitoring in Intensive Care
NLP: natural language processing
PHI: personal health information

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by S Choemprayong, SY Shin; comments to author 09.04.20; revised version received 21.04.20; accepted 27.04.20; published 15.07.20.

Please cite as:
Abdalla M, Abdalla M, Hirst G, Rudzicz F
Exploring the Privacy-Preserving Properties of Word Embeddings: Algorithmic Validation Study
J Med Internet Res 2020;22(7):e18055
URL: https://www.jmir.org/2020/7/e18055
doi:10.2196/18055
PMID:32673230

©Mohamed Abdalla, Moustafa Abdalla, Graeme Hirst, Frank Rudzicz. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Surveillance of Noncommunicable Disease Epidemic Through the Integrated Noncommunicable Disease Collaborative Management System: Feasibility Pilot Study Conducted in the City of Ningbo, China

Sixuan Li[1*], MPH; Liang Zhang[2*], BSc; Shiwei Liu[3], PhD; Richard Hubbard[4], DM; Hui Li[1], MPH

[1]Division of Chronic and Noncommunicable Disease Control and Prevention, Ningbo Municipal Center for Disease Control and Prevention, Ningbo, China

[2]Division of Big Data, Ningbo Municipal Center for Disease Control and Prevention, Ningbo, China

[3]National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

[4]Division of Epidemiology and Public Health, University of Nottingham, Nottingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Hui Li, MPH
Division of Chronic and Noncommunicable Disease Control and Prevention
Ningbo Municipal Center for Disease Control and Prevention
237 Yongfeng Road
Haishu District
Ningbo, 315010
China
Phone: 86 0574 87274542
Email: lihui4329@163.com

Abstract

Background: Noncommunicable diseases (NCDs) have become the main public health concern worldwide. With rapid economic development and changes in lifestyles, the burden of NCDs in China is increasing dramatically every year. Monitoring is a critical measure for NCDs control and prevention. However, because of the lack of regional representativeness, unsatisfactory data quality, and inefficient data sharing and utilization, the existing surveillance systems and surveys in China cannot track the status and transition of NCDs epidemic.

Objective: To efficaciously track NCDs epidemic in China, this pilot program conducted in Ningbo city by the Chinese Center for Disease Control and Prevention (CDC) aimed to develop an innovative model for NCDs surveillance and management: the integrated noncommunicable disease collaborative management system (NCDCMS).

Methods: This Ningbo model was designed and developed through a 3-level (county/district, municipal, and provincial levels) direct reporting system based on the regional health information platform. The uniform data standards and interface specifications were established to connect different platforms and conduct data exchanges. The performance of the system was evaluated based on the 9 attributes of surveillance system evaluation framework recommended by the US CDC.

Results: NCDCMS allows automatic NCDs data exchanging and sharing via a 3-level public health data exchange platform in China. It currently covers 201 medical institutions throughout the city. Compared with previous systems, automatic popping up of the report card, automatic patient information extraction, and real-time data exchange process have highly improved the simplicity and timeliness of the system. The data quality meets the requirements to monitor the incidence trend of NCDs accurately, and the comprehensive data types obtained from the database (ie, directly from the 3-level platform on the data warehouse) also provide a useful information to conduct scientific studies. So far, 98.1% (201/205) of medical institutions across Ningbo having been involved in data exchanges with the model. Evaluations of the system performance showed that NCDCMS has high levels of simplicity, data quality, acceptability, representativeness, and timeliness.

Conclusions: NCDCMS completely reshaped the process of NCD surveillance reporting and had unique advantages, which include reducing the work burden of different stakeholders by data sharing and exchange, eliminating unnecessary redundancies,

XSL•FO
RenderX

reducing the amount of underreporting, and structuring population-based cohorts. The Ningbo model will be gradually promoted elsewhere following this success of the pilot project, and is expected to be a milestone in NCDs surveillance, control, and prevention in China.

(J Med Internet Res 2020;22(7):e17340) doi:10.2196/17340

KEYWORDS

surveillance; noncommunicable diseases; regional health information platform; electronic health records.

Introduction

With the rapid socioeconomic development and acceleration of population aging and urbanization, noncommunicable diseases (NCDs), such as cardiovascular diseases, cancers, chronic obstructive pulmonary diseases (COPDs), and diabetes, have become the main public health concerns worldwide. The WHO reported that NCDs kill around 40 million people every year, equivalent to 70% of all deaths globally [1]. China, as a developing country with the largest population, has been experiencing incremental prevalence of NCDs over the past decades; the proportional burden of NCDs increased by 37.27% from 60.21% in 1990 to 82.65% in 2016 [2]. Studies have shown that the prevalence of diabetes in adults was 9.7% in 2013, COPD among people aged 40 and above was 13.6% in 2015, and hypertension in adults aged 35-75 years was 44.7% in 2014-2017 in China [3-5]. Moreover, cardiovascular diseases, cancer, and chronic respiratory disease were the leading causes of death, accounting for approximately 79.4% of total deaths [6].

Monitoring is a critical and indispensable measure for NCDs control and prevention, including providing evidence for policy making and data support for evaluation of the global and regional targets [7]. As a cornerstone of public health practice in improving disease surveillance, the US Centers for Disease Control and Prevention (CDC) launched the CDC Surveillance Strategy in 2014 to achieve the targets of advancing the use of electronic health records (EHRs), retiring redundant systems, and maximizing performance of agency resources [8]. In China, the Chronic Disease and Risk Factor Surveillance Survey was introduced in 2004 using a multistage stratified cluster random sampling method to monitor NCD epidemics nationwide every 3 years [5,9-11]. However, due to the collection of NCD information by self-reports and the heavy workload involved in field survey, the monitoring data can only be obtained through selected surveillance points and their quality varies among regions. Several surveys and registries on specific chronic diseases are available in China, including those on stroke, hypertension, COPD, diabetes, and cancer, which provide valuable national epidemiological data on NCDs [3-5,12,13]. However, because these surveys and registries are independent from each other with inconsistent research methods and different population coverage, it is difficult to compare the varied results and obtain a whole picture of current situation of NCDs in China, which also reflects the shortcomings of traditional surveillance system in data sharing and utilization [14]. A population-based data with high quality generated from a reliable surveillance system is crucial to detect the priorities to track and tackle NCDs and find the cost-effective interventions to reduce the burden of NCDs.

Ningbo is a prefecture-level city of Zhejiang province in eastern China, consists of 10 districts with a registered population of 5.93 million, and has the very advanced regional health information platform in China. In 2016, Ningbo successfully achieved the 3-level (county/district, municipal, and provincial level) direct reporting of infectious diseases through a data exchange platform, which shortened the reporting time to 2 minutes from uploading a report card in the hospital information system (HIS) to the national information system [15]. Two years later, a similar but improved reporting and management model for NCDs was successfully implemented, named integrated noncommunicable disease collaborative management system (NCDCMS). This pilot study was conducted in Ningbo by the Chinese CDC and aimed to develop an innovative model for NCDs surveillance and management in order to efficaciously tackle NCDs epidemic in China.

Methods

Ningbo Model

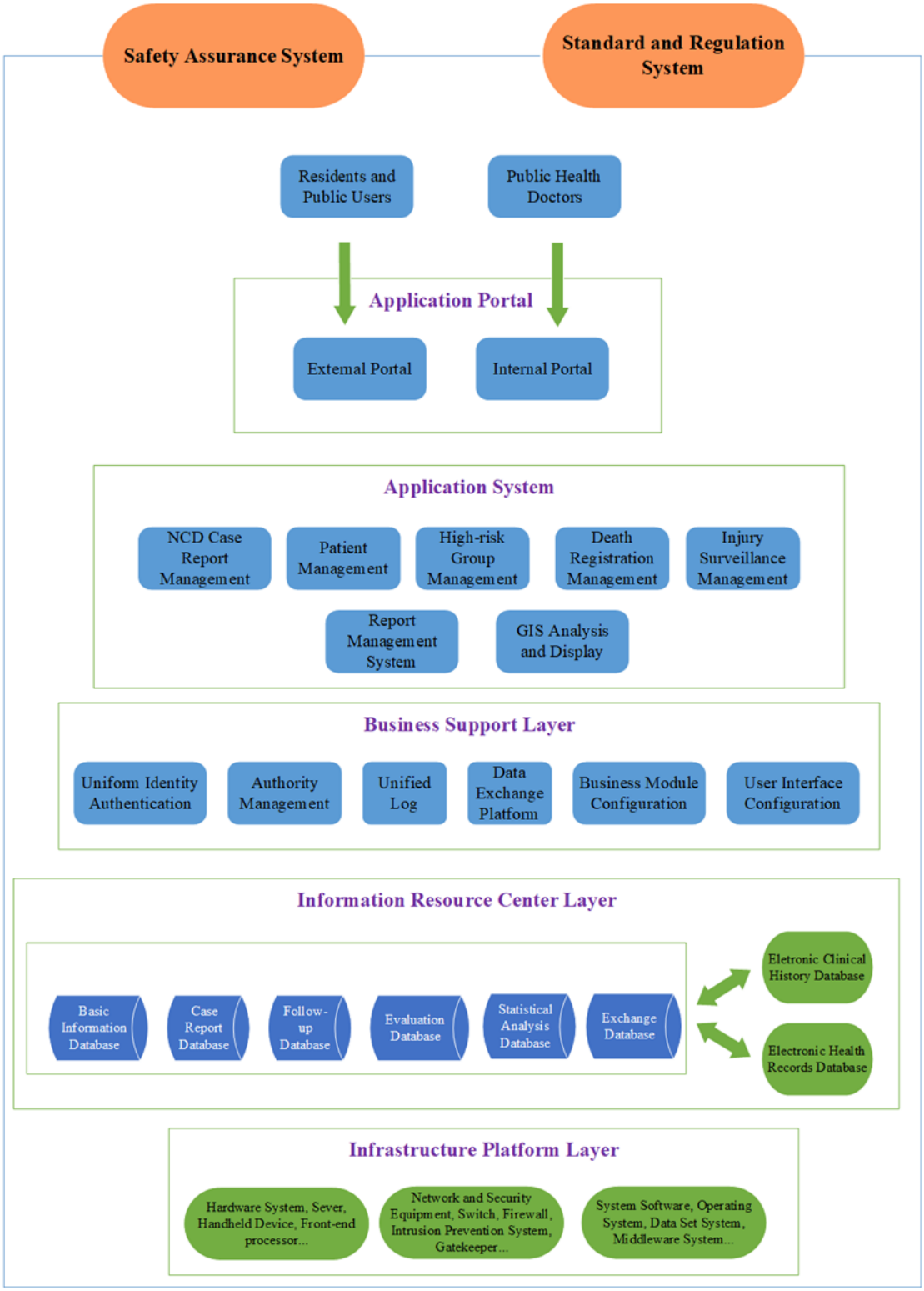
The Ningbo model was constructed through a 3-level direct reporting strategy based on the regional health information platform. The uniform data standards and interface specifications were developed to connect different platforms and conduct data exchanges. NCDs registration report and management standards were issued to clarify data quality evaluation indicators and data security management standards. 9 attributes of the surveillance system evaluation framework recommended by the US CDC were used for evaluating the performance of NCDCMS [16,17].

Architecture of the NCDCMS

With full consideration of the current situation of Ningbo and the potential requirements of system development, the SOA (service-oriented architecture) and ESB (enterprise service bus) technologies were used for building a multilayer architecture system based on the concept of integrated information system design. The architecture of NCDCMS is made up of 7 components: the infrastructure platform layer, the NCDs information resource center, the business support layer, the application system, the application portal, the safety assurance system, and the standard and regulation system. Figure 1 shows the architecture of the overall system.

XSL•FO
RenderX

Figure 1. Architecture of the integrated noncommunicable disease collaborative management system. GIS: geographic information system;



Process of Model Construction

To build an effective and reliable NCDs surveillance and management system, the NCDs 3-level direct reporting model was constructed by following these 4 aspects: (1) We completed the construction of the 3-level public health data exchange platform by connecting Ningbo Health Network (municipal level) with Zhejiang Provincial Health Network (province level) and County/District Health Network (county/district level). Health Network is the special information network for health industry in China. On this basis, the Ningbo public health data exchange platform was organized to achieve the data exchange from the county/district platform to the municipal platform, and further to the provincial platform. (2) We developed uniform data standards and interface specifications. The interface specifications were used as regulations and standards to connect different platforms and conduct data collection, exchange, inquiry, and reconciliation. (3) To connect HIS with the data

XSL•FO
RenderX

exchange platform, all medical institutions were required to adjust their public health module in accordance with the uniform data standards and the interface specifications. After the adjustment, automatic popping up of report card, automatic extraction of patient information, and real-time data exchange among platforms can be achieved. (4) We finally issued the Ningbo NCDs Registration Report and Management Standards. It elaborated the standardized workflow and mechanisms for interdepartmental collaborations between different stakeholders. The indicators of data quality evaluation and data security management were also included and elaborated in detail in these standards.

Evaluations of the System Performance

Because NCDCMS is a public health surveillance system, its performance should be evaluated to ensure the public health problems are being monitored efficiently and effectively. We evaluated the performance of the NCDCMS with the 9 surveillance system attributes based on the guidelines for evaluating public health surveillance systems developed by the US CDC: simplicity, flexibility, data quality, acceptability, sensitivity, predictive value positive, representativeness, timeliness, and stability [16]. As public health surveillance systems vary in objective, methods, purpose, and scope, attributes that are important in one system might be less important in another. Taking into account the purpose and characteristics of NCDCMS, we selected 6 attributes to evaluate the performance of NCDCMS. The definitions of these 6 attributes are presented in Table 1 [17].

Table 1. Definitions of surveillance system attributes.

Attributes	Definition
Simplicity	System's structure and ease of operation
Flexibility	Ability to adapt to changing information needs or technological operating conditions with little additional time, personnel, or allocated funds
Data quality	Completeness and validity of the data recorded in the system
Acceptability	Willingness of persons and organizations to participate in the system
Representativeness	Ability to accurately describe the occurrence of a health-related event over time and its distribution in the population by place and person
Timeliness	Speed between steps in a system

Ethical Considerations

Because the construction of NCDCMS is also a component of the Ningbo Smart City Project, it is a governmental initiative rather than a study. The ethical issues need to be addressed but could be different from general scientific studies. In the process of system construction, the protection of residents' personal privacy has been fully taken into account. Moreover, the regional health information platform of Ningbo is nationally certified and has reached the maturity level of IV, which is considered the highest level of regional information protection. Thus, it is legal and safe to collect residents' data through the system.

Ethics Approval

The study was approved by the Ethics Review Committee of Ningbo Municipal Centers for Disease Control and Prevention.

Availability of Data and Material

The data sets generated/analyzed during this study are not publicly available due to the protection of residents' personal privacy but are available from the corresponding author on reasonable request.

Results

Data Collection and Exchange

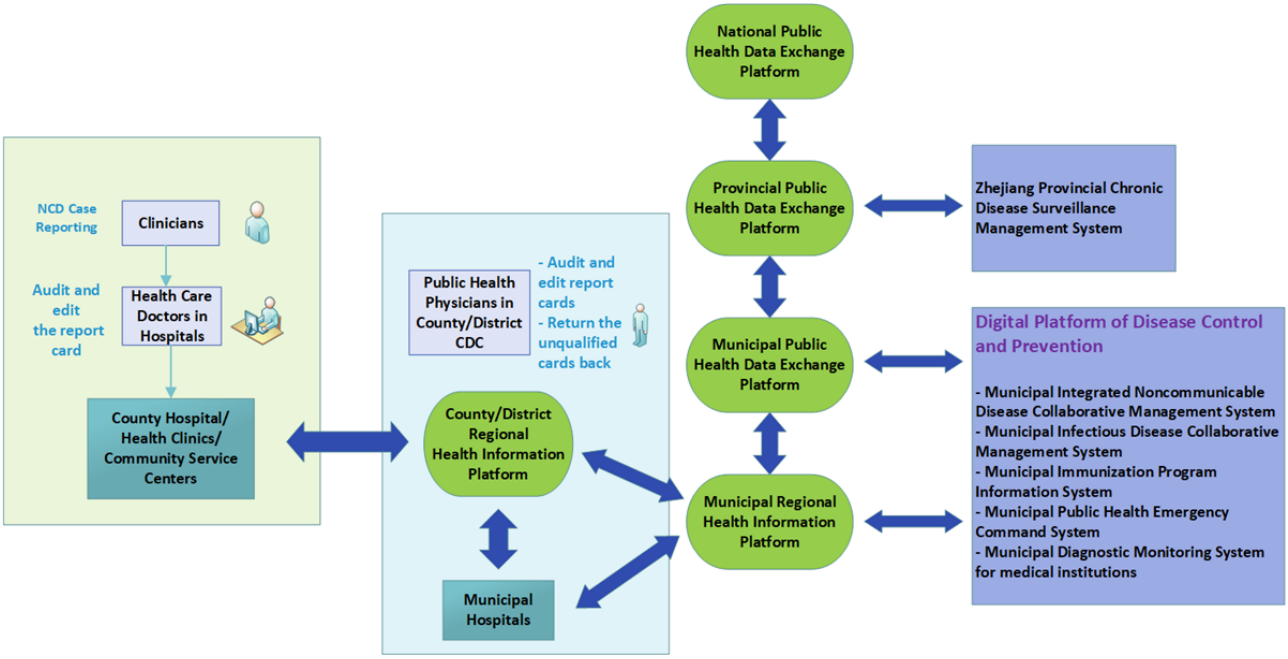
In NCDCMS, 4 types of notifiable NCDs (diabetes mellitus, ischemic heart disease and cardiac arrest, cerebrovascular disease, malignant neoplasms, and benign neoplasm of central nervous system) report cards are collected through a stepwise 3-level public health data exchange platform. The types of NCDs and the corresponding ICD-10 codes are listed in Table 2. Besides, a vital registration system is integrated into NCDCMS, which is able to collect cause-of-death data accurately and efficiently, from which we can obtain insights into the underlying causes of death.

When a clinician diagnoses a notifiable chronic disease on his/her clinic system, a disease reporting card pops up automatically that presents data about that specific patient, which are extracted from the HIS or EHR. Clinicians only need to type in their diagnosis and treatment information to complete the whole report card, which takes only a few minutes. Health care doctors working in hospitals will then review the card to ensure the accuracy and logicality of the information. Besides, public health physicians who work in county/district CDC will double check the information afterward. If the content does not meet the standard requirements, the card will be returned to the health care doctors along with the feedback so as to be further edited. Once this has been done, the data will be transmitted to the regional health information platform immediately. Through a stepwise data exchange process, report cards will be transmitted to the Zhejiang Provincial Chronic Disease Surveillance Management System, and finally to the national system. Furthermore, these data will be exchanged with the Ningbo Digital Platform of Disease Control System and be shared across branch systems as shown in Figure 2. The process of data exchange and data sharing through a 3-level public health data exchange platform is presented in Figure 2. The data type collected on report card of notifiable NCDs is shown in Table

XSL•FO
RenderX

3. Data collected on regional health information platform include residents' personal information, demographic characteristics, outpatient information, hospitalization information, results of examinations and laboratory tests, prescription lists, and immunization information. For municipal hospitals, the NCD cards reporting process is consistent with the reporting process of county hospital/health clinics/community service centers depicted in Figure 2.

Figure 2. The process of data exchange and data sharing through a 3-level (county, district, and municipal provincial level) public health data exchange platform.



Importantly, the purpose of building NCDCMS is not only to monitor the incidence of NCD but also to achieve better health management of patients with NCDs. Data exchange among platforms is not unidirectional, but rather a bidirectional data exchange process. When the NCD data are transmitted to the upper-level platform, cause of death and other branch surveillance system data (eg, infectious diseases, immunization records) of that particular patient will eventually be integrated into the EHR and transmitted back to the county/district regional information platform in order to achieve a two-way data exchange based on the principle of 1 person 1 file. Even if the NCD card is reported outside the place of residence, as long as it is reported within the Zhejiang Province, it will be returned to the place of residence following a bidirectional data exchange process.

Based on the stepwise 3-level public health data exchange platform, routine monitoring of 4 types of notifiable NCDs can be achieved on NCDCMS. Tables and charts can be automatically generated from surveillance data to track changes as needed.

Table 2. Four types of notifiable NCDs and their respective ICD-10 code(s).

Disease	ICD-10 code
Diabetes mellitus	E10-E14, O24
Ischemic heart disease and cardiac arrest	I20-I25, I46
Cerebrovascular disease	I60-I61, I63-I64
Malignant neoplasms	C00-C96
Benign neoplasm of the central nervous system	D32-D33, D42-D43

XSL•FO
RenderX

Table 3. Data types collected on report card of notifiable NCDs.

Diabetes mellitus	Ischemic heart disease and cardiac arrest	Cerebrovascular disease	Malignant neoplasms and benign neoplasm of the central nervous system
Electronic health record number	Electronic health record number	Electronic health record number	Electronic health record number
Patient name	Patient name	Patient name	Patient name
Gender	Gender	Gender	Gender
Nationality	Nationality	Nationality	Nationality
Education level	Education level	Education level	Education level
Identity card number	Identity card number	Identity card number	Identity card number
Date of birth	Date of birth	Date of birth	Date of birth
Occupation	Occupation	Occupation	Occupation
Contact number	Contact number	Contact number	Contact number
Marital status	Marital status	Marital status	Marital status
Age	Age	Age	Age
Registered address	Registered address	Registered address	Registered address
Current address	Current address	Current address	Current address
Outpatient number	Outpatient number	Outpatient number	Outpatient number
Hospitalization number	Hospitalization number	Hospitalization number	Hospitalization number
Report institution	Report institution	Report institution	Report institution
Report doctor name	Report doctor name	Report doctor name	Report doctor name
Date of report	Date of report	Date of report	Date of report
ICD-10 code of disease	ICD-10 code of disease	ICD-10 code of disease	ICD-10 code of disease
Disease type	Disease type	Disease type	Disease type
Date of diagnosis	Date of diagnosis	Date of diagnosis	Date of diagnosis
__[a]	Date of onset	Date of onset	__[a]
Symptoms	Symptoms	Symptoms	Symptoms
Examinations and tests	Examinations and tests	Examinations and tests	Examinations and tests
__[a]	__[a]	__[a]	ICD-O-3
__[a]	__[a]	__[a]	Pathologic type
__[a]	__[a]	__[a]	TMN stage
Complications	__[a]	__[a]	__[a]
Risk factors	__[a]	__[a]	__[a]
Family history	__[a]	__[a]	__[a]
__[a]	Medical history	Medical history	__[a]
Underlying cause of death	Underlying cause of death	Underlying cause of death	Underlying cause of death
Date of death	Date of death	Date of death	Date of death
ICD-10 code of death	ICD-10 code of death	ICD-10 code of death	ICD-10 code of death

[a]Not applicable.

Evaluations of System Performance

Simplicity

Previously, during the data reporting process, clinicians first filled in NCD cards (either paper card or electronic card in HIS), and then health care doctors in hospitals manually entered the data into the Zhejiang Provincial Chronic Disease Surveillance Management System. Clinicians using electronic cards reported each patient information variable one by one (ie, on separate fields) into the HIS. It is well-known that manual data entry

XSL•FO
RenderX

consumes a huge amount of human resources. Therefore, the NCDCMS was developed to offset the inefficiency of manual entries. Now, in the current system, when a clinician diagnoses a notifiable chronic disease on his/her clinic system, an NCD reporting card pops up automatically that presents much of the data about that specific patient, which are extracted from the HIS or EHR. Clinicians only need to type in their diagnosis and treatment information to complete the whole report card, which takes only a few minutes. Public health doctors in hospitals do not need to manually enter the reporting data in the surveillance system; instead, they only need to audit the report card and click the upload button, which transmits the data to the regional health information platform.

Meanwhile, because the NCDCMS realized the data collection process through a stepwise 3-level public health data exchange platform, which combined various independent platforms into one interconnected model, once a case has been reported, the data can be used in multiple ways and serve a variety of purposes. The data need to be reported by the clinicians only once and after the process is completed, the reported data can be transmitted directly to each platform in real time. This allows to eliminate duplicate entries.

Flexibility

To connect HIS with the data exchange platform, all medical institutions were required to adjust their public health module in HIS in accordance with the uniform data standards and the interface specifications; importantly, this adjustment is neither time consuming nor costly. In addition, the establishment of NCDCMS has a solid preliminary informatization foundation: (1) Ningbo is a city with a high level of informatization. (2) The 3-level direct reporting of notifiable infectious diseases by the 3-level public health data exchange platform was successfully achieved in 2016. (3) The key component of the system was structured by one company, which also increased the feasibility of the system to adapt to various requirements afterward. Because the adjustment of the public health module in HIS is neither time-consuming nor costly, with the solid informatization foundation of the whole city, the flexibility of the system can be guaranteed.

Data Quality

The quality of data collected by NCDCMS can be guaranteed in the following 4 aspects: (1) Each level of platform has a built-in logic check system and each type of NCD report card has several required items to be filled in. If the required items are not filled in completely or the contents of the report data do not meet the requirements of logical check, the report card information cannot be uploaded to the upper-level platform. (2) As mentioned above, a multilevel audit for all data reported will allow the data entered to be reviewed by health care doctors and public health physicians, which further ensures the accuracy of the collected data. (3) As the data platforms are directly related to each other, patient's EHR file can be accessed during the case review process. In this way, data errors made by clinicians can be detected on time and accuracy of the data can be improved. (4) We perform data cleaning regularly and issue data analysis reports on a monthly, quarterly, and yearly basis to ensure that data quality meets the requirements. In addition, to ensure the authenticity of the reported data and have a continuous understanding of patients' health status, a first visit within 1 month of data entry and a yearly follow-up of each NCD patient are conducted.

Acceptability

In the initial stages of the system construction, we selected 2 counties and 1 municipal-level hospital for model pilot tests. These early trials provided us with a great deal of experience in system construction. One year after the completion of the pilot tests, the system was gradually promoted and expanded to the whole city. This progressive approach of system construction also increased the acceptability of NCDCMS.

The construction of NCDCMS was government managed. The Health Commission of Ningbo explicitly demands every medical institution in Ningbo to complete the public health module adjustment to achieve the 3-level direct reporting of notifiable NCD diseases. Meanwhile, the completement of adjustment was considered as an indicator in the annual performance assessment of medical institutions. Therefore, China's national conditions determine that this model can be widely accepted among medical institutions of Ningbo. At the time of this writing, among the 205 medical institutions that are required to report notifiable chronic diseases in Ningbo, 201 have completed the module adjustment, with the completion rate of 98.1% (201/205).

Representativeness

NCDCMS is totally based on the case reporting of medical institutions and has a facility coverage rate of 98.1% (201/205), meaning that as long as patients with NCDs visit medical institutions, either in hospital or in community, they will be reported by clinicians. Therefore, the NCD cases we collected via NCDCMS can reflect the actual situation of the whole population in Ningbo over time. In addition, based on the data collected by the system, the occurrence of 4 types of notifiable NCDs over time and their distribution in the population by place and person could be accurately described. Because of the continuous monitoring, the incidence trend of NCDs could be displayed and compared.

In addition to the NCDs data collected from NCDCMS, with the patient's identity card number as a unique data matching criterion, more information can be drawn from other sources such as HIS and HER, which allows to build a more comprehensive database at the back end. Technically, from this comprehensive database, each patient's data including his/her personal information, demographic characteristics, hospitalization information, results of examinations and tests, prescriptions, and survival situations could be obtained. Furthermore, branch surveillance systems are linked by the Ningbo Digital Platform of Disease Control and Prevention, and hence the data from these branch systems can also be shared and used for multiple purposes. From this perspective, using the data collected from Ningbo NCDCMS, the priorities of public health issues can be identified and the burden of disease can be measured in an accurate and convenient approach.

XSL•FO
RenderX

Timeliness

In the past, if a clinician reported an NCD card, irrespective of whether using a paper card or an electronic card, health care doctors working in hospitals had to manually enter report data variable one by one (ie, separately on each field) into the provincial chronic disease surveillance management system within 7 days. Although these cards are not immediately audited by county/district CDC staff, according to the registration and validation procedure established by provincial CDC, the staff are required to audit these cards within 7 days after data entry. Thus, this process, from initial data reporting on cards to the final completion of review, usually took several days, with a maximum duration of up to 14 days. However, because of escalating burden of NCDs in China, this procedure consumed a great deal of human resources, time, and costs.

Following the implementation of the new model, health care doctors' working hours have been greatly reduced because of the removal of duplicate entries. Thus, more of their working time could be devoted to controlling the overall quality of the report card and training and improving the clinicians' reporting skills. Thanks to the automatic popping up of the report card and information extraction approach, clinicians could fill in the report card when patients visit their clinic, or during the patient's hospitalization, allowing simpler and faster completion. In addition, with the great improvement in the overall effectiveness of the new model, staff members from district/county CDC will review report cards on a daily basis. This new process greatly reduces the total time from reporting to reviewing of NCD cards, from the previous maximum of 14 days to the current minimum of 1 day. Compared with the past, the average card entry time is reduced from 10 minutes to 1 minute. Besides, the time taken for clinicians to enter relevant data at first visit and follow-up is reduced from 15 minutes to less than 1 minute, respectively. Excluding the time taken for manual auditing and reviewing, data are exchanged in real time between different platforms.

Data Utilization

Using the data collected from NCDCMS and EHR, one study was conducted to describe the epidemiological characteristics of patients with diabetes mellitus (DM) diagnosed with oculopathy in one district in Ningbo. The diagnosed DM cases generated from NCDCMS were matched with the oculopathy cases generated from the regional health information platform by ID number and EHR number. The results showed that 1819 patients with DM were diagnosed with oculopathy in 2015. Among them, 195 patients with DM were newly diagnosed. The oculopathy comorbidity rate of the patients with DM and those newly diagnosed with DM in 2015 were 12.78% and 19.25%, respectively. The time intervals of patients with type 1 DM and type 2 DM from being diagnosed with diabetes to comorbidity of oculopathy was 4.00 and 3.00 years, respectively [18]. By linking diabetes cases and cancer cases generated from NCDCMS, one study implied that the patients with newly diagnosed type 2 DM, who were taller and had normal BMI, had a higher risk of cancer. Compared with their overweight (standardized incidence ratio 0.62) or obese (standardized incidence ratio 0.35) counterparts, the all-cause standardized incidence ratios of the normal BMI group was 1.13. Patients with normal BMI had a high risk of liver, pancreatic, esophagus, and ovary cancers [19].

Instead of field surveys, which consumed a large amount of human resource for data collection and analysis, the ecological studies and pilot epidemiological studies can be conducted directly through this 3-level public health data exchange platform. Moreover, we have started collecting NCD surveillance data since 2009 and each NCD case is required to be followed up once per year. These data will therefore contribute to carrying out different kinds of population-based cohort studies. Different variables needed for building cohorts can be obtained from multiple sources as mentioned earlier. Meanwhile, the information on all-cause mortality is also being collected, meaning that we can obtain data on the reliable survival situation of each patient with NCD. In one district of Ningbo, a cohort based on regional health information platform has been built to predict the risk factors of cardiovascular diseases [20]. Thus, big data tools can be applied for analyzing different data collected from the 3-level exchange platforms, with the results likely to be useful for decision making on various policies and strategies related to NCDs prevention and treatments.

Discussion

To the best of our knowledge, NCDCMS is the first platform in China that achieved the 3-level direct reporting model based on regional health information platform. NCDCMS currently covers 201 medical institutions throughout the city, including municipal hospitals, county hospital, health clinics, and community service centers. The evaluations of the system performance showed that NCDCMS has high levels of simplicity, data quality, acceptability, representativeness, and timeliness. This 3-level direct reporting model helps relieve the burden of case reporting, as it allows real-time reporting, eliminates unnecessary redundancies, reduces the amount of underreporting, and provides a better database of chronic diseases.

The Ningbo Model is an innovative practice that aimed to develop a reliable population-based surveillance system via a real-time bidirectional 3-level data exchange process. It completely reshaped the traditional process of NCD surveillance data reporting and had three unique advantages: (1) It reduces the work burden of different stakeholders by data sharing and exchange. Previously, NCD reporting was completely dependent on manual data entry, which cannot guarantee the accuracy and was time-consuming. Now, clinicians could complete the case reporting during patient's visit; thus, both the patient's visit time and the work burden of clinicians are reduced due to the reduction of duplicated entries and repeated inquires. Besides, the workload of health care doctors who were responsible for checking the report card submitted by clinicians is reduced. (2) The amount of underreporting was greatly reduced through the automatic popping up of the report card. Some hospitals also link NCD reports with the discharge summary submission procedure. If the NCD cases are not reported as required, clinicians cannot eventually transfer the medical records of discharged patients to the medical record room. (3) The system

XSL•FO
RenderX

contains better databases that allow data sharing among different platforms. A fairly complete personal health file can be obtained at the back end by linking different sources mentioned above. On this basis, the follow-up data, mortality data, and EHR data can be built into different kinds of NCD cohorts and the risk prediction models can be established. Different types of risk factors can be addressed and evidence-based interventions can also be performed. These research results can help policy makers to make decisions in an evidence-based way.

In the 2 years of NCDCMS construction, one difficulty was how to convince all medical institutions to use the uniform data standards and interface specifications. Because different medical institutions used different HISs developed by various companies, coordination among multiple stakeholders was initially expected to be difficult. In this aspect, the Health Commission of Ningbo and Ningbo CDC played important roles in facilitating the construction of the whole model. However, in order to obtain more accurate and reliable databases, a great amount of effort was still needed. Although we had the assistance of computer technology that helped avoid most logic mistakes in case reporting, a standardized training for case reporting was still important, especially for the clinicians who just started to use the system. The rapid development of health informatization also increased the need to establish better requirements for data security. Although the issue of underreporting has been greatly decreased, this situation could be reduced further with the progress of system development.

Considering the high level of health informatization in Ningbo, it seems that this model might be difficult to replicate in the provinces and cities with low level of health informatization. However, based on China's current national conditions, there is an irresistible trend for improving informatization level in health industries nationwide, thus the feasibility of promoting and implementing this model nationwide is still achievable. After the completion of NCDCMS in 2018, the average number of visitors from the Health Commission and the CDCs located at different provinces and cities throughout the country was about 100 per year (as per data from Ningbo CDC), which also proved the enthusiasm for building health information platforms nationwide. At the time of writing, NCDCMS has been promoted and implemented in 5 cities in Zhejiang Province.

Driven by the development of health informatization in China, the Ningbo Model created a win–win situation for all stakeholders. Moreover, it actually enabled the possibility of performing population-based research with few human and material resources and helped achieve better data accessibility and utilization. The model has high replicability and generalizability in terms of standard establishment, workload reduction, elimination of underreporting, and improvement of data sharing and utilization. In China, it will guide the construction and establishment of systems for NCDs surveillance and management system in the future.

Acknowledgments

This study was supported by the National Key R&D Program of China (Grant Nos 2017YFC1310902 and 2018YFC1315305), National Natural Science Foundation of China (Grant No. 81872721), and Ningbo Health Branding Subject Fund (Grant No. PPXK2018-10). The authors thank all staff working in the noncommunicable disease surveillance department at the 10 county/district CDCs, general hospitals, community health service centers, and township hospitals for their hard work and assistance in collecting and revising surveillance data. We also thank all participants who contributed to build the Ningbo Model.

Authors' Contributions

SiL and HL were responsible for drafting and writing the manuscript, designing the functions of NCDCMS, and coordinating different stakeholders during the construction of the model. LZ was responsible for building the architecture of NCDCMS and data security. ShL and RH were responsible for literature collection and manuscript revision. SiL and LZ contributed equally to this work.

Conflicts of Interest

None declared.

References

1. World Health Organization. Noncommunicable Diseases. Fact Sheets. World Health Organization: Geneva; 2018. URL: https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2018-04-01]
2. GBD 2016 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
3. Fang L, Gao P, Bao H, Tang X, Wang B, Feng Y, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. Lancet Respir Med 2018 Jun;6(6):421-430. [doi: 10.1016/S2213-2600(18)30103-6] [Medline: 29650407]
4. Lu J, Lu Y, Wang X, Li X, Linderman GC, Wu C, et al. Prevalence, awareness, treatment, and control of hypertension in China: data from 1·7 million adults in a population-based screening study (China PEACE Million Persons Project). Lancet 2017 Dec 09;390(10112):2549-2558. [doi: 10.1016/S0140-6736(17)32478-9] [Medline: 29102084]

XSL•FO
RenderX

5. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. JAMA 2017 Dec 27;317(24):2515-2523. [doi: 10.1001/jama.2017.7596] [Medline: 28655017]
6. Beijing National Health Commission of the People's Republic of China. China NHCotPsRo. Diagram: China National Nutrition and Chronic Disease Status Report. 2015. URL: http://www.nhc.gov.cn/jkj/s5879/201506/4505528e65f3460fb88685081ff158a2.shtml [accessed 2018-04-30]
7. Sacco RL, Roth GA, Reddy KS, Arnett DK, Bonita R, Gaziano TA, et al. The Heart of 25 by 25: Achieving the Goal of Reducing Global and Regional Premature Deaths From Cardiovascular Diseases and Stroke: A Modeling Study From the American Heart Association and World Heart Federation. Circulation 2016 Jun 07;133(23):e674-e690. [doi: 10.1161/CIR.0000000000000395] [Medline: 27162236]
8. Atlanta Centers for Disease Control and Prevention. Improving Disease Surveillance at CDC - A Strategic Approach. 2017. URL: https://www.cdc.gov/surveillance/Improving-Surveillance-Background.html [accessed 2018-04-30]
9. Li Y, Jiang Y, Zhang M, Yin P, Wu F, Zhao W. Drinking behaviour among men and women in China: the 2007 China Chronic Disease and Risk Factor Surveillance. Addiction 2011 Nov;106(11):1946-1956. [doi: 10.1111/j.1360-0443.2011.03514.x] [Medline: 21771141]
10. Liu S, Zhang M, Yang L, Li Y, Wang L, Huang Z, et al. Prevalence and patterns of tobacco smoking among Chinese adult men and women: findings of the 2010 national smoking survey. J Epidemiol Community Health 2016 Sep 22;71(2):154-161. [doi: 10.1136/jech-2016-207805]
11. Zhao ZP, Wang LM, Li YC, Jiang Y, Zhang M, Huang ZJ, et al. [Provincial representativeness assessment of China Non-communicable and Chronic Disease Risk Factor Surveillance System in 2013]. Zhonghua Yu Fang Yi Xue Za Zhi 2018 Feb 06;52(2):165-169. [doi: 10.3760/cma.j.issn.0253-9624.2018.02.009] [Medline: 29429271]
12. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
13. Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, NESS-China Investigators. Prevalence, Incidence, and Mortality of Stroke in China: Results from a Nationwide Population-Based Survey of 480 687 Adults. Circulation 2017 Dec 21;135(8):759-771. [doi: 10.1161/CIRCULATIONAHA.116.025250] [Medline: 28052979]
14. Wang L, Deng Q, Wang LJ. Review and Prospect of comprehensive monitoring of chronic diseases in China. Chinese Journal of the Frontiers of Medical Science 2014;6(3):1-4 [FREE Full text] http://www.cnki.com.cn/Article/CJFDTotal-YXQY201403001.htm
15. Ding KQ, Zhang L, Dong HJ, Yi B, Ma JQ, Sun XD, et al. The exploration and practice on the mode of infectious diseases reporting based on intelligent platform. Chinese Journal of Health Informatics and Management 2017;14(1):50-54 [FREE Full text] http://qikan.cqvip.com/Qikan/Article/Detail?id=671702982 [doi: 10.3784/j.issn.1003-9961.2018.09.013]
16. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: 18634202]
17. Lisa MS, Teutsch SM, Thacker SB, St. Louis ME. Principles & Practice of Public Health Surveillance. New York: Oxford University Press; 2010.
18. Ying YY, Zhang L, Li L, Li H, Ji W, Li SX, et al. Epidemiological analysis on the diabetes mellitus with oculopathy based on the regional health information platform. Chinese Journal of Prevention and Control of Chronic Diseases 2017;25(3):180-183 [FREE Full text] http://d.wanfangdata.com.cn/periodical/zgmxbyfykz201703005
19. Duan D, Xu J, Feng X, Astell-Burt T, Xu G, Lu N, et al. Does body mass index and adult height influence cancer incidence among Chinese living with incident type 2 diabetes? Cancer Epidemiol 2018 Apr;53:187-194. [doi: 10.1016/j.canep.2018.02.006] [Medline: 29518753]
20. Lin H, Tang X, Shen P, Zhang D, Wu J, Zhang J, et al. Using big data to improve cardiovascular care and outcomes in China: a protocol for the CHinese Electronic health Records Research in Yinzhou (CHERRY) Study. BMJ Open 2018 Feb 12;8(2):e019698 [FREE Full text] [doi: 10.1136/bmjopen-2017-019698] [Medline: 29440217]

Abbreviations

CDC: Centers for Disease Control and Prevention
COPD: chronic obstructive pulmonary disease
EHR: electronic health records
HIS: hospital information system
NCDCMS: the integrated noncommunicable disease collaborative management system
NCDs: noncommunicable diseases

XSL•FO
RenderX

Edited by G Eysenbach; submitted 15.12.19; peer-reviewed by N Ng, W Tumuhimbise; comments to author 03.02.20; revised version received 26.03.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Li S, Zhang L, Liu S, Hubbard R, Li H
Surveillance of Noncommunicable Disease Epidemic Through the Integrated Noncommunicable Disease Collaborative Management System: Feasibility Pilot Study Conducted in the City of Ningbo, China
J Med Internet Res 2020;22(7):e17340
URL: http://www.jmir.org/2020/7/e17340/
doi:10.2196/17340
PMID:32706706

©Sixuan Li, Liang Zhang, Shiwei Liu, Richard Hubbard, Hui Li. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Considerations for Health Researchers Using Social Media for Knowledge Translation: Multiple Case Study

Sarah A Elliott[1,2,3], PhD; Michele P Dyson[1], PhD; Gilbert V Wilkes[4], PhD; Gabrielle L Zimmermann[2,5], PhD; Christine T Chambers[6,7], PhD; Kristy DM Wittmeier[8,9], PhD; Dianne J Russell[10], PhD; Shannon D Scott[11], PhD; Denise Thomson[2,3], MA; Lisa Hartling[1,2,3], PhD

[1]Alberta Research Center for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
[2]Alberta SPOR SUPPORT Unit, Department of Pediatrics, University of Alberta, Edmonton, AB, Canada
[3]Cochrane Child Health, Edmonton, AB, Canada
[4]School of Communication Studies, Mount Royal University, Calgary, AB, Canada
[5]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada
[6]Departments of Psychology & Neuroscience and Pediatrics, Dalhousie University, Halifax, NS, Canada
[7]Centre for Pediatrics Pain Research, IWK Health Centre, Hallifax, NS, Canada
[8]Department of Pediatrics and Child Health, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada
[9]Children's Hospital Research Institute of Manitoba, Winnipeg, MB, Canada
[10]CanChild Centre for Childhood Disability Research, McMaster University, Hamilton, ON, Canada
[11]Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Sarah A Elliott, PhD
Alberta Research Center for Health Evidence
Department of Pediatrics
University of Alberta
11405 - 87 Avenue Edmonton
Edmonton, AB
Canada
Phone: 1 7804926124
Email: se2@ualberta.ca

Abstract

Background: Despite extensive literature describing the use of social media in health research, a gap exists around best practices in establishing, implementing, and evaluating an effective social media knowledge translation (KT) and exchange strategies.

Objective: This study aims to examine successes, challenges, and lessons learned from using social media within health research and to create practical considerations to guide other researchers.

Methods: The Knowledge Translation Platform of the Alberta Strategy for Patient-Oriented Research SUPPORT Unit formed a national working group involving platform staff, academics, and a parent representative with experience using social media for health research. We collected and analyzed 4 case studies that used a variety of social media platforms and evaluation methods. The case studies covered a spectrum of initiatives from participant recruitment and data collection to dissemination, engagement, and evaluation. Methods and findings from each case study as well as barriers and facilitators encountered were summarized. Through iterative discussions, we converged on recommendations and considerations for health researchers planning to use social media for KT.

Results: We provide recommendations for elements to consider when developing a social media KT strategy: (1) set a clear goal and identify a theory, framework, or model that aligns with the project goals and objectives; (2) understand the intended audience (use social network mapping to learn what platforms and social influences are available); (3) choose a platform or platforms that meet the needs of the intended audience and align well with the research team's capabilities (can you tap into an existing network, and what mode of communication does it support?); (4) tailor messages to meet user needs and platform requirements (eg, plain language and word restrictions); (5) consider timing, frequency, and duration of messaging as well as the nature of interactions (ie, social filtering and negotiated awareness); (6) ensure adequate resources and personnel are available (eg, content creators, project coordinators, communications experts, and audience stakeholder or patient advocate); (7) develop

XSL•FO
RenderX

an evaluation plan a priori driven by goals and types of data available (ie, quantitative and qualitative); and (8) consider ethical approvals needed (driven by evaluation and type of data collection).

Conclusions: In the absence of a comprehensive framework to guide health researchers using social media for KT, we provide several key considerations. Future research will help validate the proposed components and create a body of evidence around best practices for using and evaluating social media as part of a KT strategy

(J Med Internet Res 2020;22(7):e15121) doi:10.2196/15121

KEYWORDS

social media; knowledge translation; health research; engagement; dissemination; exchange; evaluation

Introduction

Background

The concept of social media has been constantly evolving alongside advances in technology and the development and abandonment of various platforms. Social media has now become an integral means for sharing information while engaging and interacting with others across all sectors of society [1,2]. Subsequently, engaging with anyone, anywhere, at any time has presented opportunities for people, businesses, nonprofit organizations, and health care organizations to engage with a wide audience [3,4]. Social media have changed people's relationships with how they find and use information. Health care consumers (ie, members of the public) can now use web-based platforms to search for health information, seek others' experiences and opinions, create their own interest and advocacy groups to raise awareness, fundraise, and share information among like-minded communities. Similarly, health researchers have started to capitalize on social media for a number of purposes, from patient engagement to research dissemination and exchange, with countless applications in between [5-7]. Evidence suggests that health researchers largely hold favorable views with respect to the use of social media in a professional capacity [8,9], considering it to provide an engaging and convenient source of information [10], while meeting their need to gain and exchange knowledge with relevant stakeholders [6]. Despite the enthusiasm, there remains a gap around best practices in developing and implementing social media as part of a successful knowledge translation (KT) strategy.

Knowledge Translation Theories and Social Media Models

Many KT strategies (often guided by theories, models, or frameworks) require that the goals, audience, and messaging be carefully considered to ensure success. However, there is no current KT theory, model, or framework that addresses what makes a successful *social media* KT strategy. The unique requirements that social media has around the what (materials and procedures needed), who (provider), and how (mode of delivery and platform and to whom) are quite nuanced. Likewise, current social networking theories and models do not adequately address all facets of dissemination, exchange, and application of knowledge needed to develop (and evaluate) a KT strategy [2,11-13].

Integrated KT models emphasize that translation of knowledge is expedited when knowledge producers (eg, researchers and scientists) and knowledge users (eg, patients, caregivers, and policy makers) are known to one another and are familiar with one another's needs, preferences, and circumstances [14,15]. Pick et al [16] conceptualized the 4 Cs of social media in an attempt to describe how social media is used, proposing that *content* is developed within a given *context*, to make *connections* leading to *conversations*. However, the ability of social media users to successfully share, mobilize, and cocreate knowledge still needs to be understood.

Connecting the Dots: Social Media and Knowledge Translation

We propose that to facilitate effective use of social media as part of a KT strategy, researchers need to know how to reach and engage their target audience through social media, with the information the research teams want to share. Likewise, considerations around what stakeholders want to hear and learn need to be incorporated. There are 3 components of social media that can be classified as explanatory and predictive, which could help guide this [17]:

1. Posting frequency, which means how often a user (person, group, or enterprise) circulates content, asks for input, or responds to a comment or request, without respect to any other factor, that is, quality, depth of detail (whether a post is a simple headline or a detailed infographic), and kind of post (response to another user, announcement of an event or a resource, or key findings of a study) [17-19].
2. Whether and to what degree the user can affect or create a sense of awareness (or negotiated awareness) of who they are, their purpose, their location, the character of their expertise or experience, and the precise services or information they have on hand for their intended audience [17].
3. Whether and to what degree the user allows or affords other users to select effectively within their content and to find what they need to find and whether and to what degree the selection activity is used by the user as grounds to optimize or refine their offerings (social filtering) [20].

To date, no KT or social media framework incorporates these explanatory and predictive components. In the absence of a suitable framework, we draw on 4 case studies using social media for engagement and dissemination of health research. We summarize the successes, challenges, and lessons learned and present recommendations for others considering social media as part of a KT strategy.

XSL•FO
RenderX

Methods

General Approach

In 2016, the Alberta Strategy for Patient-Oriented Research (SPOR) SUPPORT Unit's Knowledge Translation Platform hosted a national meeting (which included a number of KT and communication experts, platform staff, and consumers) to explore the potential linkage between social media and KT research. Subsequently, a national working group was formed of representatives (n=12) from academia (ie, faculty and research staff) and community (ie, knowledge brokers and parent representative), all with experience using social media as part of a KT strategy for health research. The working group defined the approach to the topic and scope of the project. The goal was to create a document that compared researchers' experiences and provided considerations to guide other researchers wanting to utilize social media as part of a KT strategy. The working group met monthly over the course of a year (November 2017 to November 2018), collectively identified possible case studies for inclusion based on pre-established inclusion criteria (described below), and developed data collection forms for extracting key findings. Once data from each case study were collected and summarized, deliberative dialogue [21] was used to draw out important concepts and points to consider for discussion. Working group members were then responsible for creating considerations for other researchers through iterative discussions. All working group members were invited to help draft the final report and be part of the authors' team.

Social Media Case Studies

The aim was to collect case studies of social media use in health research that covered a spectrum of KT initiatives as well as a mix of social media platforms and evaluation methods. Case studies had to be focused on child health, include an evaluation component, and have been published in the past 5 years. The working group (n=12) first looked internally to gather case studies and evaluated the collective breadth of experience using social media for KT activities. An invitation was then extended to other research groups to supplement the working group's experiences. Researchers providing case studies were also invited to be part of the working group and authors' team. Four case studies were selected by the working group for inclusion.

The authors of the case studies were then asked to provide information and data about the project in a pre-established data collection form. The data collection form gathered information about the case study's topic and objectives, intended audience, project logistics (eg, social media platform(s) used, staffing requirements, and project timeline), use of any predictive or explanatory factors such as posting frequency, negotiated awareness and social filtering, dimensions of communication used (Multimedia Appendix 1), evaluation methods (including outcomes and metrics), key findings, and self-reported lessons learned (ie, successes and challenges). To ensure validity and reliability of the data collected from the case studies, collated and summarized data were presented to case study authors for review.

Group Discussions and Development of Considerations

Once data were collected, a summary document that highlighted the similarities and differences between the case studies was created. Self-reported barriers and facilitators to KT encountered when using social media were also collated. Barriers and facilitators were coded into categories based on what part of the KT strategy they impacted (eg, planning, doing, and evaluating). These categories were then broken down into specific elements that formed part of each category (eg, planning—personnel and resource requirements, platform used, and scope of messaging). This prompted focused discussions on the overall key successes and challenges. Working from the summary tables and coded elements, deliberative dialogue [21] (facilitated by SE) was used to draw out important concepts and points to consider for discussion. Through an iterative discussion process, the working group converged on considerations for researchers planning to use social media for KT, guided by the explanatory and predictive components of social media.

Results

Summary of Case Studies

The 4 case studies are summarized in Table 1. More details are presented in Multimedia Appendix 2.

Here, we synthesize the common case study elements (processes of implementing the social media KT strategies) with regard to lessons learned and challenges faced.

XSL•FO
RenderX

Table 1. Case study summaries.

Case study	Objective	Intended audience	Platform	Dimensions of communication[a]	Time frame and intensity (posting frequency)	Social filtering	Negotiated awareness	Staffing requirements	Evaluation
Cochrane Child Health [7]	Dissemination	Child health HCP[b], child health researchers, and health care organizations	Twitter and blog	Asynchronous, one to many, dynamic, pull, remote, and iconic indexical	6 months, weekly blog posts, daily tweets, monthly journal club on Twitter	Influencers were tagged in tweets; hashtags were used to identify conditions under discussion	Content shared by influencers tagged and research group stakeholders	0.2 FTE[c] RC[d], 0.2 FTE, information specialist, and 0.3 FTE graduate student or RA[e]	Twitter analytics, Bitly statistics, Altmetric scores, journal club, and participant survey feedback
#ItDoesntHaveToHurt [22]	Dissemination and engagement	Parents (primarily mothers)	Blog, Twitter, Facebook, Instagram, YouTube, and stakeholder websites	Asynchronous, one to many, dynamic, remote, recorded, focal, push, durable, phonetic syllabic, and iconic indexical	12 months, varied with approximately 3 posts per month plus amplification and sharing	Content appeared on partner website and social media with a well-established community of parents, influencer promotion and engagement (parent bloggers across the country amplified content using their own social channels), hashtag use #ItDoesntHaveToHurt	Content shared by influencers, stakeholders, research groups' social media channels, and research partners' social media channels	0.75 FTE RC, 0.2 FTE, stakeholder PC[f], and digital content creators, as needed	Web-based analytics, pre- and postsurveys and interviews, social listening and sentiment analysis, media analysis, partnership analysis, and patient engagement evaluation
Hirschsprung's Disease Community [23]	Engagement (with caregivers to identify knowledge gaps) and knowledge exchange	Caregivers of children with Hirschsprung disease	Twitter, Facebook, and blog	Asynchronous, many to one, dynamic, and remote	1-month, daily interaction from parent-led administrative team, weekly posting and reminders during research	Community was pre-established	Posts and messages were from the administrator of an established community; the community has a clear focus, consistent posting, and committed membership	0.5 FTE RA, parent partner (ongoing community management), and communication staff (ongoing community management)	Google analytics and number of surveys completed
Parents Participating in Research Facebook Group [24]	Knowledge exchange and dialogue	Parents and caregivers of special needs children and child health researchers	Facebook	Asynchronous, many to many, and synchronous	9 months, varied with approximately 4 posts per week	Parent moderators direct messaged or tagged topics to parents and research contributors with possible perspectives to add to the discussion	Content shared by moderators, parent and caregivers, or researchers	2 parent moderators—approximately 0.4-0.7 FTE; KT[g] committee and parent moderator liaison—approximately 0.5 FTE	Analytics of engagement and activity (eg, number of posts and number of comments) and web-based survey of members

[a]See Multimedia Appendix 3.

[b]HCP: health care providers.

XSL•FO
RenderX

[c]FTE: full-time equivalent.

[d]RC: research coordinator.

[e]RA: research assistant.

[f]PC: project coordinator.

[g]KT: knowledge translation.

Planning

It was unanimous that advanced planning was a key component to a successful social media KT strategy. Furthermore, a prior understanding of what theory, model, or framework best suited the goal of the KT strategy was helpful in planning the different approaches. The *Cochrane Child Health* strategy focused on disseminating information. As knowledge producers, they planned and implemented approaches to push (disseminate) knowledge toward audiences who they believed could benefit from receiving it (health care providers and researchers). The *Hirschsprung's Disease Community* and *#ItDoesntHaveToHurt* employed strategies to build upon pre-established web communities. They leveraged partnerships to foster engagement and expedite information exchange with their intended audience (parents). The *Hirschsprung's Disease Community*, *#ItDoesntHaveToHurt*, and *Parents Participating in Research Facebook Group* initiatives planned and implemented strategies to gather knowledge from sources (parents and caregivers) they identified as producing knowledge that was useful to their decision making and research agenda (priority setting and knowledge gaps). Additionally, to further build their communities, they planned to engage with their users through an interactive process (exchanging information through discussion boards, question and answer sessions, survey responses, and feedback sessions).

It was also noted that case studies focusing on a specific topic or disease led to engaging a more saturated audience. One case study reported that it was difficult to tap into any specific existing audience when posting about diverse clinical topics, whereas others that had a more focused topic (that meshed well with the focus of the pre-existing group) felt that a tighter community was developed, which enhanced information sharing.

Planning for the required staff and resources to successfully implement and maintain the KT strategy, including frequent posts and audience interactions, was challenging. The time frame from the case studies ranged from 1 month to 1 year. Case study authors reported not knowing how long a strategy needed to be for it to be effective. The main driver for the length of a strategy was the availability of financial and human resources support.

Many case studies made use of programs to help manage their social media campaigns. Tools such as Buffer (Buffer Inc, 2010) were used to preschedule tweets (the bulk of the campaigns), allowing for monitoring and interactions to be completed throughout the campaign. To keep track of this, the case studies used tools such as TweetDeck (Twitter, 2008) or Hootsuite (Hootsuite Inc, 2008), which provided a user-friendly interface to track and monitor the account, as well as log interaction notifications and collect analytics.

Doing

Platforms, Models of Communication, and Posting Frequency

A variety of different platforms were utilized among the case studies to increase reach and engagement. Three of the case studies used a multiplatform approach in which Twitter, Facebook, Instagram, and blogs were used to cross-post information and engage with multiple audiences across an expansive social network. Cochrane Child Health, whose target audience was health care providers, researchers, and organizations, focused on using Twitter and blog posts and hosting web-based journal clubs. Case studies targeting parents and caregivers, however, tended to use Facebook, YouTube, and Instagram in addition to Twitter. This was a strategic element, based on prior work to understand who their audiences were and where they gathered on the web. It was also noted that having an established platform or network (eg, Cochrane Child Health had >3500 followers at the time of the campaign) was easier than building a network from scratch for a specific campaign. Those that had a pre-established web presence or leveraged partners with such a pre-established presence reported greater interactions and reach.

A common element throughout the case studies was the dimension of communication, with all using an asynchronous approach in which messages were exchanged intermittently rather than in a steady stream. The intensity of communication varied greatly from daily to weekly between and within each case study. Across all case studies, we found that engaging the intended audience through specific posts, rather than constantly passively pushing information, amplified the sharing of information between users. However, it was noted that there was more web-based interaction overall with more frequent posts. Posting frequency was also dictated by platform use. On a platform in which information exchange occurs rapidly in real time (eg, Twitter), it was found to be important to post quite frequently; otherwise, users tended to disengage. On other platforms such as Facebook, 1 case study received feedback from their audience, suggesting that the ability for members to be able to read and contribute at any time (day or night) was what kept them engaged. Allowing users to contribute to ongoing community conversations enhanced engagement and exchange practices.

Negotiated Awareness, Social Filtering, and Content Development

Negotiated awareness and social filtering were valued by all case studies for their ability to expand the reach (numerically and geographically) of the KT strategy. Two case studies (*Cochrane Child Health* and *#ItDoesntHaveToHurt*) used specific social filtering techniques, utilizing social media influencers and hashtags to categorize messages, leading individuals to conversations and discussions pertaining to a

XSL•FO
RenderX

specific topic or theme. These techniques also helped researchers tap into existing networks that amplified sharing by engaging partners and audiences through their own social media networks and interests.

One case study highlighted that the use of an expert media partner who could create compelling digital content and who already had an established reach ensured that content creation was done more efficiently and effectively than traditional academic-led dissemination campaigns. For other case studies, developing content was often time- and resource-intensive, and there was often a trade-off between effort and return. All case study leads agreed that having a team member or partner who specialized in digital content creation is likely an important element of a successful campaign (audience engagement and resource development).

Stakeholder Engagement

The case studies in which patient or parent partners co-designed and facilitated the strategies had an extended reach and higher level of engagement from the web community than the ones in which information was just pushed out by researchers. Capitalizing on a recognized and established network, community, or partnership enabled efficient and effective content creation and sharing.

The *Hirschsprung's Disease Community* and *Parents Participating in Research Facebook Group* projects highlight that working with someone from the campaign's intended audience (eg, parent and clinician) leads to relevant content being produced, developed, and created with that specific audience in mind.

Collaborating with a parent moderator who had credibility within the parent and caregiver community and who understood the needs of the researchers and the overall purpose of the collaborative helped nurture the interconnectedness between members of those web-based communities. It was proposed that having a moderator or stakeholder known within the community added to the perceived credibility of the campaign. Additionally, peer mediation was found to be a facilitating factor in the exposure and diffusion of information on the web. The *Parents Participating in Research Facebook Group* project found a shift in emphasis from its primary goal of creating an advisory group for the researchers to being an active group of knowledge exchange for parents (ie, asking questions, deciphering the research, and providing peer support). Parents reported feeling a sense of belongingness to the community where they could safely share stories, ask questions, and provide support for other members. This provided the researchers with a deeper understanding of what issues were most important to the daily lives of families, a unique insight that might not have come to fruition if the researchers were guiding the discourse.

Resource Requirements

Project teams and personnel needs varied across the case studies. Resources, types of skills, and tasks also varied within and between campaigns. Regardless of the campaign length, all projects required at least the equivalent of one full-time position. All case studies employed a research coordinator. Digital content creators, information specialists, students, or parent partners were also involved depending on the specific nature of the project.

Evaluating

All case studies used various analytics (ie, Twitonomy, Facebook, Bitly, and Altmetric scores) to assess reach and engagement and understand how different facets of the strategies performed. A common evaluation component was web analytics to evaluate engagement (eg, likes, shares, and retweets) and uptake or use (eg, publication downloads and surveys completed) of information. However, not all web analytic programs were created equal. For example, harnessing accurate Facebook analytics with a private Facebook community was challenging because of privacy settings. Additionally, metrics such as Altmetric scores did not allow for the isolation of the study-specific impact and were time dependent (ie, older publications had more time to accumulate higher scores). The variability in the range of reporting periods for these statistics also influenced the usefulness of their collection, with availability ranging from a period of 30 days to all time. This was found to be a limiting factor when information was posted weekly or daily, making assessment of the impact of each post difficult.

A multimodal methodological approach (multiple qualitative and quantitative data collection) used by *#ItDoesntHaveToHurt*, *Hirschsprung's Disease Community*, and the *Parents Participating in Research Facebook Group* helped evaluate less tangible aspects, such as passive involvement or interactions between community members or a shift from research teams pushing information out to audiences facilitating pull. However, it was felt that there was often a trade-off between scientific rigor (ensuring appropriate methods and analytics were used) and being able to adequately evaluate and report impact (more broadly than number of citations, downloads, etc). Case study authors noted that researchers need to understand that social media KT success is not an easily quantifiable measure and that unconventional methods of evaluating success may be needed.

A common challenge was the inability to assess and gain an understanding of passive involvement and attribute knowledge sharing to behavior change and its impact on health and health system outcomes. The main goals of most KT strategies were increasing awareness, knowledge, and uptake of evidence. Although analytics can provide details on proxy indicators (ie, reach, usefulness, use, and collaboration), the ability to assess more distal health outcomes was limited and perhaps unrealistic in these case studies.

Ethics Considerations

The constraints placed by needing ethics approval to evaluate and capture participant behavior as well as to vet media posts (in some cases) was seen as a major challenge. One case study needed to have ethics approval for each social media post, which hindered their timeliness of engaging with the group on the web and highlighted a number of points for consideration. These include the value of spontaneous and organic conversation or the trade-off between not sharing information and the concern of providing health care advice rather than information to participants.

XSL•FO
RenderX

It was also noted that the need for ethics approval and informed participant consent was dictated by the outcomes evaluated. For example, the *Parents Participating in Research Facebook Group* set up rules of engagement that participants had to agree to before joining the community. However, to harvest the ideas and concerns of the group posted to the page to help other researchers know what was important to families, informed consent would have been needed. It felt that this may have stifled conversation, posing the following question: does the discourse of the web-based conversation change when participants are aware of ongoing monitoring and evaluation?

Recommendations and Considerations

From these collective experiences, we put forward recommendations and points to consider when using social media as part of a KT strategy for health research. As with any KT initiative, key areas include clearly defining the purpose (ie, engagement, dissemination, and exchange or dialogue), goals, intended audience, type and volume of resources needed, project scope, study design, and evaluation. In addition, we encourage researchers to carefully consider the 3 explanatory and predictive components of social media (ie, posting frequency, social filtering, and negotiated awareness) when developing an approach and choosing a platform.

The checklist for researchers has been provided in Table 2.

XSL•FO
RenderX

Table 2. Checklist for researchers: what to ask and consider when developing a social media knowledge translation strategy.

Phase	Probing questions	Considerations
Planning		
	Purpose and goal	
	What is the purpose of using SM[a] for this KT[b] strategy?	Are you wanting to engage, disseminate, or create a dialogue?
	What are the objectives of the campaign?	Identify a theory, framework, or model that aligns with the project's purpose[c] (eg, dissemination), objectives (eg, increase knowledge vs behavior change), and topic area (eg, health, sociology, and psychology)
	N/A[d]	The scope of messages (ie, one clear message repeated over time in different formats or messages on multiple topics)
	N/A	Create evaluation framework in planning phase
	What will success look like?	What metrics for quantitative data will you collect and when (eg, at what point, how often, and for how long)?
	How are you going to measure success?	Do you need qualitative evaluations?
	Intended audience	
	Who is the intended audience?	Conduct formal or informal research to help answer these questions
	What are the characteristics of the intended audience?	Conduct a social networking map of the intended audience
	What SM platforms do they use? How do they interact on the web?	Involve some from the intended audience in planning (concepts of integrated KT)
	How do they like to receive information?	This may differ for each end-user group
	Duration and intensity	
	How long will the campaign run?	What resources do you have available?
	How many posts a day or week or month?	This will be dependent on what platforms you use and your intended duration of interactivity
	Can you incorporate social filtering?	What hashtags are currently used for your topic area, and is there already a presence on the web?
Doing		
	Platform and messaging	
	What resonates with the intended audience?	Depending on resources and team expertise, you may use one or multiple platforms
	Are they already gathered in an SM community or are you creating a new presence or community?	Invest in a SM management tool (eg, Hootsuite, Buffer, and Twitonomy)
	What will work best for the campaign's purpose?	Tailor strategy to ensure compatibility with each platform used
	How will messages be structured and delivered (scope, format, plain language, figures, pictures, etc)?	Ensure messages are tailored to target audience needs
	Are you incorporating social filtering or negotiated awareness techniques?	Map out who the influencers are that you want to reach
	Have you reached out to stakeholders? Do you have a campaign champion (preferably a peer of the target group)?	Find someone who will act as a community or peer mediator to build trust within the target group
	Personnel, logistics, and other resources	
	Who are the key personnel involved?	Content and communications experts are minimum requirements
	What skill sets are needed?	Resources, time, and funding needed will differ across platforms
	Do you have an advocate or opinion leader from the intended audience or community?	Connect with organizations with infrastructure already in place
	How and how often will they post or interact with the audience or community?	N/A
	Can you link in with a larger organization?	N/A
	Ethics	

XSL•FO
RenderX

Phase	Probing questions	Considerations
	Is everyone a research participant or peripherally involved because of interactions with a participant?	Evaluate the nature of participant involvement and prospectively consider the potential risks to individuals
	Does the audience understand informed consent and expectations for privacy?	Tailor communication about risks and expectations
	Are you aware of the privacy policies governing the different SM platforms?	Work closely with governing bodies and develop standards of practice (eg, create SM policy or "rules of engagement" document)
Evaluating		
Planning		
	Relates back to the goals—what are you hoping to achieve (eg, increased reach, engagement, and uptake)?	Create a logic model mapping goals to output indicators and ensure output indicators are measurable
	How will you determine whether the SM campaign or strategy was successful?	Build an interim evaluation to ensure goals are being met and redeploy efforts and resources effectively
Collecting		
	Are you able to collect the data you need (eg, quantitative, qualitative, or both)?	Invest in web analytic tracking platforms

[a]SM: social media.

[b]KT: knowledge translation.

[c]Useful resources: World Health Organization communications work [25], value-added research dissemination framework [26], and a guide to KT theory [27].

[d]Not applicable.

Discussion

Case Study Findings

Social media has become prevalent in many people's lives as a means to communicate and exchange knowledge. The diverse formats of multiple, easily accessible platforms allow the creation, sharing, and exchange of ideas and information to various audiences. Users of social media can become interactive and are given equal opportunity for participation in the diffusion of information [28,29]. This shift in the communication landscape has paved the way for different stakeholders to interact and engage with intended end users in less traditional formats.

The case studies identified in this collection highlight the potential of using social media to engage with an intended audience and support collective action. Elements integral to the appeal of social media, personalization, presentation, and participation [30,31], overlap with those of KT. Furthermore, social media platforms may provide a rapid, accessible, and cost-effective means of implementing a widespread KT strategy [9]. Although there is no singular framework to guide the use of social media in a health research context, there are a number of factors that need to be considered when creating, implementing, and evaluating a social media KT strategy.

Considerations

Understand Why You Want to Use Social Media as Part of Your Knowledge Translation Strategy

Social media has the power to engage with a wide variety of end users. Depending on the purpose of the strategy (eg, increase access to health information among the intended audience, increase reach and uptake of information, increase awareness or knowledge, and engage intended audience in dialogue [eg, for priority setting or information sharing]) and process (ie, letting it happen vs making it happen), there are many models that can be applied. These focus on dissemination, communication exchange, engagement, or behavior change [25,27,32-36].

We recommend that researchers select one or more social media platforms that align best with their purpose(s). For example, to create dialogue with patients who experience a common condition, a private Facebook group may be more appropriate than a live tweet chat. However, if the goal is broad dissemination, then tapping into an existing web network such as Twitter and posting frequently using social filtering and negotiated awareness techniques would be advantageous to rapidly reach a broader audience.

The scope of the messaging will also need to be taken into consideration during the development of the strategy. Focusing on a single message (one clear message repeated over time in different formats) or multiple messages (on multiple topics) will have implications for which platform you use, resourcing needs, and campaign length.

How to Understand and Engage Web Audiences?

Understanding your intended audience (eg, those engaged in a topic that can inform priority setting, those less engaged who need to access information, or those resistant to messaging and change in behavior) can help create a targeted strategy.

An in-depth understanding of the social network that you are trying to reach is also necessary to evaluate the current web environment. Understanding the structure of the web-based social networks of the intended audience might lead to the development of web-based algorithms that can detect trusted or influential users [37]. These users can then be identified and

XSL•FO
RenderX

approached (in advance) to help engage a wider audience and disseminate information.

In the context of social media platforms, the users were self-selected; the question, therefore, becomes, how do we reach the users who were already there with the information that they need? Posting frequency is the grossest of metrics but can also be the most highly predictive [38-42]. Followers disengage when they perceive a channel or source to be dormant or dead. If you fail to post frequently in a noisy channel (eg, Twitter or any blog), you effectively do not exist. Additionally, people are more likely to draw close to those whose web activities are clear and apparent; this is why posting focused, tailored content frequently is beneficial. Users can then determine how the group of interest came to know of each other, what type of information (intent or agenda) they provide, and how it is being conveyed (personality). These factors can help them decide whether they want to connect with or follow the group. Social filtering, also collaborative filtering, is how a group organizes to allow people to find what could be useful or entertaining [43-45]. For example, if a tweet gets a lot of likes or retweets or a blog post is circulated across Twitter, Facebook, and numerous aggregators, this is interpreted as the judgment of the community.

By building on both direct (ie, proximate relationships and immediate relationships) and indirect (ie, distal relationships, connected by common factors, or influenced by external factors) social relationships, researchers can find users with similar interests. The impact this has on social capital (greater interactions generate a greater sense of community) should be considered and leveraged to enhance the success of a social media KT strategy. Targeted messaging and sharing content and knowledge that has been contributed or endorsed by other users can then be successfully capitalized on to reach the intended audience [46].

How Can a Social Media Platform Be Chosen?

Articulating the goal(s) of the strategy will help guide what platform should be used and for what purpose. Due to the unique functionality, interface, and content each platform offers, it is expected that people use and actively engage with the various platforms in different ways. For example, Voorveld et al [47] found that Facebook allowed people to respond, share, and be updated on information quickly (social interaction and topicality), whereas Twitter ensured people were quickly informed and up to date but had little topicality. Additionally, as there is often little overlap between different platforms regarding user demographics and conversations, multiple platforms may be necessary to reach all intended audiences [47,48].

Although empirical data to confirm whether cross-posting enhances reach and engagement are lacking, anecdotal evidence suggests that linking posts to multiple social media sites can enhance visibility and reach [47-49]. To gauge this, we suggest having representatives from the intended audience (eg, patient, consumer adviser, and other stakeholders) as part of the team or as a consultant. This will help determine which platform is most suitable and inform how best to engage with the intended audience from a messaging perspective. Additionally, whether one is establishing a presence on single or multiple channels, a pre-established platform or following is easier to partner with rather than building a new network during the campaign.

What Personnel and Resources Will You Require?

Many different skill sets and dedicated personnel are required to effectively execute a social media KT strategy [50]. Many grant-funded research projects are under resourced, and employing staff to come onboard for a select amount of time to execute a strategy may be unsustainable. To overcome some of the constraints placed on researchers, partnering with stakeholder organizations, networks, or advocacy units that have the necessary resources is advantageous. Additionally, media partners can alleviate some of the burden placed on research teams who do not have the necessary skill sets or time to develop multiplatform-compatible social media content or nurture a social media strategy to gain widespread traction.

Although Young et al [51] have suggested a framework for building an effective web-based health community through 4 key phases (ie, inception, establishment, maturity, and mitosis), the applicability of its use within an asynchronous or push-pull communication format via a single social media channel is yet to be determined. In other words, it seems feasible to create a closed Facebook group that is self-sustaining and that will grow organically, yet it is unclear if it is possible to create a similar model on a broader, less confined (less private) platform such as Twitter.

How Will You Evaluate Success and Impact?

The evaluation piece of a social media KT strategy needs to be embedded from the beginning (eg, to be able to see the number of times a link was accessed, the link needs to be set up as a traceable link before the strategy is implemented). There is an ongoing debate about which metrics are the most useful and relevant to determining the success of a KT strategy. Although the number of times a post is viewed or interacted with on social media may give an indication of reach, it is insufficient for assessing influence. Assessing the direct impact on change in attitude or behavior is not easily quantifiable, and alternative methods of measuring success are needed [52].

A range of quantitative analytics can be used to identify the components of a campaign that are most effective for assessing reach and uptake. However, consideration of their reporting periods and comparability across platforms is needed. If assessing reach and engagement, metrics such as likes, shares, and retweets are effective; traceable links and the number of downloads (often available from journals) are useful measures of engagement. Evaluating engagement in more depth can be done by posing questions or comments and assessing the reaction of the audience. However, unless these are monitored in real time, a retrospective evaluation of the number of comments and interactions between researchers and the consumer audience can be difficult and time-consuming (posts are often moved further down the conversation as more relevant posts are made). Although qualitative data collection could assist in this area, easily obtaining and collecting this information may not be feasible on a broad scale.

XSL•FO
RenderX

One construct that has the potential to enhance our understanding of KT success is social network analysis (SNA). SNA proposes a network-level perspective that examines how connections among individuals or entities and the nature of the associated interactions influence an outcome (eg, accessing or sharing evidence and changing practice behaviors based on evidence) [53]. However, little research evidence is available on the success of social media to improve the understanding or communication of health research findings to patients and families and the public (or those not specifically seeking treatment). Research in this area would benefit the health research community.

Do You Need Ethics Approval?

Regardless of whether evaluation is to be done formally or informally, the ethical considerations around web-based data collection and transmission need to be adhered to by researchers. As these are constantly evolving and changing to reflect both the researchers' need for information and participants' desire for privacy and confidentiality, it is imperative that research institutions and researchers are in agreement and compliance with what constitutes the need for informed consent in the digital world. As privacy policies are variable between social media sites and change frequently, they need to be checked before implementing a strategy, and the constraints imposed by varying privacy policies considered during the planning stages.

Ethics around the use of qualitative data acquired also need to be considered. Often, web-based patient groups can be a rich source of information for researchers wanting to understand the experience and views of patients around a particular condition or care need. The line between public and private domains then becomes blurred, with web-based data collection presenting unique issues during data transmission. According to Eysenbach and Wyatt [54], there are 3 main types of data collection and analysis possible from these web-based social media networks: (1) passive analysis, in which information posted on discussion groups is evaluated without researchers embedding themselves into the network; (2) active analysis, in which researchers participate in communications (response to questions by the group and posting questions to the group for discussion); (3) when researchers identify themselves and explicitly set out to gather information and data (surveys and focus groups). In the first 2 types, internet community members do not expect to be research subjects, which brings the need for and approach to consent into consideration.

Researchers and institutes need to work together to develop a code of conduct that benefits the participant as well as the project's integrity and intended purpose. In 2015, the Connected and Open Research Ethics initiative was launched, providing practical and accessible guidance for researchers designing social media–enabled studies [55]. We recommend that researchers consult these guidelines during the development of any social media KT strategy.

Limitations

Our study used a convenience sample of 4 pediatric case studies, which may not be representative of all health researchers' experiences utilizing social media for KT. However, the generalist recommendations we provided are relevant to all health researchers and propose a broad set of considerations to others wanting to utilize social media for KT.

Although we did not specifically set out to complete a formal consensus building process (eg, Delphi), we acknowledge that this may have been a useful approach to avoid any bias associated with *group think*. Instead, we utilized deliberative dialogue. This process involved listening to other points of view; exploring and searching for new ideas, perspectives, and points of agreement; and bringing unexamined assumptions into the open. The end result produced a number of considerations based on collective insight and judgment on how best to utilize social media as part of a KT strategy.

As data were collected in 2017 to 2018, the findings may not reflect health researchers' current experiences with social media. We evaluated a snapshot of social media use, which may have already changed given the direction and development of new platforms beyond Facebook and Twitter. The findings do, however, provide unique insights and ideas for future research.

Conclusions

Social media has the potential to build and link web-based communities and to engage with, disseminate, and exchange information. Interconnections between people on social network sites as well as negotiated awareness and social filtering can enhance the process of information transfer and exchange and amplify the influence of that information. However, this comes with a cost. Ensuring adequate resources and time are available is essential to ensure a successful social media KT strategy.

A greater understanding of how best to evaluate social media as a KT tool for both active and passive engagement is needed to direct researchers in planning and evaluating their intended strategy. The elements and challenges discussed herein are important for researchers to consider to plan and evaluate a strategy that harnesses the power and personal aspects of social media for KT.

Perhaps a greater challenge is understanding how knowledge sharing and engagement contribute to behavior change and health outcomes and how we can gather and evaluate such outcomes. Further research is needed to help validate the proposed components and create a body of evidence around best practices for utilizing and evaluating social media as part of a KT strategy in health research.

Acknowledgments

The authors would like to thank Ms Megan Nuspl for helping initially coordinate the project and Ms Jennifer Sprung for providing a parent perspective on important outcomes for successful end-user engagement. This work was supported by the Alberta SPOR SUPPORT Unit Knowledge Translation Platform, which is funded by the Alberta Innovates and the Canadian Institutes of Health

XSL•FO
RenderX

Research. LH is a Canada Research Chair in Knowledge Synthesis and Translation. SS is a Canada Research Chair for Knowledge Translation in Child Health. LH and SS are Stollery Science Lab Distinguished Researchers. CC is a Canada Research Chair in Children's Pain. KW is supported by Dr John M Bowman, Chair in Pediatrics and Child Health, through the RH Institute Foundation and University of Manitoba.

Authors' Contributions

SE, MD, DT, GZ, and LH designed the study. Data collection and analysis were conducted by SE with input from MD, CC, KW, and DJR. All authors contributed to iterative discussions and summary of recommendations. SE and LH drafted the manuscript with input from MD, GZ, and GW. All authors reviewed the manuscript and approved the final version. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Case study data collection form.
[DOCX File , 12 KB - jmir_v22i7e15121_app1.docx]

Multimedia Appendix 2
Case study descriptions.
[DOCX File , 23 KB - jmir_v22i7e15121_app2.docx]

Multimedia Appendix 3
Dimensions of communication.
[DOCX File , 13 KB - jmir_v22i7e15121_app3.docx]

References

1. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2019-01-02]
2. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68 [FREE Full text] [doi: 10.1016/j.bushor.2009.09.003]
3. Clark H. Canadian Social Media Stats. Sherpa Marketing: A Data Driven Marketing Company. 2018. URL: https://www.sherpamarketing.ca/blogs/canadian-social-media-stats-?updated-?2018-471 [accessed 2019-04-30]
4. Puljak L. Using social media for knowledge translation, promotion of evidence-based medicine and high-quality information on health. J Evid Based Med 2016 Feb;9(1):4-7. [doi: 10.1111/jebm.12175] [Medline: 26372327]
5. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Klassen TP, et al. Social media use by health care professionals and trainees: a scoping review. Acad Med 2013 Sep;88(9):1376-1383. [doi: 10.1097/ACM.0b013e31829eb91c] [Medline: 23887004]
6. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
7. Gates A, Featherstone R, Shave K, Scott SD, Hartling L. Dissemination of evidence in paediatric emergency medicine: a quantitative descriptive evaluation of a 16-week social media promotion. BMJ Open 2018 Jun 6;8(6):e022298 [FREE Full text] [doi: 10.1136/bmjopen-2018-022298] [Medline: 29880576]
8. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health researchers' use of social media: scoping review. J Med Internet Res 2019 Nov 13;21(11):e13687 [FREE Full text] [doi: 10.2196/13687] [Medline: 31719028]
9. Tunnecliff J, Ilic D, Morgan P, Keating J, Gaida JE, Clearihan L, et al. The acceptability among health researchers and clinicians of social media to translate research evidence to clinical practice: mixed-methods survey and interview study. J Med Internet Res 2015 May 20;17(5):e119 [FREE Full text] [doi: 10.2196/jmir.4347] [Medline: 25995192]
10. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
11. Ngai EW, Moon KK, Lam S, Chin ES, Tao SS. Social media models, technologies, and applications. Ind Manage Data Syst 2015 Jun 8;115(5):769-802. [doi: 10.1108/imds-03-2015-0075]
12. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]

XSL•FO
RenderX

13. Ryan G, Sfar-Gandoura H. Disseminating research information through Facebook and Twitter (DRIFT): presenting an evidence-based framework. Nurse Res 2018 May 02:- epub ahead of print. [doi: 10.7748/nr.2018.e1562] [Medline: 29717844]
14. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
15. Jacobson N, Butterill D, Goering P. Development of a framework for knowledge translation: understanding user context. J Health Serv Res Policy 2003 Apr;8(2):94-99. [doi: 10.1258/135581903321466067] [Medline: 12820671]
16. Pick T. The Four C's of Social Media Marketing. The B2B Marketing Blog. 2009. URL: https://webbiquity.com/social-media-marketing/the-four-cs-of-social-media-marketing/ [accessed 2019-05-01]
17. Baym N. Interpreting soap operas and creating community: inside a computer-mediated fan culture. J Folk Res 1983;30(2/3):143-176 [FREE Full text]
18. Daoud H. Social Media Marketing: Industry Report. Social Media Marketing. 2016. URL: https://www.socialmediaexaminer.com/?essential-elements-social-media-marketing-strategy/ [accessed 2018-05-01]
19. Brech FM, Messer U, Schee BA, Rauschnabel PA, Ivens BS. Engaging fans and the community in social media: interaction with institutions of higher education on Facebook. J Mark High Educ 2016 Sep 8;27(1):112-130. [doi: 10.1080/08841241.2016.1219803]
20. Lerman K. Social information processing in news aggregation. IEEE Internet Comput 2007 Nov;11(6):16-28. [doi: 10.1109/mic.2007.136]
21. Nagda B, Maxwell K. Deepening the layers of understanding and connection: a critical dialogic approach to facilitating intergroup dialogues. In: Maxwell K, Nagda B, Thompson M, editors. Facilitating Intergroup Dialogues: Bridging Differences, Catalyzing Change. Sterling, VA: Stylus Publishing; 2011:1-22.
22. Chambers C. It Doesn't Have to Hurt: Centre for Pediatric Pain Research. Code + Mortar. 2019. URL: https://itdoesnthavetohurt.ca/ [accessed 2019-04-09]
23. Wittmeier KD, Hobbs-Murison K, Holland C, Crawford E, Loewen H, Morris M, et al. Identifying information needs for Hirschsprung disease through caregiver involvement via social media: a prioritization study and literature review. J Med Internet Res 2018 Dec 21;20(12):e297 [FREE Full text] [doi: 10.2196/jmir.9701] [Medline: 30578208]
24. Russell DJ, Sprung J, McCauley D, de Camargo OK, Buchanan F, Gulko R, et al. Knowledge exchange and discovery in the age of social media: the journey from inception to establishment of a parent-led web-based research advisory community for childhood disability. J Med Internet Res 2016 Nov 11;18(11):e293 [FREE Full text] [doi: 10.2196/jmir.5994] [Medline: 27836818]
25. WHO Strategic Communications Framework for Effective Communications. World Health Organisation. 2017. URL: https://www.who.int/mediacentre/communication-framework.pdf [accessed 2019-01-03]
26. Macoubrie J, Harrison C. The Value-Added Dissemination Framework. Administration for Children and Families. 2013. URL: https://www.acf.hhs.gov/opre/resource/the-value-added-dissemination-framework [accessed 2019-01-03]
27. Estabrooks CA, Thompson DS, Lovely JJ, Hofmeyer A. A guide to knowledge translation theory. J Contin Educ Health Prof 2006;26(1):25-36. [doi: 10.1002/chp.48] [Medline: 16557511]
28. Li C, Lin Y, Yeh M. Forecasting participants of information diffusion on social networks with its applications. Inf Sci 2018 Jan;422:432-446. [doi: 10.1016/j.ins.2017.09.034]
29. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Manag Inf Syst 2014 Dec 8;29(4):217-248. [doi: 10.2753/MIS0742-1222290408]
30. Kietzmann J, Hermkens K, McCarthy I, Silvestre B. Social media? Get serious! Understanding the functional building blocks of social media. Bus Horiz 2011 May;54(3):241-251 [FREE Full text] [doi: 10.1016/j.bushor.2011.01.005]
31. The Health Communicator's Social Media Toolkit. Centers for Disease Control and Prevention. 2011. URL: https://www.cdc.gov/healthcommunication/ToolsTemplates/SocialMediaToolkit_BM.pdf [accessed 2019-01-03]
32. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
33. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
34. Neta G, Glasgow RE, Carpenter CR, Grimshaw JM, Rabin BA, Fernandez ME, et al. A framework for enhancing the value of research for dissemination and implementation. Am J Public Health 2015 Jan;105(1):49-57. [doi: 10.2105/AJPH.2014.302206] [Medline: 25393182]
35. Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009 Sep 1;19(5):503-518 [FREE Full text] [doi: 10.1177/1049731509335569] [Medline: 20976022]
36. Lin N. Social Capital: A Theory of Social Structure and Action. New York, USA: Cambridge University Press; 2002.
37. Mislove A, Marcon M, Gummadi K, Druschel P, Bhattacharjee B. Measurement and Analysis of Online Social Networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. 2007 Presented at: IMC'07; October 24-26, 2007; San Diego, California, USA. [doi: 10.1145/1298306.1298311]
38. Wei C. Formation of Norms in a Blog Community. University Digital Conservancy. 2004. URL: https://conservancy.umn.edu/bitstream/handle/11299/172809/Wei_Formation%20of%20Norms%20in%20a%20Blog%20Community.pdf?sequence=1 [accessed 2019-01-03]

XSL•FO
RenderX

39. Chen S, Xia S, Sun S. [In situ FTIRS studies of CO2 reduction on thin film alloy electrodes]. Guang Pu Xue Yu Guang Pu Fen Xi 2000 Oct;20(5):610-611. [Medline: 12945392]
40. Liu L, Liang H. Influence analysis for celebrities via public cloud and social platform. China Commun 2016 Aug;13(8):53-62. [doi: 10.1109/cc.2016.7563688]
41. Signorelli G, Fernandez M, Guerra F. Quality of Information About Physical Activity in Breast Cancer Facebook Pages: A Preliminary Content Review. In: 30th International Symposium on Computer-Based Medical Systems. 2017 Presented at: CBMS'17; June 22-24, 2017; Thessaloniki, Greece. [doi: 10.1109/cbms.2017.159]
42. Wang Y, Kraut R. Twitter and the Development of an Audience: Those Who Stay on Topic Thrive!. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5-10, 2012; Austin, TX, USA. [doi: 10.1145/2207676.2208614]
43. Chepurna I, Makrerhchi M. Social Filtering: User-Centric Approach to Social Trend Prediction. In: Proceedings of the International Conference on Web Intelligence. 2016 Presented at: WI'16; October 13-16, 2016; Omaha, NE, USA. [doi: 10.1109/wi.2016.0115]
44. Thangadurai K, Venkatesan M. Impact on Social Filtering and Meta Filtering in Recommender Systems. In: World Congress on Computing and Communication Technologies. 2014 Presented at: WCCCT'14; February 27, March 1, 2014; Trichirappalli, India. [doi: 10.1109/wccct.2014.11]
45. Chatti M, Schroeder U, Hendrik T. Harnessing Collective Intelligence in Personal Learning Environments. In: Proceedings of the 12th International Conference on Advanced Learning Technologies. 2012 Presented at: ICALT'12; July 4-6, 2012; Rome, Italy. [doi: 10.1109/icalt.2012.97]
46. Brem A, Bilgram V. The search for innovative partners in co-creation: identifying lead users in social media through netnography and crowdsourcing. J Eng Technol Manage 2015 Jul;37:40-51 [FREE Full text] [doi: 10.1016/j.jengtecman.2015.08.004]
47. Voorveld HA, van Noort G, Muntinga DG, Bronner F. Engagement with social media and social media advertising: the differentiating role of platform type. J Advert 2018 Feb 13;47(1):38-54. [doi: 10.1080/00913367.2017.1405754]
48. Hibu I. The Power of Crossposting on Social Media. Marketing Blitz. 2018. URL: https://marketingbitz.com/the-power-of-crossposting-on-social-media/ [accessed 2019-01-03]
49. To Post the Same Thing on Multiple Social Networks or Not. Modassic Marketing. 2018. URL: https://modassicmarketing.com/?to-post-the-same-thing-on-multiple-social-?networks-or-not [accessed 2019-09-01]
50. Päivi J, Asta B. How researchers use social media to promote their research and network with industry. Technol Innov Manag Rev 2017;7(8):32-39. [doi: 10.22215/timreview/1098]
51. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
52. Barton CJ, Merolli MA. It is time to replace publish or perish with get visible or vanish: opportunities where digital and social media can reshape knowledge translation. Br J Sports Med 2019 May;53(10):594-598. [doi: 10.1136/bjsports-2017-098367] [Medline: 29146754]
53. Glegg SM, Jenkins E, Kothari A. How the study of networks informs knowledge translation and implementation: a scoping review. Implement Sci 2019 Mar 27;14(1):34 [FREE Full text] [doi: 10.1186/s13012-019-0879-1] [Medline: 30917844]
54. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
55. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]

Abbreviations

KT: knowledge translation
SNA: social network analysis
SPOR: Strategy for Patient-Oriented Research

Edited by G Eysenbach; submitted 21.06.19; peer-reviewed by A Cross, S Ganesh, J Foo, G Kernohan; comments to author 03.10.19; revised version received 31.03.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Elliott SA, Dyson MP, Wilkes GV, Zimmermann GL, Chambers CT, Wittmeier KDM, Russell DJ, Scott SD, Thomson D, Hartling L
Considerations for Health Researchers Using Social Media for Knowledge Translation: Multiple Case Study
J Med Internet Res 2020;22(7):e15121
URL: http://www.jmir.org/2020/7/e15121/
doi:10.2196/15121
PMID:32706653

XSL•FO
RenderX

©Sarah A Elliott, Michele P Dyson, Gilbert V Wilkes, Gabrielle L Zimmermann, Christine T Chambers, Kristy DM Wittmeier, Dianne J Russell, Shannon D Scott, Denise Thomson, Lisa Hartling. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adherence to Blended or Face-to-Face Smoking Cessation Treatment and Predictors of Adherence: Randomized Controlled Trial

Lutz Siemer[1,2], Dipl-Psych; Marjolein G J Brusse-Keizer[3], PhD; Marloes G Postel[4,5], PhD; Somaya Ben Allouch[6], PhD; Robbert Sanderman[4,7], PhD; Marcel E Pieterse[2], PhD

[1]Technology, Health & Care Research Group, Saxion University of Applied Sciences, Enschede, Netherlands
[2]Centre for eHealth and Well-being Research, University of Twente, Enschede, Netherlands
[3]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands
[4]Department of Psychology, Health & Technology, University of Twente, Enschede, Netherlands
[5]Tactus Addiction Treatment, Enschede, Netherlands
[6]Digital Life Research Group, Amsterdam University of Applied Science, Amsterdam, Netherlands
[7]Department of Health Psychology, University Medical Center Groningen, Groningen, Netherlands

Corresponding Author:
Lutz Siemer, Dipl-Psych
Technology, Health & Care Research Group
Saxion University of Applied Sciences
MH Tromplaan 28
Enschede, 7513 AB
Netherlands
Phone: 31 17678025906
Email: l.siemer@utwente.nl

Abstract

Background: Blended face-to-face and web-based treatment is a promising way to deliver smoking cessation treatment. Since adherence has been shown to be an indicator of treatment acceptability and a determinant for effectiveness, we explored and compared adherence and predictors of adherence to blended and face-to-face alone smoking cessation treatments with similar content and intensity.

Objective: The objectives of this study were (1) to compare adherence to a blended smoking cessation treatment with adherence to a face-to-face treatment; (2) to compare adherence within the blended treatment to its face-to-face mode and web mode; and (3) to determine baseline predictors of adherence to both treatments as well as (4) the predictors to both modes of the blended treatment.

Methods: We calculated the total duration of treatment exposure for patients (N=292) of a Dutch outpatient smoking cessation clinic who were randomly assigned either to the blended smoking cessation treatment (n=130) or to a face-to-face treatment with identical components (n=162). For both treatments (blended and face-to-face) and for the two modes of delivery within the blended treatment (face-to-face vs web mode), adherence levels (ie, treatment time) were compared and the predictors of adherence were identified within 33 demographic, smoking-related, and health-related patient characteristics.

Results: We found no significant difference in adherence between the blended and the face-to-face treatments. Participants in the blended treatment group spent an average of 246 minutes in treatment (median 106.7% of intended treatment time, IQR 150%-355%) and participants in the face-to-face group spent 238 minutes (median 103.3% of intended treatment time, IQR 150%-330%). Within the blended group, adherence to the face-to-face mode was twice as high as that to the web mode. Participants in the blended group spent an average of 198 minutes (SD 120) in face-to-face mode (152% of the intended treatment time) and 75 minutes (SD 53) in web mode (75% of the intended treatment time). Higher age was the only characteristic consistently found to uniquely predict higher adherence in both the blended and face-to-face groups. For the face-to-face group, more social support for smoking cessation was also predictive of higher adherence. The variability in adherence explained by these predictors was rather low (blended R^2=0.049; face-to-face R^2=0.076). Within the blended group, living without children predicted higher adherence to the face-to-face mode (R^2=0.034), independent of age. Higher adherence to the web mode of the blended treatment

XSL•FO
RenderX

was predicted by a combination of an extrinsic motivation to quit, a less negative attitude toward quitting, and less health complaints (R^2=0.164).

Conclusions: This study represents one of the first attempts to thoroughly compare adherence and predictors of adherence of a blended smoking cessation treatment to an equivalent face-to-face treatment. Interestingly, although the overall adherence to both treatments appeared to be high, adherence within the blended treatment was much higher for the face-to-face mode than for the web mode. This supports the idea that in blended treatment, one mode of delivery can compensate for the weaknesses of the other. Higher age was found to be a common predictor of adherence to the treatments. The low variance in adherence predicted by the characteristics examined in this study suggests that other variables such as provider-related health system factors and time-varying patient characteristics should be explored in future research.

Trial Registration: Netherlands Trial Register NTR5113; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=5113

(J Med Internet Res 2020;22(7):e17207) doi:10.2196/17207

KEYWORDS

blended treatment; smoking cessation; adherence; predictors; tobacco; prevention

Introduction

Background

As smoking remains the leading cause of preventable death, cessation treatment is pivotal for public health promotion [1]. The introduction of electronic health (eHealth) [2] represents the expectation of information and communication technologies to improve health care [3]. However, adherence is generally low in web-based treatment [4] as well as in cessation treatment in general [5]. Low adherence is problematic because adherence has been shown to be an indicator of a treatment's acceptability and a determinant of treatment effectiveness [6-9]. Therefore, adherence should be optimized because—assuming a dose-response relationship [10]—patients are more likely to quit smoking if they are more exposed to active ingredients of the treatment [6]. Adherence in general can be defined as the extent to which a person's behavior (eg, taking medication, following a diet, or executing lifestyle changes) corresponds with recommendations from a health care provider [6]. In the context of behavioral change treatments such as smoking cessation, adherence issues are mainly related to premature termination of treatment and failure to perform tasks and exercises between sessions [11].

Blended Treatment

In the past few decades, a variety of effective interventions for smoking cessation have become available [12,13], including more recently developed eHealth services such as web-based interventions [14,15] or mobile phone interventions [16,17]. At present, traditional face-to-face interventions on the one hand, and both web-based and mobile phone interventions on the other hand are increasingly being transferred to blended treatment. Blended treatment is a promising eHealth service, because it is expected that the strengths of one mode of delivery will compensate for the weaknesses of the other [4,18-23]. The main strength of face-to-face treatment is to provide personal attention of a professional, which could compensate for the lack of face-to-face contact in web-based treatment. In turn, a main feature of web-based care is the accessibility anytime and anywhere, which could compensate for the time between face-to-face sessions when the user needs support. Blended treatment is applied in diverse settings (eg, individual vs group setting [24]), addresses several health issues (eg, depression [25], anxiety [26], or addiction [10,22]), uses various tools (eg, web platforms, email, SMS text messaging, apps [20,27]), and uses different modes of delivery (eg, mainly web-based [28,29] vs mainly face-to-face [26,30] or integrated [10,22] vs sequential [29]). Since a clear definition of blended interventions is still missing [18], in this paper we define blended treatment as a combination of face-to-face sessions and web-based sessions to an integrated treatment that can be delivered by health care professionals on an outpatient basis. The blended intervention adopted in this study is an integrated equal blend of face-to-face treatment and treatment via an online platform.

Adherence to Blended Treatment

Blended treatment has been shown to positively influence adherence [4,31-33]. However, to the best of our knowledge, no studies to date have directly compared adherence to a blended treatment with adherence to either a web-based treatment or a face-to-face treatment with identical active components. In this study, we used data from the LiveSmokefree study [22], which is a randomized controlled trial (RCT) comparing the effectiveness of a blended face-to-face and web-based smoking cessation treatment to a comparable face-to-face treatment. In a prior study, we explored measurement methods and levels and predictors of adherence to the blended treatment by including the blended treatment group participants of the RCT only [10]. In the current study, we extended this previous work by including participants from the face-to-face treatment, allowing for a direct comparison of the levels of adherence between blended and face-to-face treatments. Furthermore, to explore whether both modes of delivery within the blended treatment were used in equal frequency, we also focused on levels of adherence within the blended group in the two modes.

Predictors of Adherence

When adherence is low, adherence predictors become an area of interest because they may provide insight into the cause of low adherence and can help to generate new approaches to improving treatment or better alignment between the patient and treatment. Adherence, in general, is determined by provider behaviors, health system factors, and patient characteristics, and the latter have been most extensively examined as predictors of adherence to traditional interventions [6]. Within the context

XSL•FO
RenderX

of smoking cessation treatment—including both face-to-face and web-based treatments—several demographic, smoking-related, and health-related predictors of adherence have been examined. To date, several studies have indicated that the likelihood of being adherent may increase with a higher age [34,35], male gender [35], higher internet skills [36,37], negative attitude toward smoking and higher motivation to quit at baseline [38,39], higher self-efficacy at baseline [39], early success in quitting after the start of the treatment [7,34,40], and lower nicotine dependency at baseline with fewer withdrawal symptoms after quitting [35,38]. For blended treatment, our previous study showed that higher adherence was best predicted by marital status (ie, having a partner) and social modeling (ie, more nonsmoking friends/partner) [10]. Building on this work, we have expanded on these previous findings in the current study by examining the predictors for adherence to both treatment arms (blended and face-to-face) and additionally the two modes of delivery (face-to-face mode and web mode) in the blended arm.

Objectives

In detail, this study explored the following questions. With respect to adherence, we asked (1) how adherent are participants to blended compared to face-to-face treatment? and (2) within the blended treatment group, how adherent are the participants to the face-to-face mode compared to the web mode? With respect to predictors, we asked (1) which demographic, smoking-related, and health-related patient characteristics predict adherence to blended and to face-to-face treatments, and to both groups combined? and (2) within the blended group, which of these characteristics predict adherence to the face-to-face mode and to the web mode?

Methods

Study Subjects

In this study, we used the already available data from patients (blended n=130; face-to-face n=162) of a not yet completed nonblinded RCT on the effectiveness of a blended smoking cessation treatment compared with a face-to-face treatment (LiveSmokefree study, n=172 allocated per group to determine a difference in abstinence rates of 5 percentage points with a power of 80% and =.025) [22]. The patients were referred to the outpatient smoking cessation clinic at Medical Spectrum Twente hospital (Enschede, the Netherlands) by the treating physicians of the hospital or by their general practitioners, and attended the initial treatment session between May 2015 and September 2018. Inclusion criteria were: (1) willing to quit smoking, (2) aged 18 or older, and (3) current daily smoker (at least one cigarette a day). Exclusion criteria were: (1) no internet access (ie, email, websites) and (2) not able to read or write in the Dutch language. In line with the Dutch Medical Research Ethics Committee guidelines, the study was approved by the accredited MEC Twente (P14-37/NL50944.044.14). Before initiation, the study was registered in the Netherlands Trial Registry (NTR5113). All patients had to sign an informed consent form before they were randomized.

Randomization

Patients were randomly assigned to either the blended or face-to-face group. Randomization was performed at the individual level (allocation ratio 1:1) using QMinim Online Minimization [41]. The minimization was stratified according to: (1) level of internet skills [42], (2) level of nicotine dependence (Fagerstrom) [43,44], and (3) the quitting strategy favored by the patient (stop at once, gradual change, scheduled reduced smoking; for details see the description of the study intervention below). The data used for QMinim minimization were collected using the baseline questionnaire completed online by the patient at home prior to the start of treatment.

Study Interventions

The study interventions to be compared were a blended face-to-face and web-based smoking cessation treatment and a face-to-face treatment alone. Except for the differences in mode of delivery (ie, face-to-face mode and web mode), both treatments included the following same features: (1) high-intensity treatments comprising 10 sessions with a total treatment time of 230 minutes (20 minutes each, except for the first that was 50 minutes); (2) delivered by health care professionals in an outpatient cessation clinic; (3) derived from the Dutch Guideline for Tobacco Addiction [45] fulfilling the requirements of the Dutch care module for smoking cessation [46]; (4) executed by counselors registered in the Dutch quality register of qualified smoking cessation counselors; (5) treatment costs reimbursed by the patient's health insurance; (6) supporting three quitting strategies that were chosen at the start of the treatment (stop at once, change gradually by increasing the number of daily activities that are performed smoke-free, or decrease smoking at regular intervals such as scheduled smoking reduction 100%-75%, 75%-50%, etc). The chosen quitting strategy did not influence the course of the treatment in general. The order, pace, duration, and intensity were the same for all strategies.

Both the blended and face-to-face treatments included the following behavior change techniques, according to BCT taxonomy v1 of Michie et al [47]: 1.1 Goal setting (behavior), 1.2 Problem solving, 1.3 Goal setting (outcome), 1.4 Action planning, 1.5 Review behavior goal(s), 1.6 Discrepancy between current behavior and goal, 1.8 Behavioral contract, 1.9 Commitment, 2.3 Self-monitoring of behavior, 2.4 Self-monitoring of outcome(s) of behavior, 2.6 Biofeedback, 2.7 Feedback on outcome(s) of behavior, 3.1 Social support (unspecified), 4.2 Information about antecedents, 4.3 Reattribution, 5.1 Information about health consequences, 5.2 Salience of consequences, 5.3 Information about social and environmental consequences, 5.4 Monitoring of emotional consequences, 5.5 Anticipated regret, 5.6 Information about emotional consequences, 6.2 Social comparison, 6.3 Information about others' approval, 7.4 Remove access to the reward, 8.1 Behavioral practice/rehearsal, 8.2 Behavior substitution, 8.3 Habit formation, 8.4 Habit reversal, 8.6 Generalization of a target behavior, 8.7 Graded tasks, 9.1 Credible source, 9.2 Pros and cons, 9.3 Comparative imagining of future outcomes, 10.7 Self-incentive, 10.9 Self-reward, 11.1 Pharmacological support (eg, nicotine replacement therapy [patches, gum], bupropion,

XSL•FO
RenderX

varenicline), 11.2 Reduce negative emotions, 12.1 Restructuring the physical environment, 12.2 Restructuring the social environment, 12.3 Avoidance/reducing exposure to cues for the behavior, 12.4 Distraction, 13.1 Identification of self as role model, 13.2 Framing/reframing, 13.5 Identity associated with changed behavior, 14.4 Reward approximation, 14.5 Rewarding completion, 14.6 Situation-specific reward, 14.7 Reward incompatible behavior, 14.8 Reward alternative behavior, 15.1 Verbal persuasion about capability, 15.3 Focus on past success, and 16.3 Vicarious consequences.

The face-to-face treatment consisted of 10 face-to-face sessions delivered at the outpatient smoking cessation clinic. The blended treatment comprised 5 face-to-face sessions at the outpatient clinic and 5 web-mode sessions delivered via an online treatment platform. Both the face-to-face and blended treatments consisted of both counselor-dependent and counselor-independent components. The counselor-dependent web-based components of the blended treatment were interactive and relied on (asynchronous) communication (email, messaging) between the counselor and participant. The counselor-independent components such as psychoeducational content or a smoking diary were used by the participants on their own and in their own time. In the face-to-face group, these components were provided in a paper manual that the participants took home. In the blended treatment, these components were accessible online. As such, both treatments were equivalent with regard to content and intensity. An additional benefit of the blended treatment was that the content of previous counselor-dependent components remained accessible as email and messaging correspondence saved online.

The most characteristic feature of the blended treatment examined in this study is an equal balance between the face-to-face and web mode sessions; that is, the focus of the treatment is neither on face-to-face mode nor web mode. In addition, there is constant alternation and interactive use of the two modes. Table 1 shows the order, timing, main features, duration, and modes of delivery of the treatment sessions in the face-to-face and blended treatments. Although an equal distribution was planned for the blended treatment with regard to the number of sessions, there was an uneven distribution for the duration of treatment because the first session (50 minutes for face-to-face mode) was longer than the remaining sessions (20 minutes for the face-to-face mode or 20 minutes for web mode); therefore, the participants in the blended group spent 130 minutes in face-to-face mode and 100 minutes in web mode.

A detailed description of the treatments can also be found in the protocol article of the RCT [22] and in the description of the user experience of the blended smoking cessation treatment [48]. Screenshots of the web sessions of the blended treatment are shown in Multimedia Appendix 1 to provide an impression of the look and feel of the web interventions.

Table 1. Order, timing, main features, duration, and mode of delivery of the treatment sessions in the face-to-face and blended groups according to treatment protocol.

Session	Week	Main features	Duration (minutes)	Mode of delivery	
				BSCT[a]	F2F[b]
1	1	Goal setting; prompt smoking diary; measure CO[c]	50	F2F	F2F
2	3	Measures for self-control	20	Web	F2F
3	5	Dealing with withdrawal	20	F2F	F2F
4	7	Breaking habits	20	Web	F2F
5	9	Dealing with triggers	20	F2F	F2F
6	11	Food for thought	20	Web	F2F
7	14	Think differently; measure CO	20	F2F	F2F
8	18	Do differently	20	Web	F2F
9	22	Action plan; measure CO	20	F2F	F2F
10	26	Closure	20	Web	F2F

[a]BSCT: blended smoking cessation treatment; total duration=230 minutes (130 minutes F2F mode, 100 minutes web mode).

[b]F2F: face-to-face treatment; total duration=230 minutes.

[c]CO: carbon monoxide.

Data Collection

Patient Characteristics

As part of the RCT (LiveSmokefree-study), 33 demographic, smoking-related, and health-related characteristics were assessed with the intake measurement using an online questionnaire. A detailed description of these characteristics is available in the protocol article of the RCT [22].

Measuring Adherence

Established measures for adherence to a blended treatment are still lacking. Therefore, for the 2018 study [10], we constructed a customized measure for adherence by selecting 18 patient activities (eg, using a web-based smoking diary tool, responding to counselors' messages) to trace adherence to the blended treatment. Adequacy of this adherence measure was confirmed by the observed dose-response relationship between adherence and the likelihood of quitting, which is consistent with the

XSL•FO
RenderX

smoking cessation literature [6-9]. However, this activity-based method was quite detailed and labor-intensive and was particularly interesting from a methodological point of view. Since the current study was mainly focused on comparing treatment modalities in a clinical context, we used a simpler time-based method for measuring adherence, which was proven to be as suitable for clinical research as the activity-based method and was also found to be more efficient [49,50]. Although this time-based method was not as accurate as the activity-based method, it was applicable in this case because the primary goal of this study was to determine differences between the groups in terms of levels and predictors of adherence. Therefore, the analysis of relative level differences was more relevant than an exact measurement of absolute levels. Furthermore, the time-based method allowed for analysis of a larger sample and thus more accurate statistics, as it required less time and money.

For this time-based approach, we used treatment data from the hospital's electronic patient record system. This record system contains basic information of the patients' treatment status such as when the patient started treatment; which counselor was offering the treatment; time, day, and type of appointments; time and day of telephone consults; and which kind of treatment was offered in each appointment [51]. In this record system, the counselors reported, in an encoded form, which type of sessions were completed. Each code represents a fixed, average number of minutes invested in face-to-face mode or web mode, as shown in Table 2. These fixed numbers of minutes per sessions were used to calculate the total number of minutes in treatment for each patient for the blended and face-to-face treatments, as well as for the face-to-face mode and web mode in the blended treatment.

Table 2. Codes, descriptions, modes of delivery, and duration of face-to-face (F2F) and blended treatment sessions used to measure adherence.

Code	Description of the session	Mode	Duration (minutes)
RSN	First individual F2F session at treatment start	F2F	50
RSAB	Like RSN, but visiting a patient at another department of the hospital	F2F	50
RSNS	Like RSN, but with 2 patients at the same time (eg, husband and wife)	F2F	35
RSC	Usual individual F2F session	F2F	20
RSAC	Additional consult (to add to RSN/RSAB/RSNS/RSC if more time is needed)	F2F	20
RSTC	Individual telephone consult	F2F	20
RSOC	Any other individual consult	F2F	10
RSIC	Web-mode treatment session via rokendebaas.nl	Web	20
RSEC	Email consulting	Web	10

Statistical Analysis

All analyses were performed using SPSS version 24.

Patient Characteristics

For both the blended and face-to-face groups, 33 demographic, smoking-related, and health-related characteristics were measured and are reported as means (SDs) for normally distributed continuous variables and as medians (IQRs) for nonnormally distributed continuous variables. Categorical variables are reported as numbers with corresponding percentages. To identify between-group differences within the 33 demographic, smoking-related, and health-related patient characteristics, independent *t* tests or Mann-Whitney *U* tests were performed as appropriate for continuous variables; the Pearson Chi-square or Fisher exact test was performed for categorical variables.

Adherence (Time Spent in Treatment)

Based on the hospital administrational records, both the absolute treatment time (in minutes) and the proportional treatment time (in percentage) of the patients who had started treatment were calculated for the blended and face-to-face groups, as well as for the face-to-face mode and web mode of the blended treatment. Bar charts were used to compare how many patients spend how much time in the blended and face-to-face treatment on the one hand and in each mode of the blended treatment on the other hand. Mann Whitney *U* tests were performed to compare the absolute treatment time of the blended and face-to-face treatments and the proportional treatment time for the face-to-face and web modes in the blended treatment.

Predictors of Adherence

To identify the predictors of adherence (as a continuous variable) within the 33 demographic, smoking-related, and health-related patient characteristics, Pearson or Spearman correlation tests were performed as appropriate for continuous variables; independent *t* tests or Mann-Whitney *U* tests were performed for dichotomous variables. Variables with significance at $P<.15$ were considered as candidates for multivariate linear regression analyses. They were first tested with univariate linear regression analyses so that univariate and multivariate odds ratios could be compared, and were entered in the multivariate linear regression analyses after checking for multicollinearity. The variables were either all entered and removed step by step via the backward selection method (all patients; blended group; face-to-face group; face-to-face mode of the blended treatment) or entered step by step via the forward selection method (web mode of the blended treatment). Variables were entered or eliminated step by step based on the model fit. In the case of multicollinearity, the variable with the best model fit was selected for linear analyses.

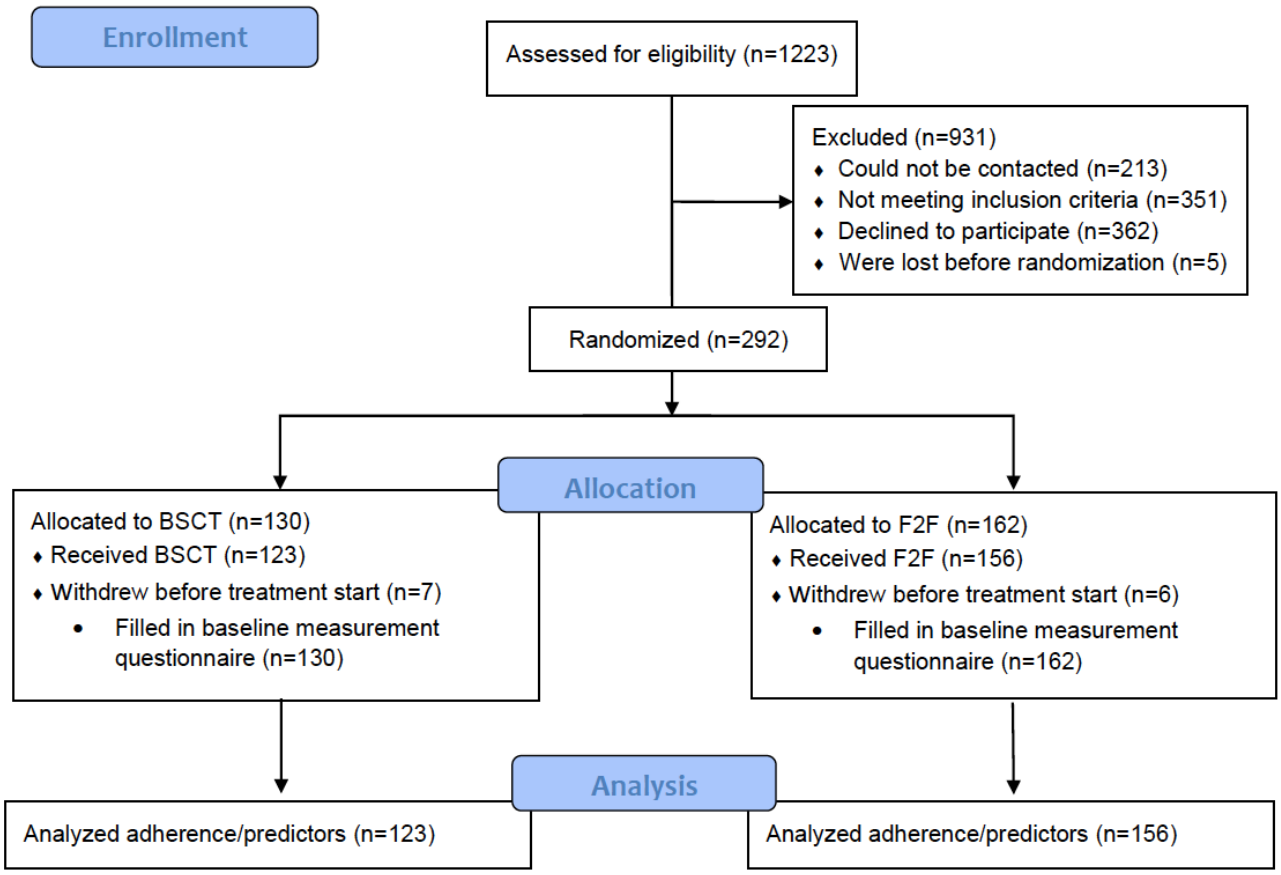
XSL•FO
RenderX

Results

Participant Flow

Figure 1 shows the flow of participants through the study. A total of 292 patients were eligible for the study, provided written consent, filled out the baseline questionnaire, and were randomized (blended n=130; face-to-face n=162). Before the start of treatment, 7/130 (5.4%) patients of the blended group and 6/162 (3.7%) patients of the face-to-face group withdrew. Finally, data from 123/130 (94.6%) patients in the blended group and 156/160 (96.3%) patients in the face-to-face group were available for adherence analysis.

Figure 1. Flow of participants through the study. BSCT: blended smoking cessation treatment; F2F: face-to-face.



Patient Characteristics

Table 3 shows the patients' characteristics for both the blended group (n=130) and the face-to-face group (n=162). Significant differences (P<.05) between the blended and face-to-face group were found for 6 of the 33 characteristics. Patients in the face-to-face group had higher internet skills, used more medication in general, reported less health complaints, and scored higher on the Depression Anxiety Stress Scale (DASS) subscales depression and anxiety, and on the total DASS score.

XSL•FO
RenderX

Table 3. Patients' characteristics of both the blended smoking cessation treatment (BSCT) and face-to-face (F2F) groups.

Characteristic	BSCT (n=130)	F2F (n=162)	*P* value
Demographic characteristics			
Sex (female), n (%)	62 (47.7)	77 (47.5)	.98
Age (years), mean (SD)	47.1 (12.8)	46.6 (13.2)	.76
Marital status, n (%)			.18
With partner	87 (66.9)	96 (59.3)	
Single	43 (33.1)	66 (40.7)	
Housing situation, n (%)			.91
Children	54 (41.5)	65 (40.9)	
No children	76 (58.5)	94 (59.1)	
Education, n (%)			.88
VET[a] or higher	82 (63.1)	101 (63.9)	
Lower than VET	48 (36.9)	57 (36.5)	
Main income, n (%)			.73
Wage or own company	64 (48.2)	83 (51.2)	
Income support	66 (50.8)	79 (48.8)	
Main day activity, n (%)			.72
Paid work	61 (46.9)	79 (49.1)	
Other	69 (53.1)	82 (50.9)	
Internet skills[b], mean (SD)	38.5 (5.64)	40.52 (8.63)	.01
Smoking-related characteristics			
Reason to start treatment, n (%)			.70
Intrinsic	83 (63.8)	107 (66.0)	
Extrinsic	47 (36.2)	55 (34.0)	
Nicotine dependency[c], mean (SD)	5.29 (2.10)	5.00 (2.18)	.59
Negative attitude toward quitting[d], mean (SD)	–5.70 (3.16)	–5.00 (2.96)	.07
Positive attitude toward quitting[e], median (IQR[c])	10 (8-12)	10 (8.75-11)	.91
Self-efficacy[f], mean (SD)	–0.37 (5.32)	–0.45 (5.02)	.89
Readiness to quit[g], median (IQR)	2 (1-3)	2 (1-3)	.31
Earlier quit attempts, n (%)	108 (83.1)	143 (88.3)	.20
Social support[h], median (IQR)	4 (3-5)	4 (3-5)	.99
Social modeling[i], median (IQR)	3.5 (1-6)	3 (1-5)	.13
Use of alcohol[j], n (%)	2 (1-3)	2 (0.75-3)	.26
Use of (recreational) drugs, n (%)	11 (8.5)	14 (8.7)	.94
Health-related characteristics			
Use of medication in general, n (%)	85 (65.4)	123 (75.9)	.05
Use of medication for addiction treatment, n (%)	0 (0.0)	0 (0.0)	N/A[k]
Use of medication for psychiatric treatment, n (%)	26 (20.0)	23 (15.1)	.28
Use of medication for physical treatment, n (%)	64 (49.2)	88 (57.9)	.15
Use of other medication, n (%)	19 (14.6)	31 (20.4)	.21

XSL•FO
RenderX

Characteristic	BSCT (n=130)	F2F (n=162)	*P* value
Health complaints (MAPHSS[l]), mean (SD)	12.58 (6.27)	10.96 (7.17)	.04
Smoking-related complaints[m], mean (SD)	20.82 (9.17)	19.95 (8.86)	.41
Health and smoking-related complaints[n], mean (SD)	33.56 (13.87)	30.91 (14.42)	.11
Depression[o], median (IQR)	4 (0-10)	4 (2-24)	.02
Anxiety[o], median (IQR)	4 (2-8)	6 (2-16.5)	.002
Stress[o], median (IQR)	8 (4-16)	10 (4-14)	.73
DASS[p], median (IQR)	18 (8-32)	22 (8-58.5)	.01
EQ-5D-3L[q], median (IQR)	0.77 (0.69-1.00)	0.77 (0.69-1.00)	.42
EQ VAS[r], mean (SD)	66.95 (16.88)	65.17 (17.56)	.38

[a]VET: vocational education and training.

[b]Scored on a scale of 10-60; a higher score indicates better skills.

[c]Fagerstroem scale (range 0-10); a higher score indicates higher nicotine dependency.

[d]Scored on a scale of –12 to 0; a lower score indicates a more negative attitude.

[e]Scored on a scale of –12 to 0; a higher score indicates a more positive attitude.

[f]Scored on a scale of –12 to 12; a higher score indicates higher self-efficacy.

[g]Scored on a scale of 0-4; a higher score indicates greater readiness to quit.

[h]Scored on a scale of 0-5; a higher score indicates more social support.

[i]Scored on a scale of 0-8; a higher score indicates more smokers in the social environment.

[j]Scored on a scale of 0-4; a higher score indicates higher alcohol consumption.

[k]N/A: not applicable; no statistical analysis performed since the variable is constant.

[l]MAPHSS: Maudsley Addiction Profile Health Symptoms Scale (range 0-40; a higher score indicates poorer health status).

[m]Scored on a scale of 0-64; a higher score indicates more smoking-related complaints.

[n]Scored on a scale of 0-104; a higher score indicates poorer health status and more smoking-related complaints.

[o]Scored on a scale of 0-42; a higher score indicates a higher level of depression/anxiety/stress.

[p]DASS: Depression Anxiety Stress Scale; sum of the Depression, Anxiety, and Stress subscale scores (range 0-126; a higher score indicates a more negative emotional status).

[q]EQ-5D-3L: societal-based quantification of health status (range 0-1; a higher score indicates better health status).

[r]EQ VAS: visual analog scale for quality of life (range 0-100; a higher score indicates better health status).

Adherence (Time Spent in Treatment)

As illustrated in Figure 2, adherence to the blended and face-to-face treatments was comparable. Patients in the blended group (n=123, 7 patients dropped out between inclusion and the first treatment session) spent a median of 246 (IQR 150-355) minutes in treatment (106.7% of the intended total treatment time); in the face-to-face group (n=156, 6 patients dropped out between inclusion and first treatment session), the patients spent a median of 238 (IQR 150-330) minutes in treatment (103.3% of the intended total treatment time). There was no significant difference between the two groups (P=.30). However, within the blended group, as shown in Figure 3, patients were more adherent to the face-to-face mode than to the web mode. Patients in the blended group (n=123) spent a mean of 198 (SD 120) minutes in face-to-face mode and 75 (SD 53) minutes in web mode. In proportion to the intended treatment time for each mode of delivery (face-to-face mode=130 minutes; web mode=100 minutes), patients in the blended group spent twice the time in face-to-face mode (mean 152%, SD 92% of 130 minutes) than in web mode (mean 75%, SD 53% of 100 minutes) (t_{122}=10.03; P<.001).

XSL•FO
RenderX

Figure 2. Adherence to blended smoking cessation treatment (BSCT) vs face-to-face (F2F) treatment.

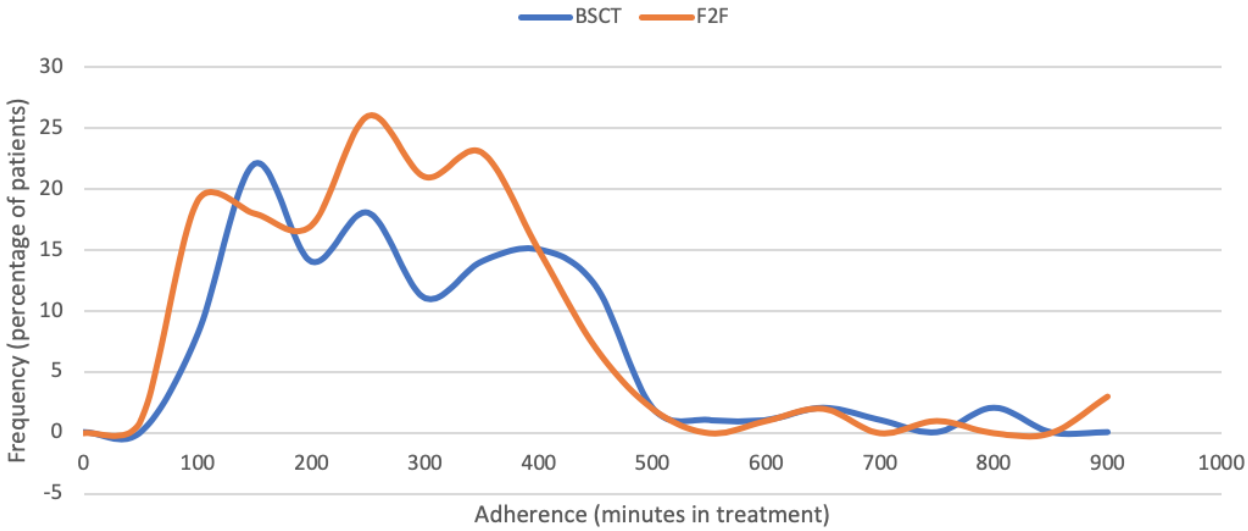
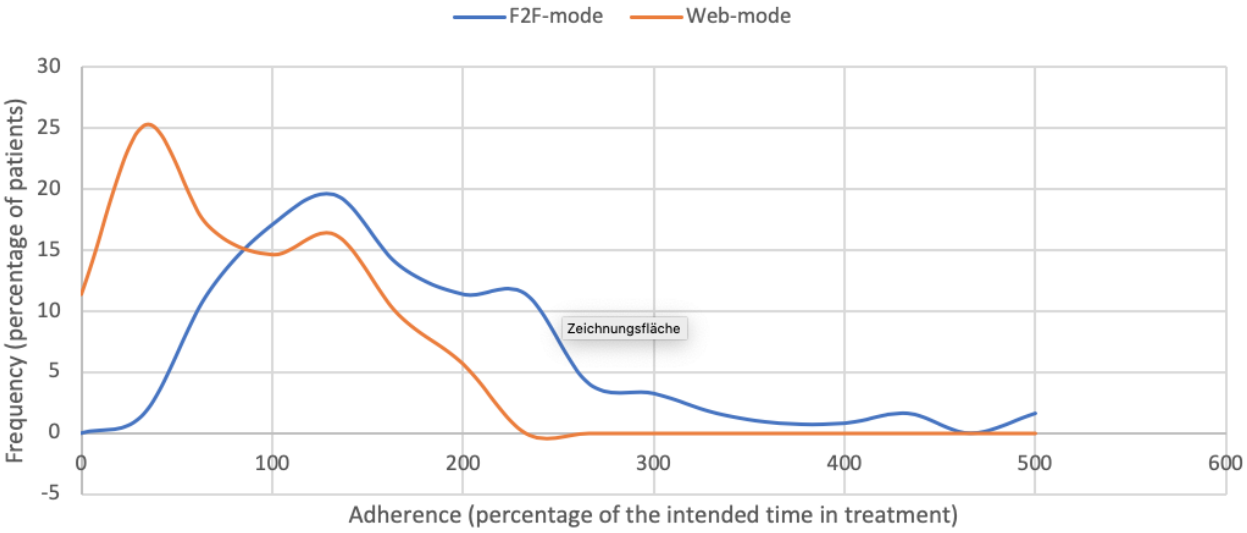


Figure 3. Adherence within the blended smoking cessation treatment group to the two modes of the treatment: face-to-face (F2F) mode vs web mode.



Predictors of Adherence

For both treatments together, 7 predictors (Table 4) were significantly associated with higher adherence in the univariate analysis (assessed at $P<.15$), including male sex, older age, housing situation (living without children), higher readiness to quit, higher social support, lower social modeling (less smokers in the social environment), and higher use of other medication. Multivariate regression analyses (Table 5) revealed that age was the best predictor of adherence ($R^2=0.047$). Per life year, patients spent 2.5 more minutes in treatment (95% CI 1.2-3.8; $P=.001$).

For the face-to-face group, 6 predictors (Table 4) were significantly associated with higher adherence in the univariate analyses (assessed at $P<.15$), including older age, higher readiness to quit, more social support, lower social modeling, higher use of other medication, and higher smoking-related complaints. Multivariate regression analyses (Table 5) revealed that age and social support together were the best predictors of adherence ($R^2=0.076$). Per life year, patients spent 2.2 minutes more in treatment (95% CI 0.4-3.9; $P=.02$). For social support, graded from 0 (low social smoking cessation support) to 5 (high social smoking cessation support), each unit increase was associated with 20.5 more minutes in treatment (95% CI 2.3-38.8; $P=.03$).

For the blended group, 8 predictors (Table 4) were significantly associated with higher adherence in the univariate analyses (assessed at $P<.15$), including higher age, housing situation (living without children), lower nicotine dependency (Fagerstroem), higher negative attitude toward quitting, lower social modeling, lower health complaints, lower anxiety, and lower stress. Multivariate regression analyses (Table 5) revealed that age was the best predictor of adherence ($P=.01$). Per life year, patients spent 2.6 more minutes in treatment (95% CI 0.5-4.6; $R^2=0.049$).

For the face-to-face mode of the blended treatment, 3 predictors (Table 6) were significantly associated with higher adherence in univariate analyses (assessed at $P<.15$), including higher age,

XSL•FO
RenderX

housing situation (living without children), and lower internet skills (Table 6). Multivariate regression analyses (Table 7) revealed that housing situation was the best predictor of face-to-face mode adherence (R^2=0.034). Patients living without children spent 49.7 more minutes in the face-to-face mode of the blended treatment (95% CI 92.7-6.8; *P*=.02) (Table 7).

For the web mode of the blended treatment, 16 predictors (Table 6) were significantly associated with higher adherence in univariate analyses (assessed at *P*<.15), including male sex, older age, main income (income support), main day activity (other than paid work), extrinsic reason to start treatment, lower nicotine dependency (Fagerstroem), higher negative attitude toward quitting, higher self-efficacy, lower social modeling, lower health complaints (assessed on the Maudsley Addiction Profile Health Symptoms Scale [MAPHSS]), lower smoking-related complaints, lower health and smoking-related complaints, lower anxiety, lower stress, lower DASS, and higher quality of life (EQ-5D-3L). Health and smoking-related complaints and the DASS were not used for multivariate regression because of multicollinearity. Multivariate regression analyses (Table 7) revealed that reason to start treatment, negative attitude toward quitting, and health complaints (MAPHSS) together were the best predictors of web mode adherence (R^2=0.164). Patients with an intrinsic motivation spent 21.5 less minutes in the web mode of the blended treatment (95% CI –39.8 to –3.3; *P*=.02). For negative attitude toward quitting (range –12 to 0; lower numbers indicate a more negative attitude toward quitting smoking), each unit increase (ie, a less negative attitude) was associated with 3.6 more minutes in web mode of the blended treatment (95% CI 0.9-6.4, *P*=.01). For health complaints (range 0-40; higher numbers indicate poorer health status), each unit increase (ie, additional complaint reported) was associated with 2.4 less minutes in the web mode of the blended treatment (95% CI –3.8 to –1.0, *P*=.001).

Table 4. Univariate predictors for adherence in all patients and in each treatment group.

Characteristic	All patients		F2F[a]		BSCT[b]	
	Regression coefficient (95% CI)	*P* value	Regression coefficient (95% CI)	*P* value	Regression coefficient (95% CI)	*P* value
Sex		.11		—[c]		—
Female (reference)	N/A[d]		—		—	
Male	28.6 (–6.4-63.6)		—		—	
Age (years)	2.5 (1.2-3.8)	.001	2.4 (0.7-4.2)	.01	2.6 (0.5-4.6)	.01
Housing situation		.13		—		.05
Children (reference)	N/A		—		—	
No children	27.8 (–8.0-63.6)		—		52.9 (105.6-0.4)	
Nicotine dependency	—	—	—	—	–10.9 (–23.3-1.4)	.08
Negative attitude toward quitting	—	—	—	—	6.6 (–1.6-14.7)	.11
Readiness to quit	16.0 (–1.5-33.6)	.07	21.1 (–3.3-45.5)	.09	—	—
Social support	13.6 (0.5-26.8)	.04	23.3 (4.5-41.7)	.01	—	—
Social modeling	–8.4 (–15.5 to –1.5)	.02	–10.1 (–20.0 to –0.1)	.05	–7.5 (–17.1-2.2)	.13
Use of other medication		.06		.06		—
Yes (reference)	N/A		N/A		—	
No	–36.8 (–75.0-1.3)		–53.2 (–108.1-1.7)		—	
Health complaints (MAPHSS[e])	—		—		–3.4 (–7.5-0.7)	.10
Smoking-related complaints	—		2.20 (–0.5-4.9)	.10	—	
Anxiety	—		—		–3.5 (–8.0-0.9)	.12
Stress	—		—		–2.8 (–6.2-0.6)	.11

[a]F2F: face-to-face treatment group.

[b]BSCT: blended smoking cessation treatment group.

[c]Data not shown, since for the sake of clarity only variables included in the multivariate regression at *P*<.15 are shown in the table.

[d]N/A: not applicable.

[e]MAPHSS: Maudsley Addiction Profile Health Symptoms Scale.

XSL•FO
RenderX

Table 5. Multivariate model of patient characteristics predicting adherence for all patients and each treatment group.

Variable	All patients		F2F[a]		BSCT[b]	
	Regression coefficient (95% CI)	*P* value	Regression coefficient (95% CI)	*P* value	Regression coefficient (95% CI)	*P* value
Age (years)	2.5 (1.2-3.8)	.001	2.2 (0.4-3.9)	.02	2.6 (0.5-4.6)	.01
Social support	__[c]	—	20.5 (2.3-38.8)	.03	—	—

[a]F2F: face-to-face treatment.

[b]BSCT: blended smoking cessation treatment.

[c]Data not shown, as for the sake of clarity only the variables of the final models are presented here.

XSL•FO
RenderX

Table 6. Univariate predictors for adherence to face-to-face (F2F) mode and web mode in the blended treatment group.

Variable	F2F mode		Web mode	
	Regression coefficient (95% CI)	*P* value	Regression Coefficient (95% CI)	*P* value
Sex				.12
Female (reference)	N/A[a]		N/A	
Male	—[b]		14.9 (33.8 to –3.9)	
Age	1.9 (0.2-3.6)	.03	0.8 (0.1-1.5)	.03
Housing situation		.02		—
Children (reference)	N/A		—	
No children	49.7 (6.8 - 92.7)		—	
Main income		—		.07
Wage or own company (reference)	—		N/A	
Income support	—		–17.2 (–36.0 to –1.5)	
Main day activity		—		.02
Paid work (reference)	—		Ref	
Other	—		–21.7 (–40.3 to –3.1)	
Internet skills	–3.0 (–6.8-0.8)	.12	—	—
Reason to start treatment		—		.09
Extrinsic (reference)	—		N/A	
Intrinsic	—		–16.9 (–36.4-2.5)	
Nicotine dependency	—	—	–6.1 (–10.5 to –1.8)	.01
Negative attitude towards quitting	—	—	3.8 (0.9-6.7)	.01
Self-efficacy	—	—	1.4 (–0.3-3.2)	.12
Social modeling	—	—	–3.1 (–6.6-0.4)	.08
Health complaints (MAPHSS[c])	—	—	–2.4 (–3.8 to –0.9)	.001
Smoking-related complaints	—	—	–1.0 (–2.0-0.1)	.06
Health and smoking-related complaints	—	—	–0.9 (–1.6 to –0.3)	.01
Anxiety	—	—	–2.2 (–3.7 to –0.6)	.01
Stress	—	—	–1.4 (–2.6 to –0.2)	.03
DASS[d]	—	—	–0.6 (–1.1 to –0.1)	.02
EQ-5D-3L[e]	—	—	28.9 (–8.1-65.8)	.13

[a]N/A: not applicable.

[b]Data not shown, since for the sake of clarity only variables that were included in the multivariate regression at *P*<.15 are shown in the table.

[c]MAPHASS: Maudsley Addiction Profile Health Symptoms Scale.

[d]DASS: Sum of Depression, Anxiety, and Stress scores.

[e]EQ-5D-3L: societal-based quantification of health status.

XSL•FO
RenderX

Table 7. Multivariate model of patient characteristics predicting adherence to face-to-face (F2F) and web mode in the blended treatment group.

Variable	F2F mode		Web mode	
	Regression coefficient (95% CI)	*P* value	Regression coefficient (95% CI)	*P* value
Housing situation		.02		—[a]
Children (reference)	N/A[b]		—	
No children	49.7 (6.8-92.7)		—	
Reason to start treatment		—		.02
Extrinsic (reference)	—		N/A	
Intrinsic	—		–21.5 (–39.8 to –3.3)	
Negative attitude toward quitting	—	—	3.6 (0.9-6.4)	.01
Health complaints (MAPHSS[c])	—	—	–2.4 (–3.8 to –1.0)	.001

[a]Data not shown, as for the sake of clarity only the variables of the final models are presented here.

[b]N/A: not applicable.

[c]MAPHSS: Maudsley Addiction Profile Health Symptoms Scale.

Discussion

Principal Findings

Since the emergence of web-based health promotion counseling a few decades ago, blended treatments have recently been introduced. The aim of the present study was to directly compare adherence to a blended treatment with a face-to-face treatment for smoking cessation with similar content.

Based on the treatment times documented in the hospital administration, we found comparable adherence levels for the blended and face-to-face treatments. However, within the blended treatment, we found that patients spent twice as much time in face-to-face mode (152% of the intended treatment time) than in web mode (75% of the intended treatment time), suggesting a tendency to substitute web sessions by additional face-to-face sessions.

Older age was the only characteristic consistently found to predict higher adherence to both the face-to-face and blended treatments. For the face-to-face group, we found that both older age and perceived social support for smoking cessation predicted higher adherence. Age is known as a relevant demographic characteristic for predicting adherence [34,35], but more social support to quit smoking has not yet been indicated as an independent predictor of adherence.

Within the blended treatment, no consistent predictor of adherence was found for its two modes of delivery. Higher adherence to the face-to-face mode was predicted by the housing situation (ie, living without children), whereas adherence to the web mode was predicted by an extrinsic motivation to quit, a less negative attitude toward quitting, and less health complaints. Although these models contained statistically significant predictive patient characteristics, the predicted proportion of variability in adherence was small, ranging from 3.4% to 16.4%. Thus, it seems immature to interpret these findings in an attempt to understand the mechanisms in adherence to blended smoking cessation treatment, and it is difficult to find a meaningful pattern in these predictors. To explain this low model fit, two aspects can be considered. First, this could indicate that the predictors examined in this study, namely only the patient characteristics, are not comprehensive. For example, it seems likely that provider-related variables and health care system factors such as treatment costs, failure to recall a receipt of a prescription, and access to free nicotine replacement therapy [6] also play a role. As no data on these factors were available in this study, this could not be further verified. Second, all patient-related predictors used in the current study were evaluated at the start of treatment, which means that changes in these characteristics during treatment (eg, due to negative treatment effects such as weight gain, adverse events, or withdrawal symptoms) were not considered. As an example of a positive treatment effect, in the context of smoking cessation treatment, the bidirectional relation between quitting success and adherence is known, in which early quitting success predicts higher adherence [34], while higher adherence predicts (long-term) abstinence [6-9]. Another example is the user experience that patients build during the course of treatment. Patients may experience the treatment as “useful,” “easy to follow,” or “stimulating” and adhere to the treatment accordingly [48,52].

In general, the finding for the blended group that treatment time not used in web mode was compensated by face-to-face mode treatment would support the expectation that in blended treatment, the strengths of one mode of delivery will compensate for the weaknesses of the other [4,18-23]. This expectation is also supported based on our recently published qualitative study on user experience with this blended smoking cessation treatment [48,52], in which we also found that the strengths of the face-to-face mode can compensate for the weaknesses of the web mode. It is noteworthy that this compensation is mainly unidirectional: face-to-face mode compensates or replaces web mode and not vice versa. By exceeding the planned face-to-face treatment time by 100 minutes on average, the vast majority of patients in the blended group (118/123, 95.9%) spent significantly more time in face-to-face mode than in web mode. By contrast, only 5/123 (4.1%) patients spent slightly more time (an additional 27 minutes on average) in web mode than in

XSL•FO
RenderX

face-to-face mode. Perhaps the new and challenging web mode is not used optimally, as it can (easily) be compensated by the traditional, familiar face-to-face mode.

Although the main objective of this study was to provide a treatment time–based comparison of adherence, we would like to briefly mention two aspects that surprised us when comparing the results with one of our previous studies [10] that used a different operationalization of adherence.

First, the current study revealed rather high adherence to both the blended and face-to-face treatments. Due to differences in interventions, measurements of adherence, adjunctive support, and investigated populations, adherence rates for smoking cessation treatment vary widely between different studies (5%-96%) [6]. This makes it difficult to compare adherence rates in general. Moreover, little is known about adherence rates for blended treatment. We only found one study that reported adherence rates: in a blended depression treatment, adherence to the blended treatment (90.5%) and the face-to-face treatment (95.1%) was comparable [53]. Our study seems to agree with this previous study, as we also found comparably high adherence to the blended (106.7%) and face-to-face (103.3%) treatments. Surprisingly, for the blended treatment, the findings in this study seem to contradict our findings from a previous study among participants of the blended treatment in the same sample [10], in which we reported that adherence to the blended treatment seemed rather low. These apparent contradictory results may be explained by different operationalizations of adherence and different measurement methods in the two studies. In the 2018 study, we traced treatment activities of the patients in detail (not only treatment time as used in the current study) and strived for a categorical threshold-based classification of patients as being either adherent or nonadherent. This activity-based method used in the 2018 study correlates with the time-based measurement applied in the current study, but it was more specific [49,50] and therefore resulted in lower absolute adherence rates [10].

Second, in our 2018 study [10], we found that in the blended treatment, based on patients' activities, there was no significant difference in adherence to the face-to-face mode compared with the web mode. Surprisingly, in the current study, based on treatment time, the adherence levels differed significantly. Patients spent only 75.2% of the intended treatment time in the web mode, but 152.3% of the intended treatment time in the face-to-face mode. This shows that in practice it is rather a 2/3 to 1/3 ratio between face-to-face mode and web mode in the blended group than the planned equal ratio. This could mean, for example, that patients in face-to-face mode need more time than planned for their activities, or that additional unplanned activities take place within the treatment time. This could be an indication of therapist drift—a known weakness of face-to-face treatment [30]—and thus bring the topic of treatment fidelity into focus. From a clinical point of view, the question then arises as to whether the planned times for face-to-face mode and web mode are appropriate.

Limitations and Implications for Future Work

To the best of our knowledge, this is the first study to compare adherence and predictors of adherence in patients randomly assigned to either a blended smoking cessation treatment or a face-to-face treatment with identical active components. Moreover, this is also the first study to compare adherence and predictors of adherence to the face-to-face mode with those to the web mode of a blended smoking cessation treatment. One limitation of this study is that the measurement of adherence was based on the treatment time documented in the hospital's administrational records, as this documentation is mainly used for financial accounting and therefore does not reflect in detail the contents and the exact temporal proportions of the treatments. Even though we assume that we have a sufficiently valid measure for the comparison of adherence and for the determination of predictors, these data unfortunately do not provide deeper insight into the adherence to the treatment process in detail. For example, the specific treatment activities carried out in different time frames remain unclear. In addition, the time data for the sessions were standard values and not exactly determined as treatment time per session. Individual sessions may therefore have been shorter or longer than evaluated. The absolute time values should therefore be interpreted with caution. Furthermore, in view of the differences between the results of this study and those of the 2018 study [10], in which different operationalizations of adherence were applied (time-based vs activity-based), the methodological question also arises as to which operationalization best reflects adherence. From our previous studies [49,50], we know that activity-based measurement has better predictive validity, which makes it seem more adequate when adherence is considered a determinant of efficacy (dose-response relationship). In this study, however, we used a time-based measurement because it requires less financial and time effort, the possibility of analyzing a larger sample size allowed us to expect more accurate statistics, and because we wanted to gain more experience with its application in clinical practice. The differences found in adherence between the 2018 study and the present study bring forth an interesting issue that deserves more attention and should be targeted in future studies, such as by addressing the research questions of this study using activity-based measurement to analyze the entire sample.

Another limitation is the low variability of adherence explained by our prediction models. The question arises as to whether the chosen predictors and their measurement are sufficient. Future research should further investigate which additional predictors (eg, provider behavior, health system factors, or other patient characteristics) should be included and how these can be measured, not only at the beginning but also over the course of treatment, so that they fit optimally to the research question.

Another point of interest should be the difference in predictors and levels of adherence to the two modes of delivery in the blended treatment. The characteristics associated with adherence are quite different (adherence to the face-to-face mode of the blended treatment was mostly associated with demographic characteristics, whereas adherence to the web mode of the blended treatment was mainly associated with smoking-related and health-related characteristics). Future research should examine the causation of these differences. For example, it is possible that web-mode treatment is better suited for patients with less health complaints because they rely less on a hospital setting and the direct contact to a health care professional such

XSL•FO
RenderX

as the smoking cessation counselor. Alternatively, web-mode treatment might be better suited for externally motivated patients because they already arrive at the treatment with a default desire to do what they are told and are therefore more likely to stick to the web mode.

Furthermore, the differences in the blended treatment adherence levels are noteworthy. In web mode, the adherence level was in the expected range, whereas there was overadherence found for the face-to-face mode. This could be related to the fact that the treatment basically starts with a rather long face-to-face session and therefore results in a type of face-to-face default mode. Therefore, it is possible that the result would have been different if the treatment had started in web mode. The overadherence raises another question as to which level of adherence is optimal to reach the treatment goals; that is, is higher adherence (152% adherence to face-to-face mode) better than lower adherence (ie, 75% adherence to web mode)?

Conclusion

This study represents one of the first attempts to thoroughly compare adherence and predictors of adherence of a blended smoking cessation treatment to a face-to-face treatment. Our results showed that the levels of adherence to both treatments were comparable. However, within the blended treatment, we found that adherence to face-to-face mode was significantly higher than that of web mode, although the intended total treatment time for the blended treatment was fairly broadly adhered to. This supports the idea that in blended treatment, one mode of delivery can compensate for the weaknesses of the other. Older age was found to be a common predictor of adherence to the treatments. However, within the blended treatment, adherence to each mode was predicted by different characteristics: adherence to the face-to-face mode was associated with demographic characteristics only, whereas adherence to the web mode of the blended treatment was also associated with several smoking-related and health-related characteristics. This may indicate that these characteristics should be taken into account when designing a blended treatment. However, the finding that only a small amount of the variance could be determined by the characteristics examined in this study suggests that provider-related health system factors and time-varying patient characteristics can also play an important role and should be explored in future research.

Acknowledgments

This research is funded by the Technology, Health & Care Research Group at Saxion University of Applied Sciences and the Centre for eHealth and Well-being Research at University of Twente. We would like to thank the directors and the personnel of Medisch Spectrum Twente Hospital and Tactus Addiction Treatment, and the patients of the outpatient smoking cessation clinic of Medisch Spectrum Twente Hospital.

Authors' Contributions

LS, MEP, MGJBK, and MGP identified the study questions, and designed the study and its measuring instruments. LS is the principal investigator and wrote the first draft of this manuscript. LS, MEP, MGJBK, MGP, RS, and SBA edited the manuscript. LS, MEP, MGJBK, MGP, and RS revised the manuscript. LS wrote the final version. All authors approved the final version of this manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the web sessions of the blended smoking cessation treatment.
[PDF File (Adobe PDF File), 1447 KB - jmir_v22i7e17207_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 250 KB - jmir_v22i7e17207_app2.pdf]

References

1. Tobacco Fact sheet. World Health Organization. 2017. URL: http://www.who.int/mediacentre/factsheets/fs339/en/ [accessed 2017-08-24]
2. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
3. Alvarez RC. The promise of e-Health - a Canadian perspective. eHealth Int 2002 Sep 17;1(1):4 [FREE Full text] [doi: 10.1186/1476-3591-1-4] [Medline: 12459044]
4. Erbe D, Eichert H, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]

XSL•FO
RenderX

5. Kemmeren LL, van Schaik DJF, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. BMC Psychiatry 2016 Apr 21;16:113 [FREE Full text] [doi: 10.1186/s12888-016-0818-5] [Medline: 27102812]
6. Sabaté E, editor. Adherence to long-term therapies: evidence for action. Geneva: World Health Organization; 2003.
7. Alterman AI, Gariti P, Cook TG, Cnaan A. Nicodermal patch adherence and its correlates. Drug Alcohol Depend 1999 Jan 07;53(2):159-165. [doi: 10.1016/s0376-8716(98)00124-0] [Medline: 10080041]
8. Westman EC, Behm FM, Simel DL, Rose JE. Smoking behavior on the first day of a quit attempt predicts long-term abstinence. Arch Intern Med 1997 Feb 10;157(3):335-340. [Medline: 9040302]
9. Fish LJ, Peterson BL, Namenek Brouwer RJ, Lyna P, Oncken CA, Swamy GK, et al. Adherence to nicotine replacement therapy among pregnant smokers. Nicotine Tob Res 2009 May;11(5):514-518 [FREE Full text] [doi: 10.1093/ntr/ntp032] [Medline: 19351783]
10. Siemer L, Brusse-Keizer MG, Postel MG, Ben Allouch S, Patrinopoulos Bougioukas A, Sanderman R, et al. Blended Smoking Cessation Treatment: Exploring Measurement, Levels, and Predictors of Adherence. J Med Internet Res 2018 Aug 01;20(8):e246 [FREE Full text] [doi: 10.2196/jmir.9969] [Medline: 30068503]
11. Bosworth H, editor. Improving Patient Treatment Adherence A Clinician's Guide. Heidelberg: Springer; 2010.
12. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016 Mar 24;3:CD008286. [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
13. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
14. Civljak M, Stead LF, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013 Jul 10(7):CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
15. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
16. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2009 Oct 07(4):CD006611. [doi: 10.1002/14651858.CD006611.pub2] [Medline: 19821377]
17. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
18. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
19. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
20. Kemmeren LL, van Schaik DJF, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. BMC Psychiatry 2016 Apr 21;16:113 [FREE Full text] [doi: 10.1186/s12888-016-0818-5] [Medline: 27102812]
21. Postel MG, Witting M, van Gemert-Pijnen JEWC. Blended behandeling in de geestelijke gezondheidszorg (Blended treatment in mental health care). Dth: kwartaaltijdschrift voor directieve therapie en hypnose 2013;33(3):210-221.
22. Siemer L, Pieterse ME, Brusse-Keizer MGJ, Postel MG, Ben Allouch S, Sanderman R. Study protocol for a non-inferiority trial of a blended smoking cessation treatment versus face-to-face treatment (LiveSmokefree-Study). BMC Public Health 2016 Nov 24;16(1):1187 [FREE Full text] [doi: 10.1186/s12889-016-3851-x] [Medline: 27881108]
23. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJEWC. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
24. Schuster R, Leitner I, Carlbring P, Laireiter A. Exploring blended group interventions for depression: Randomised controlled feasibility study of a blended computer- and multimedia-supported psychoeducational group intervention for adults with depressive symptoms. Internet Interv 2017 Jun;8:63-71 [FREE Full text] [doi: 10.1016/j.invent.2017.04.001] [Medline: 30135830]
25. Kooistra LC, Wiersma JE, Ruwaard J, van Oppen P, Smit F, Lokkerbol J, et al. Blended vs. face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial. BMC Psychiatry 2014 Oct 18;14:290 [FREE Full text] [doi: 10.1186/s12888-014-0290-z] [Medline: 25326035]
26. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. [An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences]. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]

XSL•FO
RenderX

27. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
28. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CLH, Burger H. Blended care vs. usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017 Jun 13;17(1):218 [FREE Full text] [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]
29. Harrington KF, McDougal JA, Pisu M, Zhang B, Sadasivam RS, Houston TK, CHART Collaborative Group. Web-based smoking cessation intervention that transitions from inpatient to outpatient: study protocol for a randomized controlled trial. Trials 2012 Aug 01;13:123 [FREE Full text] [doi: 10.1186/1745-6215-13-123] [Medline: 22852802]
30. Månsson KNT, Skagius Ruiz E, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an Internet-based support system for face-to-face cognitive behavior therapy: a proof of concept study. J Med Internet Res 2013 Dec 10;15(12):e280 [FREE Full text] [doi: 10.2196/jmir.3031] [Medline: 24326065]
31. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011 Jan 27;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
32. Hantsoo L, Epperson CN, Thase ME, Kim DR. Antepartum depression: treatment with computer-assisted cognitive-behavioral therapy. Am J Psychiatry 2013 Aug;170(8):929-930 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020254] [Medline: 23903340]
33. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen ML, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 07;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
34. Hays JT, Leischow SJ, Lawrence D, Lee TC. Adherence to treatment for tobacco dependence: association with smoking abstinence and predictors of adherence. Nicotine Tob Res 2010 Jun;12(6):574-581. [doi: 10.1093/ntr/ntq047] [Medline: 20457644]
35. Ben Taleb Z, Ward KD, Asfar T, Bahelah R, Maziak W. Predictors of adherence to pharmacological and behavioral treatment in a cessation trial among smokers in Aleppo, Syria. Drug Alcohol Depend 2015 Aug 01;153:167-172 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.05.038] [Medline: 26077603]
36. Mathew M, Morrow JR, Frierson GM, Bain TM. Assessing digital literacy in web-based physical activity surveillance: the WIN study. Am J Health Promot 2011;26(2):90-95. [doi: 10.4278/ajhp.091001-QUAN-320] [Medline: 22040389]
37. Gerhards SAH, Abma TA, Arntz A, de Graaf LE, Evers SMAA, Huibers MJH, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. J Affect Disord 2011 Mar;129(1-3):117-125. [doi: 10.1016/j.jad.2010.09.012] [Medline: 20889214]
38. Asfar T, Al Ali R, Rastam S, Maziak W, Ward KD. Behavioral cessation treatment of waterpipe smoking: The first pilot randomized controlled trial. Addict Behav 2014 Jun;39(6):1066-1074 [FREE Full text] [doi: 10.1016/j.addbeh.2014.02.012] [Medline: 24629480]
39. Heffner JL, Lewis DF, Winhusen TM. Preliminary evidence that adherence to counseling mediates the effects of pretreatment self-efficacy and motivation on outcome of a cessation attempt in smokers with ADHD. Nicotine Tob Res 2013 Feb;15(2):393-400 [FREE Full text] [doi: 10.1093/ntr/nts135] [Medline: 22949577]
40. Busnello RG, Melchior R, Faccin C, Vettori D, Petter J, Moreira LB, et al. Characteristics associated with the dropout of hypertensive patients followed up in an outpatient referral clinic. Arq Bras Cardiol 2001 May;76(5):349-354 [FREE Full text] [doi: 10.1590/s0066-782x2001000500001] [Medline: 11359183]
41. Saghaei M, Saghaei S. Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. JBiSE 2011;04(11):734-739. [doi: 10.4236/jbise.2011.411090]
42. van Deursen AJAM, Courtois C, van Dijk JAGM. Internet Skills, Sources of Support, and Benefiting From Internet Use. Int J Hum Comput Interact 2014 Mar 07;30(4):278-290. [doi: 10.1080/10447318.2013.858458]
43. Mudde AN, Willemsen NC, Kremers SPJ, de Vries H. Meetinstrumenten voor onderzoek naar roken en stoppen met roken. In: Den Haagtivoro. The Hague: Stivoro; 2000:9071263096-9789071263095.
44. Mudde AN, Willemsen MC, Kremers S, de Vries H. Meetinstrumenten voor onderzoek naar roken en stoppen met roken [Measuring instruments for smoking and smoking cessation research] 2nd edition. The Hague: Stivoro; 2006.
45. Richtlijn behandeling van tabaksverslaving, herziening 2009 Guidelines for treating tobacco addiction. Utrecht: Centraal BegeleidingsOrgaan; 2009. URL: https://www.nhg.org/sites/default/files/content/nhg_org/uploads/tabaksverslaving-herziening.pdf [accessed 2020-05-02]
46. Zorgmodule Stoppen met Roken. Den Haag: Partnership Stop met Roken,; 2009. URL: https://www.nhg.org/themas/publicaties/nhg-zorgmodule-leefstijl-roken-volledige-tekst [accessed 2020-05-02]
47. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

XSL•FO
RenderX

48. Siemer L, Ben Allouch S, Pieterse M, Brusse-Keizer M, Sanderman R, Postel M. Patients' User Experience of a Blended Face-to-Face and Web-Based Smoking Cessation Treatment: Qualitative Study. JMIR Form Res 2020 Jun 03;4(6):e14550 [FREE Full text] [doi: 10.2196/14550] [Medline: 32343245]
49. Siemer L, Brusse-Keizer M, Postel M, Ben AS, Patrinopoulos A, Sanderman R. How do we measure adherence to a blended smoking cessation treatment? 2019 Presented at: Supporting Health by Technology IX; 2019; Groningen. [doi: 10.1183/13993003.congress-2018.pa1227]
50. Siemer L, Brusse-Keizer M, Postel M, Ben AS, Patrinopoulos A, Sanderman R. Adherentie aan een blended stoppen-met-roken interventie: hoe meten we adherentie, hoe adherent zijn patienten en wat voorspelt adherentie? 2019 Presented at: NNvt Congres; 2019; Utrecht. [doi: 10.2196/preprints.9969]
51. Patrinopoulos BA. Adherence to blended smoking cessation treatment. Enschede: University of Twente; 2017.
52. Siemer L, Ben AS, Pieterse M, Brusse-Keizer M, Sanderman R, Postel M. Blended face-to-face and web-based smoking cessation treatment (BSCT): a description of patients? user experience (UX). Eur Resp J 2019;54(Suppl 63):PA1690. [doi: 10.2196/preprints.14550]
53. Wright JH, Wright AS, Albano AM, Basco MR, Goldsmith LJ, Raffield T, et al. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry 2005 Jun;162(6):1158-1164. [doi: 10.1176/appi.ajp.162.6.1158] [Medline: 15930065]

Abbreviations

DASS: Depression Anxiety Stress Scale
eHealth: electronic health
MAPHSS: Maudsley Addiction Profile Health Symptoms Scale
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 03.12.19; peer-reviewed by R Shan, S Wen; comments to author 27.01.20; revised version received 23.03.20; accepted 15.04.20; published 23.07.20.

Please cite as:
Siemer L, Brusse-Keizer MGJ, Postel MG, Ben Allouch S, Sanderman R, Pieterse ME
Adherence to Blended or Face-to-Face Smoking Cessation Treatment and Predictors of Adherence: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17207
URL: http://www.jmir.org/2020/7/e17207/
doi:10.2196/17207
PMID:32459643

©Lutz Siemer, Marjolein G J Brusse-Keizer, Marloes G Postel, Somaya Ben Allouch, Robbert Sanderman, Marcel E Pieterse. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Incentives on Adherence to a Web-Based Intervention Promoting Physical Activity: Naturalistic Study

Ramona Wurst[1], MA; Anja Maliezefski[2], BSc, PhD; Christina Ramsenthaler[1], MSc, PhD; Judith Brame[1], MSc; Reinhard Fuchs[1], PhD

[1]Department of Sport and Sport Science, University of Freiburg, Freiburg, Germany

[2]IGEL-M GbR, Bad Oldesloe, Germany

Corresponding Author:
Ramona Wurst, MA
Department of Sport and Sport Science
University of Freiburg
Schwarzwaldstrasse 175
Freiburg, 79117
Germany
Phone: 49 7612034563
Email: ramona.wurst@sport.uni-freiburg.de

Abstract

Background: Despite many advantages of web-based health behavior interventions such as wide accessibility or low costs, these interventions are often accompanied by high attrition rates, particularly in usage under real-life conditions. It would therefore be helpful to implement strategies such as the use of financial incentives to motivate program participation and increase adherence.

Objective: This naturalistic study examined real-life usage data of a 12-week web-based physical activity (PA) intervention (Fitness Coach) among insurants who participated in an additional incentive program (incentive group) and those who did not (nonincentive group). Users in the incentive group had the perspective of receiving €30 (about US $33) cash back at the end of the intervention.

Methods: Registration and real-life usage data as part of routine data management and evaluation of the Fitness Coach were analyzed between September 2016 and June 2018. Depending on the duration of use and the weekly recording of tasks, 4 adherence groups (low, occasional, strong, and complete adherence) were defined. Demographic characteristics were collected by a self-reported questionnaire at registration. We analyzed baseline predictors and moderators of complete adherence such as participation in the program, age, gender, and BMI using binary logistic regressions.

Results: A total of 18,613 eligible persons registered for the intervention. Of these, 15,482 users chose to participate in the incentive program (incentive group): mean age 42.4 (SD 14.4) years, mean BMI 24.5 (SD 4.0) kg/m^2, median (IQR) BMI 23.8 (21.7-26.4) kg/m^2; 65.12% (10,082/15,482) female; and 3131 users decided not to use the incentive program (nonincentive group): mean age 40.7 (SD 13.4) years, mean BMI 26.2 (SD 5.0) kg/m^2, median BMI 25.3 (IQR 22.6-28.7) kg/m^2; 72.18% (2260/3131) female. At the end of the intervention, participants in the incentive program group showed 4.8 times higher complete adherence rates than those in the nonincentive program group (39.2% vs 8.1%), also yielding significantly higher odds to complete the intervention (odds ratio [OR] 12.638) for the incentive program group. Gender significantly moderated the effect with men in the incentive group showing higher odds to be completely adherent than women overall and men in the nonincentive group (OR 1.761). Furthermore, older age and male gender were significant predictors of complete adherence for all participants, whereas BMI did not predict intervention completion.

Conclusions: This is the first naturalistic study in the field of web-based PA interventions that shows the potential of even small financial incentives to increase program adherence. Male users, in particular, seem to be strongly motivated by incentives to complete the intervention. Based on these findings, health care providers can use differentiated incentive systems to increase regular participation in web-based PA interventions.

(J Med Internet Res 2020;22(7):e18338) doi:10.2196/18338

XSL•FO
RenderX

KEYWORDS

internet-based intervention; adherence; incentive; reward; mHealth; eHealth; exercise; dropout rate; usage; attrition; telemedicine

Introduction

Promoting a healthy lifestyle, especially reducing sedentary behavior and increasing physical activity (PA), is crucial in the prevention of chronic diseases such as cardiovascular disease, type 2 diabetes, breast and colon cancer as well as hypertension [1]. Persons who follow the PA recommendations of the World Health Organization [2] show a 27% decreased risk of mortality [3]. Besides numerous face-to-face PA interventions [4,5], electronic health (eHealth) and mobile health (mHealth) interventions have become increasingly popular to promote health behavior change such as increasing PA levels [6]. These interventions have the potential to reach large numbers of people in a cost- and time-efficient way [6]. Testing the effectiveness of internet interventions for behavior change, in particular for promoting PA, has yielded positive results despite small effect sizes with standardized mean differences ranging from 0.14 to 0.20 for self-reported PA [7-10].

Based on this evidence, many stakeholders in health care including health insurers have developed web-based interventions promoting PA as a public health approach [11,12]. However, *program adherence* to web-based interventions, that is, using the program as intended by the developers over the course of the intervention, has always been a major challenge (law of attrition [13]). Despite the positive effects on PA described above, internet interventions suffer from high dropout rates of up to 80% at the end of the intervention [14-16]. This lack of program adherence is likely to negatively influence *behavior adherence* [17,18], that is, the maintenance of newly acquired PA behavior in the real life after the program is finished.

To increase program and behavior adherence of web-based interventions, many providers such as health insurers revert to financial incentives such as cash, bonus points, or charity rewards [12,19]. The rationale behind using such external rewards to change health behavior is based on the assumption that this strategy may trigger an increasingly intrinsic motivation for PA that stabilizes the behavior even after the incentive removal [19]. In *face-to-face* settings, financial incentives have been shown to increase real-life exercise adherence (behavior adherence) in the short term and partially even in the long term among children, adolescents, and adults [20,21]. Similar results have also been found for *web-based* interventions with increased PA levels reported by incentivized groups after the intervention [12,19,22-24]. It appears that those incentives do not need to be of high value. Small financial incentives of about US $1.50 per day or collecting loyalty points for coupons have been shown to substantially impact on uptake rates of eHealth or mHealth interventions [12,25,26].

Only a few studies using incentives for *web-based* PA interventions report *program adherence* rates for incentivized groups [12,23]. Results are equivocal: A web-based step count intervention for patients with ischemic heart disease lasting 24 weeks found a substantial difference between the incentivized group and the control group regarding program adherence (62.1% vs 51.2%) [23]. The incentive group was given US $14 per week and they could lose US $2 per day if they did not achieve their step goals. The results of this loss-framed incentive study point toward a beneficial effect of financial incentives to increase program adherence in web-based PA interventions. In another randomized controlled trial with insurants in Switzerland, differences between the incentivized and control groups were not significant after 3 months. However, program adherence rates were rather high in both groups with 92.1% in the incentive group (US $10 per month) and 90.5% in the control group [12]. Both groups were given discount to an activity tracker and they had to wear the tracker every day which probably also motivated the control group to participate regularly.

Data on program adherence rates of *app-based* PA interventions (daily step counts) are provided by the Carrot Rewards App Study (N=32,229) in which all users of the app collected loyalty points that could be redeemed for travel, groceries, or movies [22]. Results revealed that 61.9% of those who had initially registered as users completed the intervention after 12 weeks [22]. Unfortunately, there was no control group without incentives in the study, and therefore, their specific effect remains unclear here. From other app-based PA interventions, however, it is known that program adherence is markedly lower when no incentives are provided. For instance, in a naturalistic study by Guertler and colleagues [15], only 25% continued to use the PA promotion app after 6 weeks.

In sum, using incentives in web- or app-based PA interventions has the potential to increase program adherence in clinical trials. However, little is known about the effects of incentives on program adherence in real-life settings. All of the above reported results on program adherence were taken from studies in which users knew that they were part of a study and their data were tracked for study purposes. There is evidence that usage behavior differs substantially between people who use the programs on their own (ie, open-access users) and those who know that they are participants of clinical trials [6,27,28]. For example, Wanner and colleagues [27] showed that only 25.8% of the open-access users visited the PA internet intervention repeatedly compared with 67.3% of the trial participants [27]. This result reflects the high discrepancy of trial and real-life usage and points toward the need for naturalistic studies investigating the effects of incentives among open-access participants [6].

Besides real-life adherence rates, there is also no research on the factors that moderate program adherence in web- or app-based PA interventions using financial incentives. From studies without incentives, we know that age and gender may be significant predictors of PA program adherence. Thus, older age groups have been shown to participate more frequently than younger adults in web-based PA interventions with odds ratios (ORs) ranging between 1.02 and 2.61 [15,27,29,30]. Regarding gender, results are more inconsistent. Whereas some reviews

XSL•FO
RenderX

and studies revealed that women have a higher chance to complete an internet-based behavior change intervention (OR 2.24) [14,29], other investigations demonstrated that men use such interventions more continuously (OR 1.2; reduced hazard ratio to drop out 0.85) [15,27]. Studies that also included BMI in their analyses found that this parameter did not predict program adherence [27,29,31,32]. Given these age and gender differences in web-based program adherence, it is interesting to see how these differences are affected by the application of incentives.

This study's purpose was to expand upon previous research about incentives in web-based PA interventions by examining naturalistic data (ie, real-life usage data) of the largest health insurance company in Germany. Members interested in the web-based intervention *Fitness Coach* could choose to participate in an additional incentive program at the beginning of the intervention. Based on earlier studies [22], we hypothesized that real-life adherence would be higher in the incentive than in the nonincentive group. Furthermore, we considered the effects of moderators and predictors such as age, gender, and BMI on program adherence. This is the first study to examine the role of incentives for program adherence to a web-based PA intervention in a naturalistic, large-scale study of over 18,000 participants.

Methods

Study Design, Setting, and Participants

This investigation is a naturalistic study examining the real-life usage data of open-access users who participated in a web-based intervention called *Fitness Coach*. Naturalistic (observational) studies do not "involve any intervention [...] on the part of the investigator" [33] and the data are not affected by the actions of researchers [34]. Using such a naturalistic study design might help us better understand real-life program adherence to web-based PA interventions. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [35].

This study includes usage data from September 2016 to June 2018 of the Fitness Coach intervention provided by a large German health insurance company. The Fitness Coach intervention was freely accessible to everyone in Germany, but only insured persons could participate in the incentive program and noninsured persons could just use the intervention for at least four weeks. Most of the users in this study were members of the insurance company providing the intervention. In the first session of the Fitness Coach, they were screened for their suitability to exercise with the Physical Activity Readiness Questionnaire (PAR-Q +) [36]. Exclusion criteria were being younger than 18 years or being pregnant. If users met one or both exclusion criteria, they were advised to ask a physician about participating in the program.

All data entered by the open-access users including information about website use were collected and stored under a pseudonym by Vilua Healthcare GmbH (Berlin, Germany). Prior to the intervention, all users had to sign a data protection declaration in which they were informed that all data obtained could be further analyzed for research purposes. The Guidelines of the Declaration of Helsinki were followed. Given the retrospective examination of the real-life usage data, no ethical approval was obtained prior to data collection.

Intervention

The Fitness Coach is a 12-week long, tailored, and technically guided internet intervention promoting PA. It is based on the MoVo concept consisting of motivational and volitional strategies [37] as well as training progression concepts [38]. The intervention was developed in 2005 by an insurance company in Germany and was relaunched in September 2016 in an updated version [11]. All users of the intervention receive an individualized training program tailored to their personal goals and daily life. This consists of training videos that match the person's goals, that is, becoming stronger, improving one's endurance, or becoming more flexible (Multimedia Appendices 1 and 2). Users also receive psychological input to help them initiate and maintain their behavior change. The intervention contains multiple behavior change techniques such as goal setting, action planning, self-monitoring, and feedback as well as coping planning [39].

Incentive Program

As a form of primary prevention, health insurance companies in Germany are legally obliged to use incentives for promoting healthy behavior (eg, increasing PA and reducing weight). There is also evidence that investing approximately €30 (about US $33) per person and per year in incentives for health behavior change reduces the overall costs for the insurance company by approximately €100 (about US $112) per person and per year [40,41]. Based on these findings, all users of the web-based intervention who were insured by the insurance company could register for the incentive program at the beginning of the intervention. In order to receive the incentive, the users had to log the completion of at least three tasks (eg, strength, endurance, or mobility exercises; evaluating one's well-being; reading background information) per week in 10 of 12 weeks. In case of adherence to the intervention, they received 500 credit points on their health insurance account. Upon the collection of 500 further credit points within 1 year (earned via membership in sports clubs or gyms, regular medical check-ups, or participation at a sporting event), they could choose between €30 (about US $33) cash back and €60 (about US $68) health credit. This health credit could be used for additional medical and health services such as professional tooth cleaning, acupuncture, osteopathy, and fees for gyms or sports clubs.

Variables and Statistical Analyses

Demographic characteristics of all users including age, gender, and BMI were obtained from the self-report questionnaire included in the first session of the intervention. Each login to the intervention per day was captured, reported as mean (SD), resulting in a minimum number of 1 and a maximum number of 84 logins over the 12-week intervention span (16.9 [SD 16.8]). In order to categorize adherence, login data were summarized per week. If a user had at least one login per week with at least three logged tasks in the week, the user was adherent and participation in that week was defined as successful

XSL•FO
RenderX

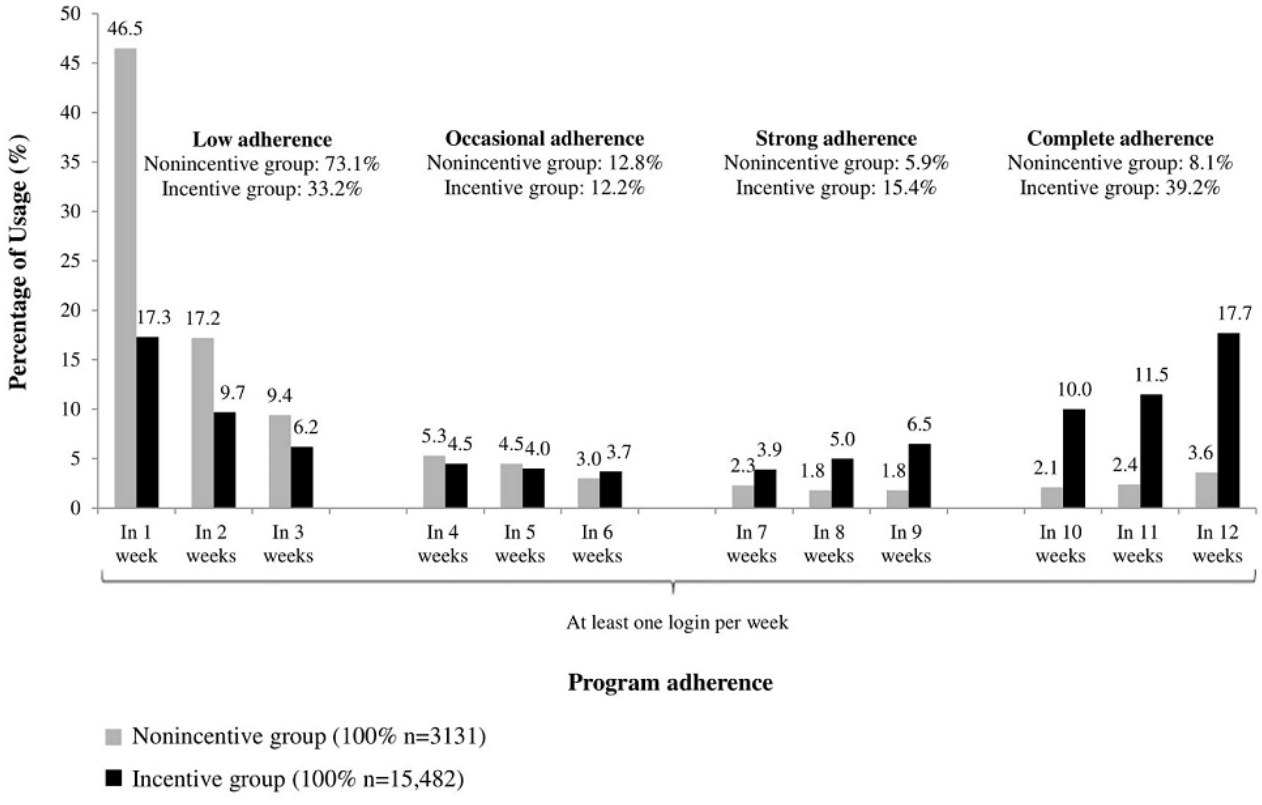
[42]. Based on the health insurance company's definition of successful participation (ie, ≥10 weeks), we classified the number of weeks with successful participation into 4 categories or adherence groups:

The *low adherence* group included all users participating for 1-3 weeks during the 12-week intervention span. *Occasional adherence* was defined as participation between 4 and 6 weeks. If users had 7-9 weeks of successful participation, we categorized them as *strongly adherent*. All persons completing 10 or more weeks successfully were defined as the *complete adherence* group (Figure 1). The last group was incentivized by the health insurance company.

Given the naturalistic sample, group sizes differ substantially in all 4 adherence groups as well as in the incentive and nonincentive groups (Table 1). We compared demographic variables between incentive and nonincentive groups using two-tailed Welch *t* tests and chi-square tests.

To investigate which characteristics would increase the chance of complete adherence and thus allowing the users to receive the incentive, binary logistic regressions were conducted using SPSS Statistics version 25 (IBM Corporation). We combined the first 3 adherence groups (low, occasional, and strong; =0) and compared them with the complete adherence group (=1) to best display the incentive effect on adherence. Incentive participation (0=no incentive, 1=incentive), gender (0=female, 1=male), and age and BMI (both continuous variables) were considered as baseline predictors. Furthermore, the interactions of incentive × age, incentive × gender, and incentive × BMI were considered as moderators for the association between incentives and program adherence. There were no missing data for adherence, incentive participation, age, and gender. Because it was not mandatory to enter height and weight, BMI data were missing in 530 cases, resulting in a total number of 18,083 users included in the logistic regression. After checking for multicollinearity between incentive participation, age, gender, and BMI (variable inflation factors between 1.024 and 1.066), all demographic variables and interactions were included in the regression model. Model fit was judged using Nagelkerke R^2. The level of significance was set at $P<.05$.

Figure 1. Percentage of Fitness Coach usage over 12 weeks divided into 4 adherence groups (N = 18,613).



XSL•FO
RenderX

Table 1. Demographic characteristics of the sample, grouped by adherence and incentive status.

Groups and characteristics	Nonincentive	Incentive
Overall		
N	3131	15,482
Age (years), mean (SD)	40.7 (13.4)	42.4 (14.4)
BMI (kg/m^2), mean (SD)	26.2 (5.0)	24.5 (4.0)
BMI (kg/m^2), median (IQR)	25.3 (22.6-28.7)	23.8 (21.7-26.4)
Gender		
Female, n (%)	2260 (72.18)	10,082 (65.12)
Male, n (%)	871 (27.82)	5400 (34.88)
Low adherence		
n (%)	2290 (73.14)	5133 (33.15)
Age (years), mean (SD)	39.9 (13.3)	40.2 (13.7)
BMI (kg/m^2), mean (SD)	26.2 (5.1)	24.9 (4.3)
BMI (kg/m^2), median (IQR)	25.1 (22.5-28.8)	24.0 (21.8-26.9)
Gender		
Female, n (%)	1639 (71.57)	3452 (67.25)
Male, n (%)	651 (28.43)	1681 (32.75)
Occasional adherence		
n (%)	401 (12.81)	1895 (12.24)
Age (years), mean (SD)	42.9 (12.9)	41.4 (13.8)
BMI (kg/m^2), mean (SD)	26.4 (4.8)	24.6 (4.1)
BMI (kg/m^2), median (IQR)	26.0 (22.7-29.0)	24.0 (21.7-26.7)
Gender		
Female, n (%)	288 (71.82)	1281 (67.60)
Male, n (%)	113 (28.17)	614 (32.40)
Strong adherence		
n (%)	187 (5.97)	2377 (15.35)
Age (years), mean (SD)	42.0 (13.0)	41.9 (14.0)
BMI (kg/m^2), mean (SD)	25.9 (4.8)	24.4 (3.8)
BMI (kg/m^2), median (IQR)	25.1 (22.4-28.3)	23.8 (21.7-26.2)
Gender		
Female, n (%)	139 (74.33)	1542 (64.87)
Male, n (%)	48 (25.66)	835 (35.13)
Complete adherence		
n (%)	253 (8.08)	6077 (39.25)
Age (years), mean (SD)	43.9 (14.2)	44.9 (14.3)
BMI (kg/m^2), mean (SD)	26.0 (4.9)	24.3 (3.8)
BMI (kg/m^2), median (IQR)	25.2 (22.6-28.4)	23.7 (21.6-26.0)
Gender		
Female, n (%)	194 (76.67)	3807 (62.65)
Male, n (%)	59 (23.32)	2270 (37.35)

XSL•FO
RenderX

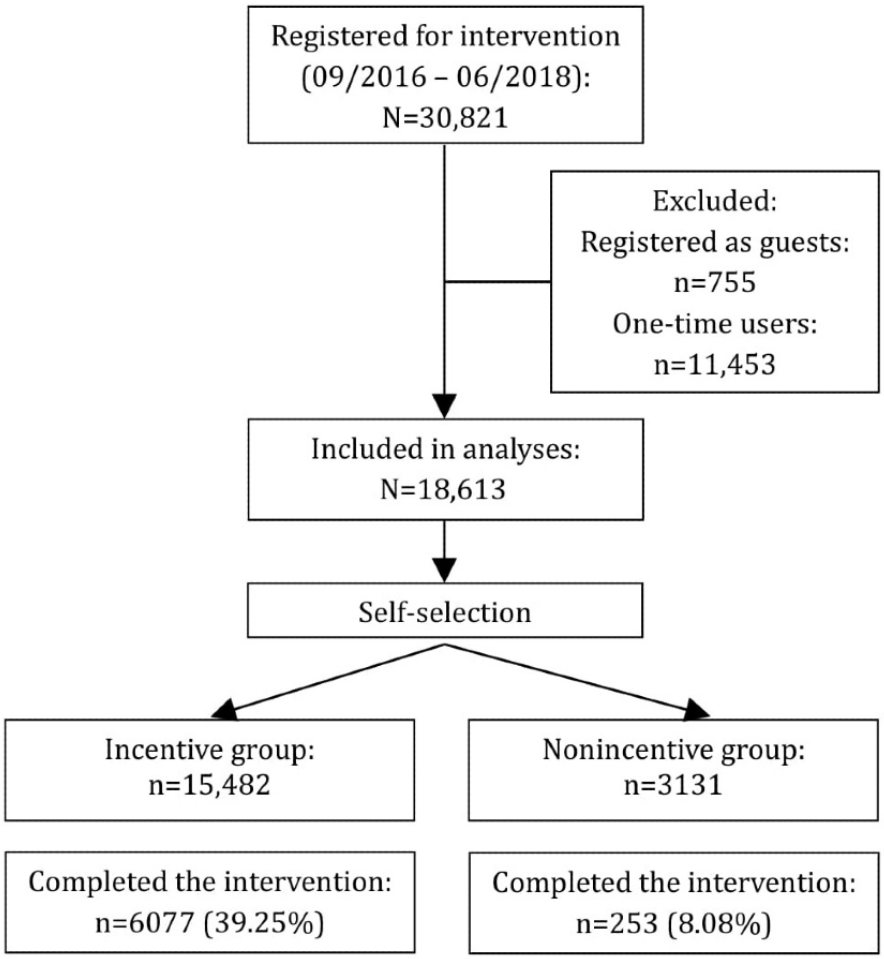
Results

Participants

From September 2016 to June 2018, a total of 30,821 persons registered for the Fitness Coach intervention. Of these, 755 (2.45%) registered as guests who were not insured by the health insurance company and were therefore unable to participate in the incentive program. In addition, there were 11,453 (37.16%) insured persons who visited the internet intervention only once and had no further login. All guests and one-time users were excluded from further analyses, so the demographic analyses included the total number of 18,613 insured users (Figure 2).

Demographic characteristics differed significantly between the incentive and nonincentive groups. Age (P<.001), BMI (P<.001), and gender (P<.001) differences indicated that older and more male users with a lower BMI register for the incentive (Table 1). Table 1 also displays the total sample size for each group at the start and at the end of the intervention (complete adherence).

Figure 2. STROBE flow chart of Fitness Coach participants.



Participation and Attrition Rates

In total, 34.01% of users (6330/18,613) were completely adherent (Figure 2). Among those complete adherers, the participation rate in the incentive group was 4.8 times higher than that in the nonincentive group: 39.25% (6077/15,482) of all members in the incentive group versus 8.08% (253/3131) of all members in the nonincentive group. By contrast, 73.14% (2290/3131) of members in the nonincentive group were classified as being lowly adherent, which means participating in less than 4 out of 12 weeks of the intervention, in comparison to 33.15% (5133/15,482) of members in the incentive group. The overall nonusage attrition rate (low, occasional, and strong adherence combined) in the nonincentive group reached 91.92% (2878/3131) compared with 60.75% (9405/15,482) in the incentive group (Table 1). The cutoff point of nonusage attrition was set at 9 or less successful intervention weeks.

Predictors of Intervention Completion

Binary logistic regressions revealed that participating in the incentive program was a statistically significant predictor (P<.001) at baseline for completing the intervention (OR 12.638, 95% CI 5.614-28.449; Table 2). There were also moderating effects on the association between incentive participation and

XSL•FO
RenderX

program adherence. Male users in the incentive group showed significantly higher odds to complete the intervention than female users overall and male users in the nonincentive group (OR 1.761, 95% CI 1.272-2.438). By contrast, the interactions between the incentive group and age or BMI did not yield significant odds for intervention completion.

Without taking the financial incentive participation into account (main effect), both age (P<.001) and gender (P=.017) proved to be significant baseline predictors of complete adherence to the Fitness Coach intervention for all users (Table 2), with older users having higher odds of Fitness Coach completion than younger ones (OR 1.023, 95% CI 1.013-1.033) and women having lower odds to be completely adherent than men (OR 0.679, 95% CI 0.494-0.933).

Table 2. Predictors of complete adherence (N=18,083).[a]

Characteristics	Logit odds	Standard error	Wald χ^2	*P* value	Odds ratio	95% CI for odds ratio	
						Lower	Upper
Intercept	–2.847	0.398	51.046	<.001	0.058		
Incentive program	2.537	0.414	37.545	<.001	12.638	5.614	28.449
Gender	–0.387	0.162	5.712	.017	0.679	0.494	0.933
Age	0.022	0.005	19.594	<.001	1.023	1.013	1.033
BMI	–0.017	0.014	1.503	.220	0.983	0.956	1.010
Gender × Incentive	0.566	0.166	11.608	.001	1.761	1.272	2.438
BMI × Incentive	–0.027	0.015	3.205	.073	0.974	0.946	1.003
Age × Incentive	–0.001	0.005	0.072	.788	0.999	0.988	1.009

[a]Reference category is low, occasional, and strong adherence groups combined; Nagelkerke R^2 (0.130), *df* (degrees of freedom) for all=1.

Discussion

Principal Findings

This large-scale, naturalistic study with a nationwide sample of 18,613 users examined the effect of an incentive program and its moderators on program adherence to a web-based PA intervention. Participants were members of the largest German health insurance who either decided to take part in an additional incentive program at the beginning of the intervention or rejected the use of the incentive program. Results showed that complete program adherence was 4.8 times higher in the incentive group than in the nonincentive group. In line with this finding, binary logistic regressions revealed incentive program participation to be a strong and significant predictor for complete adherence (OR 12.638, 95% CI 5.614-28.449, P<.001). This incentive effect on program adherence was moderated by gender, but not by BMI and age. Males in the incentive group showed the highest odds for program completion. Furthermore, gender and age also exhibited significant main effects on program adherence: being male and being of older age were associated with higher program adherence. However, BMI was not a significant predictor of program adherence.

Our findings demonstrate that even a small incentive (chance of €30 [about US $33] cash back at the end of the intervention) was able to enhance program adherence by almost five times, emphasizing the important motivating role of external rewards, at least in the initial phase of a new health behavior [19,39,43]. This result gains in importance because it is based on real-life usage data, which allows for a higher ecological validity. To the best of our knowledge, this is the first study investigating the effects of incentives on program adherence in a web-based PA intervention on the basis of naturalistic data. Therefore, no direct parallels can be drawn with data from the literature. However, comparisons can be made with similar real-life studies in which no incentives were used [27,28]. Those studies reported program adherence rates of 10%-25.8%. Compared with these findings, the program adherence rate of our nonincentive group (8.08%, 253/3131) is at the lower end of this range. By contrast, the program adherence rate of our incentive group (39.25%, 6077/15,482) substantially exceeded the program adherence rates of those two studies, suggesting once again that incentives might make a clear difference in the compliance with web-based PA interventions.

This study also found a *moderating effect* of gender on the association between incentive participation and program adherence. Men in the incentive group had significantly higher odds to complete the intervention than female users overall and male users in the nonincentive group, indicating that financial incentives may appeal particularly to men. This is supported by Czap and colleagues [44] who reported men to be more attracted to financial incentives than to other nonfinancial nudges. Given this clear gender difference in the motivational power of financial incentives, future research could investigate more closely which equivalent incentives might be more appealing to women.

By looking at the *main effects* of age on program adherence, we found older users to be significantly more inclined to complete the intervention than their younger counterparts, yet with a small OR of 1.023 (95% CI 1.013-1.033; P<.001). This result is well known from the literature, which indicates that older users have higher chances to be adherent than younger users [15,27,29,30]. It appears that on average users in their mid-40s are the most adherent ones to the web-based PA intervention. Regarding the main effect of gender on program adherence, there were more women than men participating in

XSL•FO
RenderX

the Fitness Coach, suggesting that improving one's fitness via internet appeals at first sight to women. This is in line with previous research reporting higher uptake rates of web-based PA interventions among women [6,8]. However, if we do not look at uptake rates but at adherence rates, the picture is quite different: women in our sample showed significantly lower odds to be completely adherent than men. There are several studies which support the finding that men are more adherent to these kinds of interventions [15,27]. For example, a large-scale, open-access evaluation from Switzerland yielded a significantly higher chance (OR 1.23) for men to participate repeatedly in a web-based intervention than for women [27].

The scope of this study was on *program adherence*; therefore, no definite conclusions can be drawn on the actual adherence to regular PA after the end of the 12-week intervention (behavior adherence). However, it is in the logic of time-limited behavior change programs that after finishing the program the newly acquired health behavior has to be continued in everyday life on one's initiative. We cannot be certain how successful the health behavior change has been in the present sample, yet there is evidence that a higher level of program adherence also leads to sustained health behavior change in everyday life [17].

Strengths and Limitations

This study is the first large-scale examination that investigated whether financial incentives increase program adherence to a web-based PA intervention under real-life usage conditions. Because of the naturalistic study design, the results of this investigation have high ecological validity.

However, some limitations need to be addressed. Internal validity might be threatened as the assessment of program adherence depends on self-reported participation behavior. There is also a risk that self-reported data on age and BMI may be distorted by socially desirable response tendencies or even worse, false data, to reveal as little information about oneself as possible. The study uses a naturalistic design with self-selection of participants into the incentive or nonincentive conditions. Given this design, it was not possible to randomize users into these two groups. Overall, both groups differed substantially regarding age, gender, and BMI with older users, more men, and users with lower BMI choosing to participate in the incentive program. It might be hypothesized that participants in the incentive group were more motivated from the beginning to use the program than users who did not register for the incentive.

Given the unequal sample sizes for incentive and nonincentive groups (15,482 vs 3131) as well as for the low, occasional, and strong adherence groups combined and the complete adherence group (12,283 vs 6330), the results of binary logistic regressions might be overestimated. This is reflected, for example, in very large CIs and statistical significance despite small OR (eg, OR_{Age} 1.023, 95% CI 1.013-1.033; $P<.001$). However, the decision for the large comparison group was made from a content perspective focusing on representativeness and external validity. The unbalanced analyses took account of the naturalistic data and better displayed how complete adherers differ from the other users who did not successfully finish the intervention. The logistic regression was also based on large sample size. While the unbalanced distribution might have affected the width of the confidence intervals, it has not affected the ORs in the logistic regression because there is enough power in this analysis.

Because only age, gender, and BMI were assessed at registration, we cannot exclude that other factors which were not investigated, such as education or socioeconomic status, may have influenced the results. The user's socioeconomic status, in particular, might be of great interest for incentive participation and adherence as users with lower income are more likely to be appealed by the incentives [19] and are more adherent to interventions changing health behavior [12,45].

Conclusions

This examination points out that financial incentives in the form of cashback or health credits have the potential to increase program adherence in a web-based PA intervention. Incentives are likely to act as an external motivator that enhances program adherence, which in turn can be the basis for a more intrinsically motivated, longer-lasting PA behavior in real life after the program has ended (behavior adherence). Therefore, health care providers may encourage program participation by providing differentiated incentives depending on gender, age, or BMI in web-based interventions promoting PA. Further investigations are needed to support these findings; in particular, it is necessary to further clarify the moderating role not only of gender, age, or BMI, but also of other potentially relevant factors such as socioeconomic status or education. The study results are based on naturalistic data with a high ecological validity; in the next step, it is necessary to confirm these findings in more controlled studies (eg, randomized controlled trials) with higher levels of internal validity.

Acknowledgments

The project was funded by Techniker Krankenkasse (German Health Insurance Company) and the article processing charge was funded by the University of Freiburg in the funding program Open Access Publishing. The project funder had no role in study design, management, analysis and interpretation of data, writing of the report, and the decision to submit the report for publication. We thank the Techniker Krankenkasse, in particular Dr Nicole Knaack, Dagmar Koeppel, and Dr Kerstin Hofreuter-Gaetgens for supporting this project. Besides, we thank Dr Boris Gauss, Susanne Dworschak, and Valeska Zarwell of Vilua Healthcare GmbH for providing the data and for their technical support. Finally, a big thank you goes to Lisa Sophia Schmiedeshoff for her excellent work on data preparation and data management.

XSL•FO
RenderX

Authors' Contributions

RW, AM, CR, and RF designed the study; RW, AM, CR, JB, and RF analyzed and interpreted the data; RW drafted the manuscript; AM, CR, JB, and RF revised the manuscript.

Conflicts of Interest

RW, AM, and JB report grants for the development and evaluation of the Fitness Coach by Techniker Krankenkasse. CR and RF do not report conflicts of interest.

Multimedia Appendix 1
Dashboard Fitness Coach mobile.
[PNG File , 240 KB - jmir_v22i7e18338_app1.png]

Multimedia Appendix 2
Knowledge Section Fitness Coach.
[PNG File , 590 KB - jmir_v22i7e18338_app2.png]

References

1. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
2. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; 2010.
3. Leitzmann MF, Park Y, Blair A, Ballard-Barbash R, Mouw T, Hollenbeck AR, et al. Physical activity recommendations and decreased risk of mortality. Arch Intern Med 2007 Dec 10;167(22):2453-2460. [doi: 10.1001/archinte.167.22.2453] [Medline: 18071167]
4. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. Soc Sci Med 2017 Nov;192:125-133. [doi: 10.1016/j.socscimed.2017.09.021] [Medline: 28965003]
5. Carr RM, Prestwich A, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Quested E, et al. Dyadic interventions to promote physical activity and reduce sedentary behaviour: systematic review and meta-analysis. Health Psychol Rev 2019 Mar;13(1):91-109. [doi: 10.1080/17437199.2018.1532312] [Medline: 30284501]
6. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
7. Broekhuizen K, Kroeze W, van Poppel MNM, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
8. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
9. Joseph RP, Durant NH, Benitez TJ, Pekmezi DW. Internet-Based Physical Activity Interventions. Am J Lifestyle Med 2014 Dec;8(1):42-68 [FREE Full text] [doi: 10.1177/1559827613498059] [Medline: 25045343]
10. Foster C, Richards J, Thorogood M, Hillsdon M. Remote and web 2.0 interventions for promoting physical activity. Cochrane Database Syst Rev 2013 Sep 30;9:CD010395 [FREE Full text] [doi: 10.1002/14651858.CD010395.pub2] [Medline: 24085594]
11. Knaack N, Gauss B, Blieske P. Mit digitalen Gesundheitszielen in ein gesundes Leben starten. In: Duesberg F, editor. e-Health 2017: Informations- und Kommunikationstechnologien im Gesundheitswesen. Solingen: Medical Future Verlag; 2017:133-135.
12. Kramer J, Tinschert P, Scholz U, Fleisch E, Kowatsch T. A Cluster-Randomized Trial on Small Incentives to Promote Physical Activity. Am J Prev Med 2019 Feb;56(2):e45-e54 [FREE Full text] [doi: 10.1016/j.amepre.2018.09.018] [Medline: 30553693]
13. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
14. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
15. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and Nonusage Attrition With a Free Physical Activity Promotion Program: The Case of 10,000 Steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]

XSL•FO
RenderX

16. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]
17. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
18. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions: a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
19. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2019 May 15:100633. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
20. Strohacker K, Galarraga O, Williams DM. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med 2014 Aug;48(1):92-99 [FREE Full text] [doi: 10.1007/s12160-013-9577-4] [Medline: 24307474]
21. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
22. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the Carrot Rewards App, a Population-Level Incentive-Based Intervention Promoting Step Counts Across Two Canadian Provinces: Quasi-Experimental Study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
23. Chokshi NP, Adusumalli S, Small DS, Morris A, Feingold J, Ha YP, et al. Loss-Framed Financial Incentives and Personalized Goal-Setting to Increase Physical Activity Among Ischemic Heart Disease Patients Using Wearable Devices: The ACTIVE REWARD Randomized Trial. J Am Heart Assoc 2018 Jun 13;7(12):e009173 [FREE Full text] [doi: 10.1161/JAHA.118.009173] [Medline: 29899015]
24. Herman CW, Musich S, Lu C, Sill S, Young JM, Edington DW. Effectiveness of an incentive-based online physical activity intervention on employee health status. J Occup Environ Med 2006 Sep;48(9):889-895. [doi: 10.1097/01.jom.0000232526.27103.71] [Medline: 16966955]
25. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an Incentive-Based mHealth App: Process Evaluation of the Carrot Rewards App. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
26. Liu S, Hodgson C, Zbib AM, Payne AYM, Nolan RP. The effectiveness of loyalty rewards to promote the use of an Internet-based heart health program. J Med Internet Res 2014 Jul 02;16(7):e163 [FREE Full text] [doi: 10.2196/jmir.3458] [Medline: 24989982]
27. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
28. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
29. van der Mispel C, Poppe L, Crombez G, Verloigne M, de Bourdeaudhuij I. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
30. Kelders SM, van Gemert-Pijnen JEWC, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
31. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
32. Gomez Quiñonez S, Walthouwer MJL, Schulz DN, de Vries H. mHealth or eHealth? Efficacy, Use, and Appreciation of a Web-Based Computer-Tailored Physical Activity Intervention for Dutch Adults: A Randomized Controlled Trial. J Med Internet Res 2016 Nov 09;18(11):e278 [FREE Full text] [doi: 10.2196/jmir.6171] [Medline: 27829576]
33. Porta M, editor. A Dictionary of Epidemiology. Oxford: Oxford University Press; 2014.
34. Given LM, editor. The Sage Encyclopedia of Qualitative Research Methods. Los Angeles, CA: Sage Publications; 2008.
35. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
36. Warburton DER, Jamnik VK, Bredin SSD, Gledhill N. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). The Health & Fitness Journal of Canada 2011;4(2):3-17 [FREE Full text] [doi: 10.14288/hfjc.v4i2.103]

XSL•FO
RenderX

37. Fuchs R, Goehner W, Seelig H. Long-term effects of a psychological group intervention on physical exercise and health: the MoVo concept. J Phys Act Health 2011 Aug;8(6):794-803. [doi: 10.1123/jpah.8.6.794] [Medline: 21832294]
38. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
39. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
40. Stock S, Schmidt H, Büscher G, Gerber A, Drabik A, Graf C, et al. Financial incentives in the German Statutory Health Insurance: new findings, new questions. Health Policy 2010 Jun;96(1):51-56. [doi: 10.1016/j.healthpol.2009.12.015] [Medline: 20106543]
41. Stock S, Stollenwerk B, Klever-Deichert G, Redaelli M, Büscher G, Graf C, et al. Preliminary analysis of short-term financial implications of a prevention bonus program: first results from the German Statutory Health Insurance. Int J Public Health 2008;53(2):78-86. [doi: 10.1007/s00038-008-7026-0] [Medline: 18681336]
42. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
43. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
44. Czap N, Czap H, Khachaturyan M, Burbach M. Comparing Female and Male Response to Financial Incentives and Empathy Nudging in an Environmental Context. RBE 2018 Mar 26;5(1):61-84 [FREE Full text] [doi: 10.1561/105.00000079]
45. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JPT, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. Prev Med 2015 Jun;75:75-85 [FREE Full text] [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]

Abbreviations

OR: odds ratio
PA: physical activity

Edited by G Eysenbach; submitted 20.02.20; peer-reviewed by T Muto, SGS Shah, A De Ramón Fernández, K Alipasandi; comments to author 30.03.20; revised version received 02.05.20; accepted 13.05.20; published 30.07.20.

Please cite as:
Wurst R, Maliezefski A, Ramsenthaler C, Brame J, Fuchs R
Effects of Incentives on Adherence to a Web-Based Intervention Promoting Physical Activity: Naturalistic Study
J Med Internet Res 2020;22(7):e18338
URL: http://www.jmir.org/2020/7/e18338/
doi:10.2196/18338
PMID:32729835

©Ramona Wurst, Anja Maliezefski, Christina Ramsenthaler, Judith Brame, Reinhard Fuchs. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Synthesizing Multiple Stakeholder Perspectives on Using Virtual Reality to Improve the Periprocedural Experience in Children and Adolescents: Survey Study

Naseem Ahmadpour[1], BSc, MSc, PhD; Andrew David Weatherall[2,3], BMed; Minal Menezes[2], BSc, MSc, PhD; Soojeong Yoo[1], PhD; Hanyang Hong[4], MD; Gail Wong[3], BSc (Med), MBBS

[1]Design Lab, School of Architecture, Design and Planning, The University of Sydney, Darlington, NSW, Australia

[2]Division of Child and Adolescent Health, The University of Sydney, Sydney, Australia

[3]Department of Anaesthesia, The Children's Hospital at Westmead, Westmead, Australia

[4]Sydney Medical School, The University of Sydney, Sydney, Australia

Corresponding Author:
Naseem Ahmadpour, BSc, MSc, PhD
Design Lab
School of Architecture, Design and Planning
The University of Sydney
148 City Road
Darlington, NSW, 2006
Australia
Phone: 61 286275649
Email: naseem.ahmadpour@sydney.edu.au

Abstract

Background: Virtual reality (VR) technology is a powerful tool for augmenting patient experience in pediatric settings. Incorporating the needs and values of stakeholders in the design of VR apps in health care can contribute to better outcomes and meaningful experiences for patients.

Objective: We used a multiperspective approach to investigate how VR apps can be designed to improve the periprocedural experiences of children and adolescents, particularly those with severe anxiety.

Methods: This study included a focus group (n=4) and a survey (n=56) of clinicians. Semistructured interviews were conducted with children and adolescents in an immunization clinic (n=3) and perioperative setting (n=65) and with parents and carers in an immunization clinic (n=3) and perioperative setting (n=35).

Results: Qualitative data were examined to determine the experience and psychological needs and intervention and design strategies that may contribute to better experiences for children in three age groups (4-7, 8-11, and 12-17 years). Quantitative data were used to identify areas of priority for future VR interventions.

Conclusions: We propose a set of ten design considerations for the creation of future VR experiences for pediatric patients. Enhancing patient experience may be achieved by combining multiple VR solutions through a holistic approach considering the roles of clinicians and carers and the temporality of the patient's experience. These situations require personalized solutions to fulfill the needs of pediatric patients before and during the medical procedure. In particular, communication should be placed at the center of preprocedure solutions, while emotional goals can be embedded into a procedure-focused VR app to help patients shift their focus in a meaningful way to build skills to manage their anxiety.

(J Med Internet Res 2020;22(7):e19752) doi:10.2196/19752

KEYWORDS

virtual reality; periprocedural anxiety; children; adolescents; stakeholder perspective; design; VR; pediatrics; patient experience; app; eHealth

XSL•FO
RenderX

Introduction

A fundamental role of clinicians looking after children having medical procedures in hospitals is to try to minimize any distress associated with that procedure. A negative experience of a medical procedure has implications more extensive than the immediate distress of that event. The imprint will be left on the child for all subsequent medical procedures. A range of approaches may be attempted as a periprocedural intervention (ie, intervention prior, during, or after the medical procedure) to mitigate such negative experiences, especially in the context of significant anxiety associated with needles in children requiring vaccinations [1]. In-hospital sedation services facilitate vaccination in patients who are otherwise unable to complete their immunization schedule due to severe avoidance behavior. Improved immunization experiences for patients, their caretakers, and staff may reduce the need for pharmacological intervention in the form of sedation, which may be difficult to administer or sometimes harmful [2]. Other approaches, such as exposure therapy, require long-term time commitments and significant resources [3]. In this paper, we present an exploratory study into ways through which virtual reality (VR) might best be employed to enhance the pediatric patient experience.

Technology-mediated solutions for focus-shifting or distraction positively alter the periprocedural experiences of children, to reduce anxiety and decrease pain response [4]. Current evidence suggests efficacy for technologies such as hand-held video games [5], tablet computers [6], immersive 360 videos [7], and interactive VR games [8]. VR technology has been explicitly explored as an anxiety management intervention in the context of vaccination [9] and acute procedures such as laceration repair and wound care [10]. With recent advances in commercial VR Head-Mounted Display (HMD) devices (eg, Oculus Quest by Oculus VR), the availability of this technology has dramatically increased. Recent market research [11,12] found that of 1917 children (aged <15 years) surveyed in the United States of America, only 19% were unaware of VR technology in the Spring of 2017. This mainstream awareness of VR technology presents an opportunity that we can leverage to develop effective interventions to support pediatric treatment.

The efficacy of VR apps is often attributed to the distraction created by the immersive environment, which creates an illusion of presence in the virtual world [13] and therefore reduces the user's cognitive resources to attend to distressing stimuli. Distraction in VR may be achieved in two ways: passive and active [14]. Passive engagement can be in the form of viewing an immersive 360-degree video without guiding the user's attention. Active distraction engages the user in gameplay or cognitive tasks in order to shift their focus to virtual objects and away from the negative stimuli [15]. Distraction can be more effective if patients are engaged in emotionally relevant VR experiences [16] as it enables them to dissociate their virtual and physical bodies [17].

Current examples of VR for managing pediatric anxiety do not exploit the multitude of strategies possible in VR, particularly active distraction. Additionally, there is a lack of knowledge about the needs of pediatric patients as well as the role of other stakeholders involved in patient care, including clinicians and parents, that can be incorporated into future pediatric VR solutions [18]. This study aimed to investigate opportunities for designing VR apps for children and adolescents in order to improve their experience and help them manage anxiety in the context of perioperative care and immunization.

To achieve this aim, we followed a unique multiperspective and qualitative approach. We collected in-depth data from multiple stakeholders to explore factors that can inform the design of future VR solutions. We conducted a focus group and surveyed clinicians to collect their views on typical strategies for managing pediatric anxiety in periprocedural settings. To understand patient responses to current anxiety management strategies, we conducted semistructured interviews with children and their carers.

Methods

This study included a focus group and survey of clinicians, in addition to interviews with children and parents using protocols approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (HREC Reference Number–LNR/18/SCHN/160 HREC/17/SCHN/429). All study participants were recruited from The Children's Hospital at Westmead, NSW in Australia and provided written informed consent before participation. Focus groups and semistructured interview sessions were audio-recorded.

Clinicians

Focus Group

Four health professionals were recruited to participate in a 90-minute focus group convened by two researchers; one anesthetist and one human-computer interaction (HCI) researcher. The aim was to gain a better understanding of the strategies used by clinicians to manage periprocedural anxiety in their pediatric patients. The session started with a brief explanation of VR technology, followed by questions that guided the discussion, like "Do you use any particular strategies to shift the focus of children during the procedures?" and "How do you choose which approach will work for which patient?"

Survey

A total of 56 clinicians were recruited by email and direct approach to complete a survey, repeated three times, and focusing on children in three age groups: 4-7, 8-11, and 12-17 years old. The questionnaire was developed by researchers with a background in clinical anesthesia based on their experience. Participants rated the likelihood of 13 items to induce anxiety in children and adolescents in the preoperative, operative, and postoperative phases (5-point scales, ranging from "not at all" to "very much"). Participants were then asked to describe strategies based on their experience that made the patient feel good, calm, or reassured before going in for an operation.

Children and Adolescents

Semistructured Interview in the Immunization Clinic

Three participants were recruited from the immunization clinic for 90-minute interview sessions convened by two HCI

XSL•FO
RenderX

researchers and a child-life therapist. Participants attended the meetings with their parent(s) but were interviewed without them being present in the room. The interview started with, "Can you tell us about what it was like last time you came in for immunizations?" We then explained the VR technology and asked, "what would you like to see or do in VR?" Drawing and clay play were used to help participants communicate their ideas.

Semistructured Interviews in the Perioperative Setting

Patients presenting for day surgery were approached in the hospital preoperative waiting room for 15-minute interview sessions. Members of the research team with a background in clinical anesthesia developed questions for semistructured interviews that included a series of rating scales to capture how much the participant knew about the hospital visit on the day of the procedure (4-point scale, ranging from "nothing at all" to "everything"). Additional open-ended questions inquired about things that make children feel good about going to sleep for an operation and things about which they were concerned. A final question asked about the quality of previous experiences of anesthesia where relevant (4-point scale, ranging from "very bad" to "very good").

Parents and Carers

Semistructured Interview in the Immunization Clinic

Three parents were recruited (children were interviewed separately) for 90-minute interviews aimed to gain insight into immunization experiences and strategies that helped their child, like "can you tell us what it was like for your child last time you come in for immunizations?" and "what sort of things have you or someone tried to make the experience easier?" Participants were then briefed on VR technology and asked if they thought it could help their child.

Semistructured Interviews in the Perioperative Setting

Parents and carers were recruited while in the postoperative waiting area, shortly before discharge from the hospital. Each interview lasted about 10 minutes. We used rating scales and open-ended questions, which were developed by researchers with a background in clinical anesthesia based on issues raised by parents and carers in their clinical practice. Participants rated how much they knew about various aspects of the appointment (4-point scale, ranging from "nothing at all" to "everything"). Open-ended questions included "what were good or difficult aspects of their hospital visit and postoperative time?" and "what can be improved?" and whether they were worried about the operation (yes/no). When children had an experience of prior anesthetics, participants rated the quality of that experience (5-point scale, from "very bad" to "very good").

Results

All interviews and focus group recordings were transcribed. Two researchers performed thematic analysis separately and discussed the outcomes until they agreed on a coding scheme based on similarity in meanings associated with participants' statements, following what Braun and Clarke [19] identify as a bottom-up or inductive approach. The coding scheme included four themes based on the interview data obtained from children and adolescents in the immunization group, which exhibited similarity to the parents' group in the same setting. The themes were then used as a coding scheme to analyze interview data in all three groups in that setting. Survey data was tabulated in Excel (Microsoft Corporation), and average ratings defined. Participants' comments in the perioperative setting were analyzed to identify areas of interest and strategies that helped with calming children based on their experiences.

Clinicians

Focus Group

Participants (2 males, 2 females) were anesthetists (n=2), a registered nurse (n=1), and a child life therapist (n=1). Participants worked with children in a variety of settings, including perioperative care, procedural sedation settings, and immunization services. All participants were familiar with VR technology, but none had used it in their practice. Four themes were identified in the participants' statements (Textbox 1).

XSL•FO
RenderX

Textbox 1. A summary of clinicians' perspectives on design and intervention strategies that fulfill the experience and psychological needs of pediatric patients to manage their periprocedural anxiety.

Experience needs

- Patient's prior experience
- Parent's attitude
- Patient's attitude

Psychological needs

- Sense of agency
- Privacy

Intervention strategies

- Set achievable goals for the patient
- Provide distraction
- Improve communication
- Personalize intervention
- Educate

Design strategies

- Provide positive emotional experiences
- Embed narrative storytelling
- Accomplishment and satisfaction through game tasks
- Match contextual requirements

Experience Needs

The patient's prior experience and their parent's attitude towards their child's anxiety can hinder a child's ability to manage their periprocedural anxiety or even promote anxiety. In response to prior negative experiences, some children become anxious as soon as the nurse or doctor enters the room. This response is often exacerbated if children detect parental anxiety. Our participants identified three typical personas based on the patient's attitude: *trusting child* (who is easy to connect with and instruct), *timid child* (engaging with whom requires additional effort by clinicians), and *anxious child* (engaging with whom requires the most effort, and sometimes multiple visits to the hospital).

Psychological Needs

Providing a sense of agency (eg, offering realistic choices to the patient) and ensuring privacy throughout the appointment are essential considerations for fulfilling the psychological needs of pediatric patients to manage their anxiety.

Intervention Strategies

Typical strategies for pediatric anxiety management include setting achievable goals, so the patient gains a sense of accomplishment, distraction through activities, such as singing songs, blowing bubbles, playing games, communication of information, and allowing time for familiarization. One participant noted, "I gave a mask to a patient to take home, and they practiced diligently every day before the next appointment and were then able to complete the procedure more comfortably." Participants suggested they tailored their strategies to the specific needs of the patient, such as by adjusting the level or form of distraction. A fundamental difference was evident among the strategies sought by the anesthetists, the child life therapist, and the nurse. Anesthetists' interactions with children typically last a few minutes. They are focused on negotiating alternative scenarios with the child (eg, a choice between taking the mask or needle injection), whereas the therapist and the nurse in the group wanted to communicate, educate, and prepare children for future appointments.

Design Strategies

Participants suggested future VR solutions should provide enjoyable emotional experiences (eg. the thrill of riding a roller coaster), embed narrative storytelling (eg, adventure and exploration), and create a sense of accomplishment in children. Additionally, participants suggested creating virtual experiences that correspond to the events of the physical environments (eg, seeing the wind blow in VR precisely at the time as receiving gas during induction of anesthesia). Physical interaction with the device was noted as important, eg, operating a device with both hands is impractical if one hand must be still for the clinician to establish intravenous access.

Survey

Participants were anesthetists (n=23), anesthesia recovery nurses (n=19), and preoperative nurses (n=8). The clinical experience of participants varied from 1 to 30 years (mean 11, SD 8.28). Average clinician ratings indicate their views on what pediatric patients worry about before, during, and after an operation. As

XSL•FO
RenderX

shown in Table 1, the top three worries for the 4-7 year age group were: going to sleep with a needle (mean 4.62, SD 0.60), being surrounded by strangers (mean 3.76, SD 0.92), and not knowing what is about to happen (mean 3.68, SD 1.02). In the 8-11 age group, the top three worries were: going to sleep with a needle (mean 4.08, SD 0.72), not knowing what is about to happen (mean 3.76, SD 0.80), and possible pain after the procedure (mean 3.48, SD 0.89). For patients aged 12-17, the top worries were: pain after the procedure (mean 3.94, SD 0.68), waking up during anesthesia (mean 3.92, SD 0.85), and not knowing what is about to happen (mean 3.64, SD 0.98).

When asked what things could make the child feel good, calm, or reassured before an operation, clinicians specified some strategies that were common to all age groups:

- Being friendly and creating a positive experience to set a favorable precedent,
- Creating a calm perioperative environment,
- Using distraction techniques (eg, blowing bubbles) to provide experiences of joy or surprise, was stated as particularly important for younger children (4-7 and 8-11),
- Keeping familiar items such as toys that are comforting to pediatric patients,
- Effective communication and providing adequate, truthful and age-appropriate preoperation information was described as more relevant to older children (12-17),
- Positive and friendly engagement, eg, sitting with the patient rather than standing, particularly with younger age groups (4-7 and 8-11),
- Allowing parents to be present, keeping parents calm, and
- Acknowledging the child's agency by offering them choices (eg, gas or needle).

Table 1. The average rating (likelihood of each item occurring on a 5-point scale ranging from "not at all" to "very much") of clinician's responses to the repeated survey on the experience of pediatric patients in three age groups (n=56).

Survey question	Patients		
	4-7 years, mean (SD)	8-11 years, mean (SD)	12-17 years, mean (SD)
When it comes to the preoperative phase, do you think children are most worried about…			
Not knowing what is about to happen?	3.70 (1.02)	3.8 (0.80)	3.6 (0.98)
Waiting in the pre-op area?	2.6 (0.84)	2.8 (0.97)	2.6 (0.95)
Going into theatres?	3.6 (0.73)	3.4 (0.25)	3.1 (0.87)
Being surrounded by strangers?	3.8 (0.92)	2.9 (0.79)	2.2 (0.71)
Going to sleep with a mask?	3.5 (0.76)	2.9 (0.74)	2.4 (0.86)
Going to sleep with a needle?	4.6 (0.60)	4.1 (0.72)	3.4 (0.70)
Reflecting on how kids think about the actual operation, do you think they are worried about…			
Complications that might happen during the operation?	2 (0.82)	2.7 (0.92)	3.5 (1.01)
Waking up during anesthesia?	2.4 (1.14)	3.3 (1.04)	3.9 (0.85)
Reflecting on what kids think about what will happen after the surgery itself, do you think they are worried about…			
Pain after the procedure?	3.2 (0.98)	3.5 (0.89)	3.9 (0.68)
Feeling sick after the procedure?	2.1 (0.95)	2.6 (0.99)	3.1 (0.97)
Being okay to go home?	2.5 (1.18)	2.8 (0.91)	3.2 (0.87)
Complications after the procedure?	1.7 (0.82)	2.3 (0.89)	3.2 (0.90)
Having a scar after the operation?	1.8 (0.71)	2.6 (1.13)	3.4 (0.92)
Making a full recovery?	2.1 (1.02)	2.6 (1.16)	3.2 (1.00)

Children and Adolescents

Semistructured Interviews in the Immunization Setting

Three participants (2 males, 1 female) aged 12-15 years (mean 13) were recruited. All participants said they had prior experience of severe anxiety associated with needles. All participants were familiar with VR technology. Four themes were identified (Textbox 2).

XSL•FO
RenderX

Textbox 2. Children and adolescents' perspectives of their experience and psychological needs and intervention and design strategies that can help them manage their periprocedural anxiety.

Experience needs

- Disconnect from the surroundings through an immersive experience
- Engage in simple activities that are easy to learn
- Engage in recreational activities such as creating art

Psychological needs

- Sense of agency
- Experience a range of emotions
- Set emotional goals

Intervention strategies

- Provide distraction
- Time the distraction well to start distraction before the procedure begins
- Involve other stakeholders (such as parents) in the intervention
- Build resilience through empathic relationships (eg, with pets)
- Personalize intervention

Design strategies

- Embed narrative storytelling
- Familiar yet novel design elements
- Mask unpleasantness, particularly noises

Experience Needs

Participants suggested that prior negative experience impedes their ability to manage their anxiety during hospital visits. VR was suggested to be helpful as it is "immersive," unlike tablet or mobile technology, and it can disconnect them from their physical environment. One participant said, "I just want to get out of there" and wanted to wear the VR headset at the beginning of the appointment to avoid seeing the needle. Simple VR games that are easy to learn and play in a short time were suggested, eg, creating art, playing a favorite sport, singing, and playing music.

Psychological Needs

Participants recognized that a useful VR experience must help them calm down autonomously through gameplay. They gave examples of including emotional goals, starting with "excitement" to match their state before receiving an injection (eg, going on exciting VR adventures) and ending with a "relaxing" experience. One participant said, "I know I can never be relaxed immediately, so I want something exciting to begin with."

Intervention Strategies

Distraction strategies were discussed, such as involving other stakeholders in the game experience (eg, an accompanying parent or carer). Participants noted the importance of personalization if these were to occur in VR. One participant wanted to take their pet to the appointment for comfort, which may suggest the relevance of empathic interactions for coping strategies.

Design Strategies

Several themed narratives were suggested for a VR app, such as nature, superpowers, learning something new, and playing familiar games. Participants also discussed interactive sensory experiences such as "fast-paced" music to mask procedure noises.

Semistructured Interviews in the Perioperative Setting

A total of 65 (44 males, 21 females) participants aged 4-16 (mean 10, SD 3.89) were recruited. Overall, 49 (75%) participants had previous experiences of anesthetics (1 to 17 times, mean 4 times) and rated the quality of those experiences at a mean score of 2, SD 0.81 (4-point scale, "very bad" to "very good") (Table 2).

Overall, 37 (56%) participants claimed they had worries "about the operation or any other things that might happen before, during, or after the operation." Worries about what happens while they are asleep were: having bad dreams, inadequate analgesia, duration or depth of sleep, a repeat of a bad experience, and the surgery itself. Worries about the anesthetics were: change of location, loss of sensation, uncontrolled movement while asleep, sharp objects such as needles, and missing friends and school. Finally, worries related to the postoperative experience were: being alone, emotional lability, the presence of an intravenous cannula, overnight admission, removal of tapes, and surgical outcomes.

XSL•FO
RenderX

Concerning children's knowledge about their hospital visit, the highest rating was given to knowing what happens during the anesthesia (mean 2.28, SD 0.97) and the lowest rating to knowledge about going into the rooms where the operation happens (mean 2.11, SD 0.99).

Table 2. Children's average ratings of how much they knew about the hospital visit on the day of the procedure (4-point scale, ranging from 1="nothing at all" to 4="a lot or everything").

Survey question	Mean (SD)
How much did you know about what it would be like when you got to the hospital?	2.28 (1.02)
How much did you know about where you would spend time before the surgery?	2.19 (1.03)
How much did you know about who you would meet before the operation?	2.19 (0.90)
How much did you know about what it would be like going into the operation room?	2.11 (0.99)
How much do you know about what happens during an anesthetics?	2.28 (0.97)
How much do you know about what happens during surgery?	2.16 (1.01)
How much do you know about what happens to you after the surgery?	2.19 (0.94)

Additional comments revealed opportunities to improve preoperative experiences across all age groups, such as effective communication of preoperative information (knowing they will not be able to feel the pain or needles during the surgery), having a parent present, and anticipation of positive surgical outcomes. Distraction was mentioned by children in age groups 4-7 and 8-11, but not by those age 12+, who additionally mentioned "staff attitude" and "a child-friendly environment." When asked what could be fun with the anesthetics and operation, responses varied from having access to toys, games, and music, getting well, sleeping, and the hospital facility itself.

Parents

Semistructured Interview in the Immunization Clinic

All three participants identified their child as having severe anxiety related to needles. Only one parent was familiar with VR technology. Details of the four themes are discussed below:

1. *Experience needs*. Participants discussed their emotional experiences, such as helplessness due to their perceived inability to help with their children's anxiety. One parent felt that they did not have the means or time to prepare their child for appointments at the hospital.
2. *Psychological needs*. Parents mentioned that having information about the procedure affords their child a level of control, whereas feeling that someone is trying to control their child's response to anxiety only diminishes their ability to manage it. They suggested that being allowed to take personal items to the appointment (eg, their favorite music) is often helpful.
3. *Intervention strategies*. Two strategies were suggested for regulating anxiety during immunization appointments. One was distraction (eg, playing games) so that the child does not "see" the needle, and the other was goal setting to shift their focus from the unpleasant elements.
4. *Design strategies*. Parents described their children as being sensitive (eg, building things up in their head that adds to their anxiety) who sometimes feel guilty for their anxiety. They suggested that a VR experience could help them build a sense of accomplishment and pride.

Semistructured Interview in the Perioperative Setting

A total of 35 parents and carers were recruited. Their children were aged 4-16 (mean 9.86, SD 3.85). Overall, 23 (65%) parents identified their children as having prior experiences of anesthesia (prior instances mean 4.3, SD 4.66), with an average quality of prior experiences rated 3.91, SD 1.04 (5-point scale, "very bad" to "very good"). In total, 12 parents were concerned about what might happen before, during, or after the operation.

When asked how they might describe their child at the beginning of anesthesia (4-point scale, "anxious" to "calm"), the average rating was 3.31 (SD 0.76). The average knowledge about various aspects of the procedure is shown in Table 3. Most items were rated above 3, with the lowest ratings given for knowledge about waiting before the operation (mean 3.09, SD 1.03) and what happens during surgery (mean 3.09, SD 0.97).

When asked about what might improve their child's preparation for anesthesia, parents suggested making children feel safe to set a better precedent for future appointments. The also suggested creating a positive environment (eg, smaller waiting rooms), distracting and keeping the child occupied (eg, gameplay), clear communication of what to expect, positive engagement with trusting and caring staff and allowing a parent to be present at the time of induction.

When asked about what they found difficult at the hospital, participants cited preoperative anxiety and fasting, worrying about the child will wake up during the surgery, the postoperative state of their child (eg, being disoriented), long waiting times, the environment not being child-friendly, and lack of information about the procedure and logistics (eg, car park).

XSL•FO
RenderX

Table 3. Parents' (n=35) average ratings of how much they knew about the procedure before arriving at the hospital (4-point scale, ranging from 1="nothing at all" to 4="everything").

Survey question	Mean (SD)
Before arriving today, how much did you know about what to do when you got to the hospital?	3.32 (0.64)
Before arriving today, how much did you know about what the room where you wait before the operation would be like?	3.09 (1.03)
Before arriving today, how much did you know about who you would meet before the operation?	3.30 (0.92)
Before today, how much did you know about what it would be like going into the actual operating room?	3.18 (0.87)
Before today, how much did you know about what happens during an anesthetic?	3.15 (0.82)
Before today, how much did you know about what happens during surgery?	3.09 (0.97)
Before today, how much did you know about what happens after surgery?	3.21 (0.88)

Discussion

There is currently a substantial gap in meaningful and truly human-centered applications of VR for managing pediatric anxiety, which can be addressed by putting design and the needs of the patients at the heart of VR development [20]. In order to characterize effective VR design strategies that are suitable for this population, we conducted a multiperspective investigation. Data obtained through a focus group, survey, and interviews allowed us to synthesize the perspectives of three stakeholder groups: patients, parents, and clinicians. Based on the findings, we propose 10 factors that highlight opportunities for future VR design beyond simple distraction, to enhance the experiences of children and adolescents in the periprocedural setting. These are discussed next.

Factors Determining a Positive Experience for Pediatric Patients

Table 4 summarizes 10 factors to guide the design of future pediatric VR solutions.

Table 4. A list of 10 design factors that can be used as input when devising a solution to assist children with managing their periprocedural anxiety. Groups of participants (clinicians, children, and parents) who have mentioned each factor are identified with an "x."

Design factors	Example	Clinicians	Children	Parents
Empathic experience	Diversifying a range of positive experiences such as reassurance, empathy, calm	x	x	x
Welcoming Environment	Creating a child-friendly and warm environment with minimal complexity in procedures	x	x	x
Stimulation & distraction	Offering fun and enjoyable experiences (eg, positive surprise) that distract from negative stimuli	x	x	x
Personalized strategies	Setting achievable goals that are tailored to the child's ability and providing a sense of accomplishment	x	x	x
Effective communication	Providing useful information and answering questions helps children deal with their anxiety and increases their confidence	x	x	x
Engagement with staff	Positive engagements with staff improve emotional support and reduce anxiety	x	x	x
Involving parents	Feeling that parents are close and part of the procedure provides a feeling of safety and diminishes anxiety in children	x	x	x
Acknowledging agency	Feeling in control and having some choices provides a sense of agency and confidence In managing anxiety	x	x	x
Fulfilling emotional needs	Acknowledging the child's emotional expectations helps them cope with their anxiety (eg, not expecting them to calm down immediately and helping them to work through their emotions)		x	x
Familiar design	Familiar forms of technology and design can improve the child's belief that they have the tools to manage their anxiety	x	x	x

Experience

VR can offer a positive experience of reassurance, empathy, and calm to negate children's anxiety. Empathy, in particular, is demonstrated valuable in *Farmoo*, a VR app that allows pediatric cancer patients to care for a virtual farm and companion character [21].

Welcoming Environment

Immersive VR environments can establish a sense of presence in a child-friendly, intimate, and age-appropriate virtual space with proportionality cultivating confidence and preventing the child from feeling small.

Stimulation and Distraction

Our findings suggest distraction paired with stimulating emotional experiences such as positive surprise, is desirable

XSL•FO
RenderX

and may help children disconnect from distressing elements of the environment. Positive emotions are shown to alleviate psychological stress [22] and improve the efficacy of distraction through VR [18].

Personalized Strategies

Distraction in VR can be tailored to a patient's age and abilities and preferences to set achievable goals and elicit a sense of accomplishment, as noted by parents and clinicians in our study. All three stakeholder groups agreed that VR solutions should affirm the child's belief that they have the tools to manage their anxiety. From our findings, we can infer that the clinicians' perception of "what would make a child feel good, calm and reassured" is primarily based on their existing knowledge of an average child's cognitive abilities appropriate to that age group. Therefore, for the oldest age group (12-17), clinicians cited "effective communication such as providing adequate, truthful and age-appropriate preoperative information" while considering "using distraction techniques" and "positive staff engagement" as vital to the 8-11 and 4-7 age groups. However, children from all three age groups cited that it was effective communication that made them feel good about going to sleep for the operation. Although distraction was mentioned by the younger children (4-7 and 8-11), it was not mentioned by those in the adolescent group. This result was different from the clinicians' response that distraction was a priority for the 12-17 age group.

Effective Communication

Our findings demonstrated the importance of providing adequate information for reducing pediatric anxiety. In our study, 56% (n=36) of children and nearly 30% (n=12) of parents in the preoperative setting expressed worry about all stages of an operation when their knowledge about the operating room was lowest. This result confirms the need for better communication of information with patients and their carers. There have been some examples in this area, with mixed results. Ryu et al [7] showed a VR tour of the operating room to children aged 4-10 and found it significantly reduced their anxiety. However, Liszio and Masuch's [23] playful VR simulation of an MRI procedure for children aged 8-15 did not produce a significant reduction in patients' anxiety. Instead, it reduced the parents' anxiety and enhanced clinician satisfaction. Eijlers et al [24] similarly used VR to show the operating room to children (n=191, aged 2-12) before surgeries. They also found no significant changes in self-reported anxiety; however, patients needed rescue analgesia significantly (P=.002) less often (55%) than the control group (95.7%). The challenges of measuring the efficacy of VR intervention are not trivial. Although self-reported anxiety was not reduced in these cases, meaningful clinical measures (ie, rescue analgesic requirement) did confirm a significant difference. Future studies may need to examine the impact of integrating multiple design factors (in Table 4) into a holistic VR plan for children to manage their anxiety.

Engagement With Staff

In addition to supporting the implementation of VR solutions, staff can partake in a virtual experience to further connect with the patient and even personalize the VR app based on the patient's age and abilities. *Voxel Bay,* for example, is a VR app where some game features are controlled by the clinician to provide a positive surprise to the child [8].

Involving Parents

Similar to engagement with staff, VR provides an excellent platform to create a parental presence during the procedure. All three groups of stakeholders in our study stated this is crucial for managing pediatric patient's anxiety.

Acknowledging Agency

All groups involved in our study confirmed that fostering agency is vital to pediatric patient's ability to manage their anxiety. Agency is exemplified in *Voxel Bay* VR app [8] in the form of several choices afforded to the patient from customizing their cardboard VR headset to picking a companion game character.

Fulfilling Emotional Needs

Our research identified emotional goal-setting as a unique strategy to support anxiety management. Children interviewed in our study independently identified that at the commencement of procedures, significant anxiety results in a state of heightened arousal, which a VR experience should match before any attempt to induce relaxation or distraction. To our knowledge, this remarkable finding has not been implemented in any study or existing app. Surprisingly, this strategy was not suggested by clinician participants in our research, further illustrating the relevance of a multiperspective approach for mapping future design opportunities.

Familiar Design

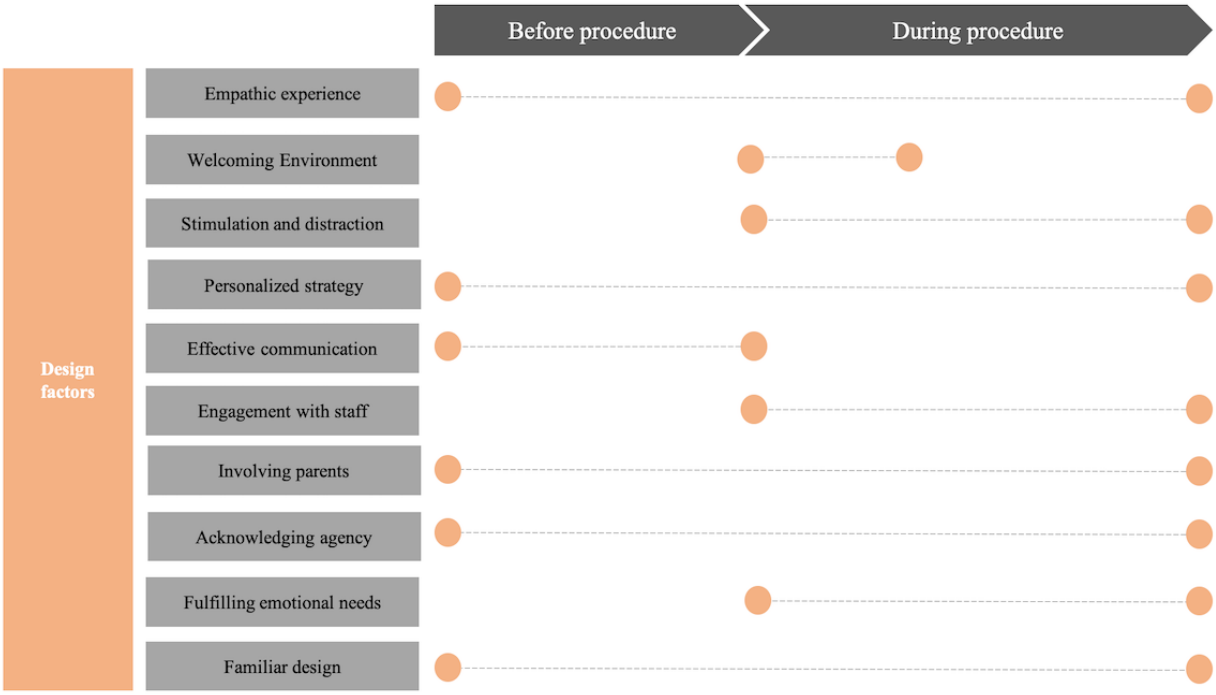
Both clinicians and children in our study stated that having familiar items at the hospital is comforting. This need for a comfort item can be achieved in VR by embedding a familiar design or music in the app, potentially reducing barriers to app learnability, as noted by our young participants.

Future Design Directions for Pediatric VR Solutions

The diversity of the 10 design factors generated based on our findings reveals the complexities and potential challenges in meeting the experiential, psychological, and intervention needs of pediatric patients and their carers through one VR solution. Ahmadpour et al [18] called for skill-building in VR to allow patients to become active agents in their care in a variety of ways. Building meaningful skills in VR necessitates a holistic approach due to the temporality of periprocedural experience with different needs becoming prominent at different stages of the procedure. Therefore, any individual VR solution cannot encompass all 10 factors. Instead, the factors can be mapped to address specific needs at different points of the patient's journey (before the medical procedure and during the procedure), as shown in Figure 1.

XSL•FO
RenderX

Figure 1. The patient journey mapped across the medical procedure phases (before and during) with design factors that can make an impact in either or both of those phases.



This effort may start with a *communication-focused* VR app to familiarize the patient and carers with the details of their procedure in advance. This can then be coupled with an *procedure-focused* VR solution to strategically shift the patient's focus and set emotional goals to help them manage their anxiety autonomously. Design elements that are stimulating, welcoming, and engaging of staff or parents through gameplay, may enhance the effectiveness of these apps.

The research presented in this paper is part of the larger Kids Immersive VR (KiDiVR) project, a collaboration between The University of Sydney and The Children's Hospital at Westmead, which aims to establish a suite of VR apps to augment the pediatric patient journey. This study helped us scope and define the opportunities within the KiDiVR suite. The next phase will involve ideation and prototyping. Specifically, we seek to create a novel VR app to test the efficacy of using emotional goals to guide patient arousal from high to low in order to reduce their anxiety. Our approach is iterative, and therefore we will seek feedback from multiple stakeholders throughout our process to ensure their values are captured, and their needs are met.

Limitations

This study has several limitations. We collected qualitative and quantitative data using tailored questions. This decision was made due to the exploratory nature of our study as a preliminary effort to capture the values and needs of multiple stakeholder groups. The findings should, therefore, be verified using validated tools in the future. The validity of the 10 proposed design factors should be examined using VR prototypes and through controlled clinical trials. We identified a limitation in methods used to assess the efficacy of pediatric VR apps in the literature when testing mainly relied on self-reported anxiety. Future research could address this limitation by investigating the efficacy of VR apps using a combination of meaningful clinical measures (eg, need for rescue analgesia), physiological measures of anxiety, and qualitative assessment (eg, satisfaction), as suggested by Ahmadpour et al [18].

Our participants were recruited from one hospital in Australia. The sample size in our immunization group was small, which limited our interactions with this potential but important target user group. Although we were able to recruit a wide range of respondents in our perioperative group, more diversity in the sample and recruitment from other institutions may result in different perspectives. There is also the potential for selection bias to be introduced, as volunteers for this form of research may represent a more engaged and active group than the general population. All clinicians in the focus group were familiar with VR technology, as were the children and adolescents in the immunization group. All but one parent from the latter group were also familiar with VR. The clinicians and participants in the immunization group were recruited based on direct contact, which may be classified as a convenience sampling approach. Our participants in the immunization group self-identified as having severe anxiety, which was not formally assessed. These may be considered as limitations of our study and challenge the reliability of our findings. We will address these limitations in future studies by recruiting a larger sample using random methods to ensure participants represent a variety of subgroups (eg, those not familiar with VR technology) and performing data triangulation.

Conclusions

This research highlights the value of obtaining multiple perspectives for identifying the needs and values of patients. We identified 10 factors that may inform future VR solutions to enrich the periprocedural experiences of children and adolescents. A striking similarity was evident in the key factors identified by children, parents, and clinicians, including

XSL•FO
RenderX

personalization of strategies to help children build skills to deal with their periprocedural anxiety. Many existing VR apps are used to mediate distraction techniques equally to all users irrespective of their abilities, needs, and preferences. There is an opportunity to maximize the utility of VR as a procedural support tool at different points of patients journey, particularly by using a communication-focused solution before the appointment and an procedure-focused solution during the appointment to set meaningful goals for the patient.

Acknowledgments

This research was made possible through funding by the Westmead Program, Faculty of Engineering and IT, and the Sydney School of Architecture, Design, and Planning. The authors thank all participants and staff at The Children's Hospital at Westmead for their help and support. The authors also thank the Kids Research Institute for providing access to infrastructure for the studies.

Conflicts of Interest

None declared.

References

1. Jacobson RM, Swan A, Adegbenro A, Ludington SL, Wollan PC, Poland GA. Making vaccines more acceptable — methods to prevent and minimize pain and other common adverse events associated with vaccines. Vaccine 2001 Mar;19(17-19):2418-2427. [doi: 10.1016/s0264-410x(00)00466-7]
2. National Academies of Sciences and Medicine, Engineering. Pain management and the opioid epidemic : balancing societal and individual benefits and risks of prescription opioid use. In: Richard J B, Morgan A F, Jonathan K P, editors. National Academies Press. Washington, DC: The National Academies Press; 2017.
3. McMurtry CM, Taddio A, Noel M, Antony MM, Chambers CT, Asmundson GJG, et al. Exposure-based Interventions for the management of individuals with high levels of needle fear across the lifespan: a clinical practice guideline and call for further research. Cognitive Behaviour Therapy 2016 Mar 23;45(3):217-235. [doi: 10.1080/16506073.2016.1157204]
4. Arane K, Behboudi A, Goldman RD. Virtual reality for pain and anxiety management in children. Canadian family physician Medecin de famille canadien 2017;63(12):932-934 [FREE Full text]
5. Patel A, Schieble T, Davidson M, Tran MCJ, Schoenberg C, Delphin E, et al. Distraction with a hand-held video game reduces pediatric preoperative anxiety. Paediatr Anaesth 2006 Oct;16(10):1019-1027. [doi: 10.1111/j.1460-9592.2006.01914.x] [Medline: 16972829]
6. Seiden SC, McMullan S, Sequera-Ramos L, De Oliveira GS, Roth A, Rosenblatt A, et al. Tablet-based Interactive Distraction (TBID) vs oral midazolam to minimize perioperative anxiety in pediatric patients: a noninferiority randomized trial. Paediatr Anaesth 2014 Jul 17;24(12):1217-1223. [doi: 10.1111/pan.12475]
7. Ryu J, Park S, Park J, Kim J, Yoo H, Kim T, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Oct 04;104(12):1628-1633. [doi: 10.1002/bjs.10684]
8. Grishchenko A, Luna J, Patterson J. Voxel bay-VR as a distraction for pediatric pain. In: Proceedings of ACM International Conference on Computer Graphics and Interactive Techniques.: ACM; 2016 Presented at: SIGGRAPH'16; 2016; Anaheim, CA p. a94. [doi: 10.1145/2897839.2927403]
9. Meindl JN, Saba S, Gray M, Stuebing L, Jarvis A. Reducing blood draw phobia in an adult with autism spectrum disorder using low‐cost virtual reality exposure therapy. J Appl Res Intellect Disabil 2019 Jun 13;32(6):1446-1452. [doi: 10.1111/jar.12637]
10. Jeffs D, Dorman D, Brown S, Files A, Graves T, Kirk E, et al. Effect of Virtual Reality on Adolescent Pain During Burn Wound Care. Journal of Burn Care & Research 2014;35(5):395-408. [doi: 10.1097/bcr.0000000000000019]
11. Ashfaq A, Diane G, Xin T, Chris S. Immersion in Cardboard VR Compared to a Traditional Head-Mounted Display. : Springer International Publishing; 2017 Jul 17 Presented at: International Conference on Virtual, Augmented and Mixed Reality, in HCI International 2016; 2016; Toronto, Canada p. 269-276. [doi: 10.1007/978-3-319-39907-2_25]
12. Yamada-Rice D, Mushtaq F, Woodgate A, Bosmans D, Douthwaite A, Douthwaite I, et al. Children and Virtual Reality: Emerging Possibilities and Challenges. In: Technical Report. London: Royal College of Art; 2017.
13. Lorenz M, Brade J, Diamond L, Sjölie D, Busch M, Tscheligi M, et al. Presence and User Experience in a Virtual Environment under the Influence of Ethanol: An Explorative Study. Sci Rep 2018 Apr 23;8(1):6407. [doi: 10.1038/s41598-018-24453-5]
14. Dahlquist LM, McKenna KD, Jones KK, Dillinger L, Weiss KE, Ackerman CS. Active and passive distraction using a head-mounted display helmet: Effects on cold pressor pain in children. Health Psychology 2007;26(6):794-801. [doi: 10.1037/0278-6133.26.6.794]
15. Ahmadpour N, Randall H, Choksi H, Gao A, Vaughan C, Poronnik P. Virtual Reality interventions for acute and chronic pain management. The International Journal of Biochemistry & Cell Biology 2019 Sep;114:105568. [doi: 10.1016/j.biocel.2019.105568]
16. Parsons TD, Riva G, Parsons S, Mantovani F, Newbutt N, Lin L, et al. Virtual Reality in Pediatric Psychology. Pediatrics 2017 Nov 01;140(Supplement 2):S86-S91. [doi: 10.1542/peds.2016-1758i]

XSL•FO
RenderX

17. Bailey JO, Bailenson JN. Considering virtual reality in children's lives. Journal of Children and Media 2017 Jan 02;11(1):107-113. [doi: 10.1080/17482798.2016.1268779]
18. Ahmadpour N, Keep M, Janssen A, Rouf AS, Marthick M. Design Strategies for Virtual Reality Interventions for Managing Pain and Anxiety in Children and Adolescents: Scoping Review. JMIR Serious Games 2020 Jan 27;8(1):e14565. [doi: 10.2196/14565]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
21. Ng J, Lo H, Tong X, Gromala D, Jin W. Farmooo , a Virtual Reality Farm Simulation Game Designed for Cancer Pediatric Patients to Distract their Pain during Chemotherapy Treatment. In: Proc Engineering Reality of Virtual Reality. 2018 Jan 28 Presented at: Society for Imaging Science and Technology; 2018; Burlingame, CA p. 432-1-432-4. [doi: 10.2352/issn.2470-1173.2018.03.ervr-432]
22. Folkman S. Positive psychological states and coping with severe stress. Social Science & Medicine 1997 Oct;45(8):1207-1221. [doi: 10.1016/s0277-9536(97)00040-3]
23. Stefan L, Maic M. Virtual Reality MRI: Playful Reduction of Children's Anxiety in MRI exams. In: Proc ACM International Conference on Interaction Design and Children.: ACM; 2017 Presented at: IDC '17; 2017; Stanford, CA p. 127-136. [doi: 10.1145/3078072.3079713]
24. Eijlers R, Dierckx B, Staals LM, Berghmans JM, van der Schroeff MP, Strabbing EM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children. European Journal of Anaesthesiology 2019;36(10):728-737. [doi: 10.1097/eja.0000000000001059]

Abbreviations

HMD: head-mounted display
HCI: human-computer interaction
KiDiVR: Kids Immersive VR
MRI: magnetic resonance imaging
VR: virtual reality

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by L McCann, B Nievas Soriano, J Hamid; comments to author 22.05.20; revised version received 13.06.20; accepted 14.06.20; published 17.07.20.

Please cite as:
Ahmadpour N, Weatherall AD, Menezes M, Yoo S, Hong H, Wong G
Synthesizing Multiple Stakeholder Perspectives on Using Virtual Reality to Improve the Periprocedural Experience in Children and Adolescents: Survey Study
J Med Internet Res 2020;22(7):e19752
URL: http://www.jmir.org/2020/7/e19752/
doi:10.2196/19752
PMID:32706671

©Naseem Ahmadpour, Andrew David Weatherall, Minal Menezes, Soojeong Yoo, Hanyang Hong, Gail Wong. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 17.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Combination of 3-Dimensional Virtual Reality and Hands-On Aromatherapy in Improving Institutionalized Older Adults' Psychological Health: Quasi-Experimental Study

Vivian Ya-Wen Cheng[1], MEd; Chiu-Mieh Huang[2], PhD; Jung-Yu Liao[3], PhD; Hsiao-Pei Hsu[2], PhD student; Shih-Wen Wang[4], MEd; Su-Fei Huang[5], PhD; Jong-Long Guo[4], PhD

[1]PureAroma Healing Academy, Taipei, Taiwan

[2]Institute of Clinical Nursing, School of Nursing, National Yang-Ming University, Taipei, Taiwan

[3]Institute of Population Health sciences, National Health Research Institutes, Miaoli, Taiwan

[4]Department of Health Promotion and Health Education, College of Education, National Taiwan Normal University, Taipei, Taiwan

[5]Department of Senior Citizen Service, Mackay Junior College of Medicine, Taipei, Taiwan

Corresponding Author:
Jong-Long Guo, PhD
Department of Health Promotion and Health Education, College of Education
National Taiwan Normal University
162, Section 1, He-ping East Road
Taipei, 10610
Taiwan
Phone: 886 277493705
Email: jonglong@ntnu.edu.tw

Abstract

Background: In Taiwan, which has one of the most rapidly aging populations in the world, it is becoming increasingly critical to promote successful aging strategies that are effective, easily usable, and acceptable to institutionalized older adults. Although many practitioners and professionals have explored aromatherapy and identified its psychological benefits, the effectiveness of combining 3-dimensional (3D) virtual reality and hands-on aromatherapy remains unknown.

Objective: A quasi-experimental trial was designed to evaluate the effectiveness of this combination in lowering perceived stress and promoting happiness, sleep quality, meditation experience, and life satisfaction among institutionalized older adults in Taiwan.

Methods: A total of 60 institutionalized elderly participants either received the combined intervention or were in a control group. Weekly 2-hour sessions were implemented over 9 weeks. The outcome variables were happiness, perceived stress, sleep quality, meditation experience, and life satisfaction, which were assessed at baseline and after the intervention.

Results: Generalized estimating equation (GEE) analyses indicated that the experimental group showed significant post-intervention improvements in terms of scores for happiness, perceived stress, sleep quality, meditation experience, and life satisfaction (n=48; all $P<.001$). Another GEE analysis showed that the significant improvements in the 5 outcome variables persisted in participants aged 80 years and older (n=35; all $P<.001$).

Conclusions: This is the first trial to explore the effectiveness of a combination of 3D virtual reality and hands-on aromatherapy in improving older adults' psychological health. The results are promising for the promotion of psychological health in institutionalized older adults.

Trial Registration: ClinicalTrials.gov NCT04324216; https://clinicaltrials.gov/ct2/show/NCT04324216.

(J Med Internet Res 2020;22(7):e17096) doi:10.2196/17096

KEYWORDS

three-dimensional; virtual reality; aromatherapy; older adult; happiness; stress; sleep quality; meditation; life satisfaction

XSL•FO
RenderX

Introduction

Scientific and technological advancements and the resulting improvements in human living environments and medical treatments have resulted in a gradual aging of the human population. In 2016, the global average life expectancy at birth was 72.0 years. It is estimated that by 2050, the proportion of adults over 60 will double [1,2]. Taiwan's elderly population is one of the fastest growing in the world, and with a 14.1% elderly population rate in 2018, the country has become an aged society. This figure is expected to exceed 20% by 2026, with Taiwan then becoming a super-aged society. In order to actively plan and prepare for successful aging under these circumstances, evidence-based health promotion programs for elderly persons have gained top priority so that elderly persons can have healthy and productive lives.

However, previous studies showed poor psychological health among older residents of Taiwanese nursing homes. For example, the prevalence rates of unhappiness, poor sleep, depression, and anxiety among surveyed residents of Taiwanese nursing homes were 50% [3], 46.4% [4], 37% [5], and 26.3% [6], respectively. In addition, diminished life satisfaction and higher depression [7] were identified in nursing home residents compared with their community-dwelling counterparts, which indicates the need for psychological health interventions for the institutionalized population.

Aromatherapy, also known as essential oil therapy, is a complementary treatment that uses ingredients from different parts of plants, such as leaves, flowers, and seeds, to yield aromatic essential oils using different extraction techniques. Aromatherapy is widely used clinically in the treatment of chronic pain, anxiety, depression, cognitive disorders, insomnia, and stress-related diseases [8-10]. Studies have shown that essential oils can alleviate participants' depression [11,12] and perceived stress [11,12] and improve sleep quality [13], meditation experience [14], and quality of life [15].

A descriptive exploratory study investigating the prevalence and type of complementary and alternative medicine (CAM) use among older Taiwanese patients with depression found that 69.6% of participants reported using at least one form of CAM, and 20.9% used aromatherapy in the past 12 months; in addition, 6.8% and 7.3% reported using aromatherapy for treating their depression weekly and daily, respectively [16]. The researchers concluded that the popularity of CAM appears to be a consequence of individuals' preferences for a more holistic approach to health care.

Another study in Taiwan indicated that among 3 alternative remedies, aroma massage was a more effective intervention than cognitive stimulation therapy and reminiscence therapy in alleviating the agitated behavior and depressive symptoms of residents with dementia in 10 nursing homes [17]. Aroma massage has the advantage of being an easy-to-learn intervention for staff working with persons with dementia. The findings of this study contributed to clinical practice in nursing homes [17].

Three-dimensional (3D) virtual reality (VR) involves participants using devices, such as helmets and joysticks, to observe a virtual scene. This approach allows situational teaching; it provides an interactive learning environment that is not limited by time and space, thereby increasing convenience in learning and allowing real-time practice. The 3D virtual world provides rich interactions to maintain users' attention in environments similar to the real world. Furthermore, 3D VR is interactive, integrated, and imaginative, and can be used to aid learning [18]. A strong relationship was found between depth of interaction and engagement duration, with user engagement time increasing due to the high interaction with the environment [19]. The advantages of 3D VR educational activities include avoiding the laborious traffic of learners, increasing social participation and interpersonal communication, and not being limited by time. Since these activities can be experienced and practiced any time, they are convenient for older adults with mental disabilities. Furthermore, the effectiveness of 3D VR programs for the psychological health improvement of older adults is well supported by prior literature [20].

The combination of 3D VR and hands-on aromatherapy allows for a powerful learning experience and facilitates the construction of a 3D space for aromatherapy products. There is some evidence that VR can help increase learners' interest and motivation and effectively support knowledge transfer, since the learning process can be settled within an experiential framework [21]. In this regard, 3D VR could provide experiences with new technologies through actual use—learning in VR environments requires interaction, thus encouraging active participation rather than passivity.

Until now, it has been difficult to provide elderly persons with an opportunity to practice before performing hands-on aromatherapy activities. However, the emergence of 3D VR can solve this problem. If 3D VR can be successfully used, the difficulties elderly participants experience when engaging in hands-on aromatherapy activities might be overcome. Another advantage of a 3D VR educational activity is that it can prevent the waste of materials in hands-on activities for institutionalized older adults.

Because older adults' hearing and hand-eye coordination are relatively poor compared with young and middle-aged adults, they are more likely to struggle with the hands-on aromatherapy activities, based on our practical experience. Thus, older adults need more workers and material support for successful engagement in hands-on activities. If the elements of 3D VR technology can be integrated into the aromatherapy intervention and reduce the burden of human and material resources, it will significantly contribute to the CAM literature. It is reasonable to conduct research that explores 2 effective strategies combined to improve the psychological health of institutionalized older adults. Therefore, the purpose of the present study was to explore the effectiveness of a combined program of 3D VR and hands-on aromatherapy in improving institutionalized older adults' psychological health.

XSL•FO
RenderX

Methods

Participants

We used a quasi-experimental design, which was found to be a common study design in aromatherapy studies in a meta-analysis [22]. The research team visited 2 nursing homes and gave an introduction to the purpose and methods of the study. The nursing home which provided the consent earlier than the other was assigned as the experimental group. The other nursing home became the control group and received a compact intervention program after completion of the study. The experimental nursing home included 432 beds, and residents had a mean age of 85.2 years (male: 134/432, 31.0%). The control nursing home included 248 beds, and residents had a mean age of 84.6 years (male: 82/248, 33.1%).

A total of 30 participants were recruited from each nursing home through posters and advertisements, and the total number of participants was 60. According to Kirk [23], for an estimated effect size (estimated population mean group difference divided by the estimated population standard deviation) of .80, the approximate sample size required is 26 for each group when α is set at .05 and the power is set at .80. A previous aromatherapy study yielded significant pre-post intervention improvements in psychological health with an identical sample size [24]. Therefore, the sample size in this study (N=60) was large enough to detect intervention effects. At the end of the intervention, 6 older adults withdrew from the intervention because of serious sickness, hospitalization, etc.

The elderly participants in this study were all older than 65 years. The selection criteria included having the ability to understand verbal instructions, provide simple responses, and operate a joystick freely with at least one hand. The exclusion criteria were (1) a history of severe psychiatric conditions, (2) dementia, (3) significant visual or hearing impairment, (4) marked upper motor difficulties that could affect the participant's ability to participate in the study, and (5) currently suffering from severe illnesses (eg, stroke, Parkinson disease).

Ethics Statement

The study received approval from the Research Ethics Review Committee of National Taiwan Normal University (201903HM012). We confirm that we have obtained verbal permission to use images of the individuals included in this article.

Recruitment and Baseline Procedure

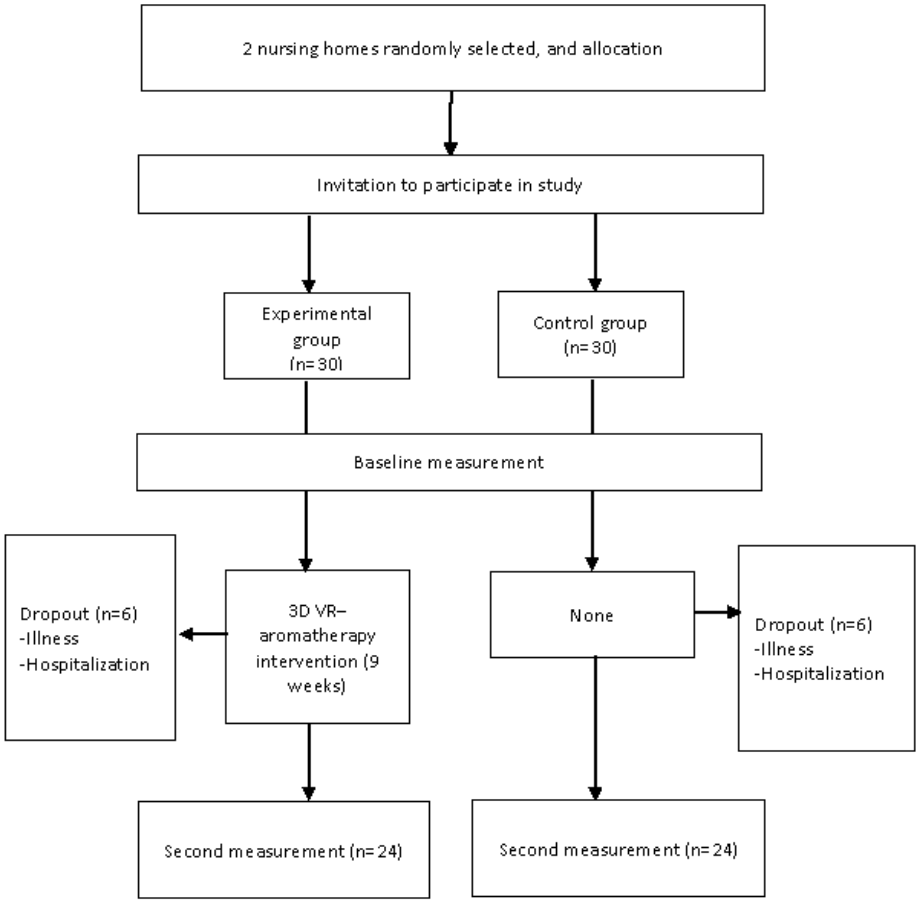
A flowchart outlining participant enrollment and assessments is provided in Figure 1. After selecting the elderly care institutions, the research team approached the executive director and staff to explain the research purpose, method, and protocol. After obtaining permission to conduct the study, we posted recruitment messages to invite potential participants who met the inclusion criteria to participate in this study, and we scheduled one-on-one visits. Before obtaining written consent, we provided face-to-face explanations to potential participants, and participants then signed the consent form.

After potential participants were identified, we provided an orientation session with 3D VR and hands-on aromatherapy to test the feasibility and acceptance of the combination program. Participants indicated their appreciation of this arrangement and reported that the 3D VR program could help them perform better in the subsequent hands-on aromatherapy activities. The research team members collected their baseline data in a quiet room provided by the nursing home.

During the implementation period of the intervention, a medical professional, staff of the nursing home, and an aromatherapy professional were available to ensure the safety of the participants. The counterparts in the control group did not receive the aromatherapy intervention at the same time.

XSL•FO
RenderX

Figure 1. Flowchart of participant enrollment and assessment. 3D: 3-dimensional. VR: virtual reality.



Combination of 3D VR and Hands-On Aromatherapy

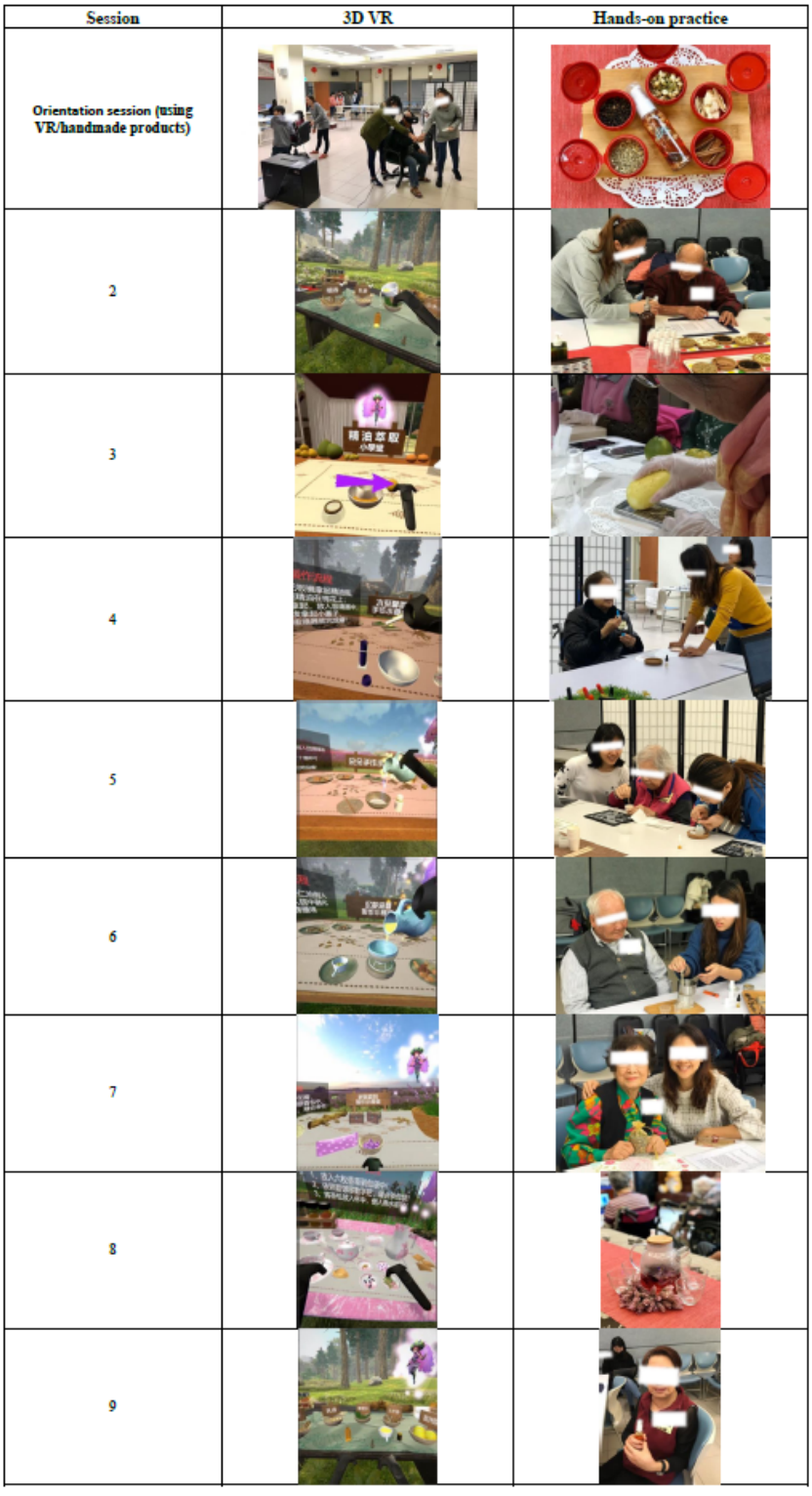
The intervention consisted of 2-hour weekly sessions over 9 weeks. The first week involved ice-breaking activities, during which the participants were taught to wear 3D VR helmets and operate VR handles with familiar VR scenes so that the participants could practice multiple times to avoid dizziness. The contents of the program are shown in Figure 2.

The research team developed the combination program, involving aromatherapy, long-term care, elderly health promotion, and health education professionals. The characteristics of the 3D VR and hands-on aromatherapy program were appropriate for the psychological status of the elderly participants. The final versions of each session (see Multimedia Appendix 1) were prepared after multiple careful revisions of the program. Each session was delivered over 2 hours once a week, and in each session, the participants received hands-on guidance for preparing an aromatherapy product that they could use for the next 7 days. Conducted by 2 trained aromatherapy professionals with certifications from both the UK International Federation of Aromatherapists and the US National Association of Holistic Aromatherapy, with graduate students working as facilitators, the intervention was designed to promote happiness, stress relief, sleep quality, meditation experience, and life satisfaction. The prior findings in the "Introduction" section provided the evidence that helped us determine the program components and select the responsive psychological outcomes in the study.

The participants in the experimental group were divided into multiple groups with a facilitator to enhance individual engagement and solve the problems associated with operating the 3D VR device. A staff member of the nursing home asked participants to record a 7-day log of daily usage of the hands-on aromatherapy product, seen in Figure 2. This log was aimed at increasing the intervention intensity of the program to secure the expected effects. The staff checked participants' logs weekly during the intervention period. All the participants successfully finished the 8-week logs with the support of the assisting staff. The participants were paired with the same group members and facilitator throughout the intervention in order to build rapport. Facilitators received a 2-day training workshop (16 hours total) to acquire knowledge of 3D VR and aromatherapy skills to assist in the activities performed by the participants.

XSL•FO
RenderX

Figure 2. Examples of the 3D VR and hands-on aromatherapy program. 3D: 3-dimensional. VR: virtual reality.



Measures

Sociodemographic variables assessed at baseline included age, gender, and educational level. The psychological outcome variables are presented below.

Happiness

The Oxford Happiness Inventory was used to measure happiness. It consists of 29 items and assesses the following 7 concepts: positive cognition, social commitment, positive affect, sense of control, physical fitness, satisfaction with self, and mental alertness [25]. Each item is scored on a Likert-type scale ranging from 1 to 4, with higher scores indicating a higher level of happiness. The Cronbach α coefficient was .92 during its development [26] and .90 in this study.

XSL•FO
RenderX

Perceived Stress

The Perceived Stress Scale is a self-reported scale that measures the degree of stress experienced by an individual over the last month. In comparison with life events, it has shown good predictive validity for various health outcomes [27]. The scale consists of 14 items [27], 7 positive and 7 negative, and was translated into Chinese. Each item is scored on a Likert-type scale that ranges from 0 (never) to 4 (very often). The scale has been shown to have good internal reliability, with a Cronbach α coefficient of .84 during its development [27] and .85 in this study.

Sleep Quality

The Pittsburgh Sleep Quality Index [28] was used to measure sleep quality. It is an effective instrument used to measure the quality and patterns of sleep in older adults. It differentiates poor from good sleep by measuring 7 domains over the last month: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. The scoring plan was followed according to the description provided in a previous study [28].

Meditation Experience

A shortened version of the 10-item Experiences During Meditation (EOM-DM) scale [25] was used to measure the experience of meditation. The original EOM-DM contains the following 5 subscales: cognitive effects, emotional effects, mystical experiences, relaxation, and physical discomfort. The first 2 subscales were appropriate for our study and were thus selected to measure the meditation experience. Each item was scored on a Likert-type scale from 1 to 5, with higher scores indicating a higher level of experience during meditation. The Cronbach α coefficients for the 2 subscales were .83 and .87 during its development [25] and .90 and .89 in this study.

Life Satisfaction

The life satisfaction scale for older adults developed by the Health Promotion Administration of the Ministry of Health and Welfare [29] was used to measure participants' life satisfaction. This scale is a short-form version containing 10 items from the 20-item Life Satisfaction Index A [30]. Items are scored as 0 (disagree) and 1 (agree). The total raw score ranges from 0 to 10, with higher scores indicating a higher level of life satisfaction. The reliability of the Kuder-Richardson Formula 20 was 0.72.

Data Analyses

Descriptive analyses were conducted for demographic and outcome variables. A 2-tailed *t* test and a chi-square test were used to compare differences in age, education level, and gender between the experimental and control groups. Because the use of a fragmented univariate test may lead to an inflated overall type 1 error, Hotelling T^2 test (multivariate 2-group test) was performed for group comparisons of the 5 outcome measures at baseline [31]. A generalized estimating equation (GEE) was used to investigate the effect of time point, group, and their interaction on the outcome variables; GEEs enable an understanding of the patterns of change and their effects at both the individual and group levels by estimating the average response of the population (the population average effect), rather than regression parameters that would enable prediction of the effect of changing one or more covariates on a given individual [32]. Statistical analyses were conducted using SPSS (version 20.0; IBM Corp).

After the first-round analysis, we found that all outcome variables showed significant improvements (P<.001). This indicated that we could conduct a further analysis for the smaller sample of participants aged 80 and older (very old adults) in order to explore the intervention effects for this group. This further analysis is meaningful because a growing number of residents in nursing homes are in this age group, and these results are important to secure information on the implementation of the combined program for this population.

Results

Sociodemographic Data

The participants' average age was 83.03 (SD 7.6) years and 81.92 (SD 9.0) years in the experimental and control group, respectively. There were no statistically significant differences in participants' education levels and gender distribution between the experimental and control groups. Hotelling T^2 results revealed that the 5 outcome measures at baseline were not significantly different between the groups (T^2=10.95; $F_{5,40}$=2.00; P=.10). Since the overall result was not statistically significant, we did not analyze each outcome variable separately.

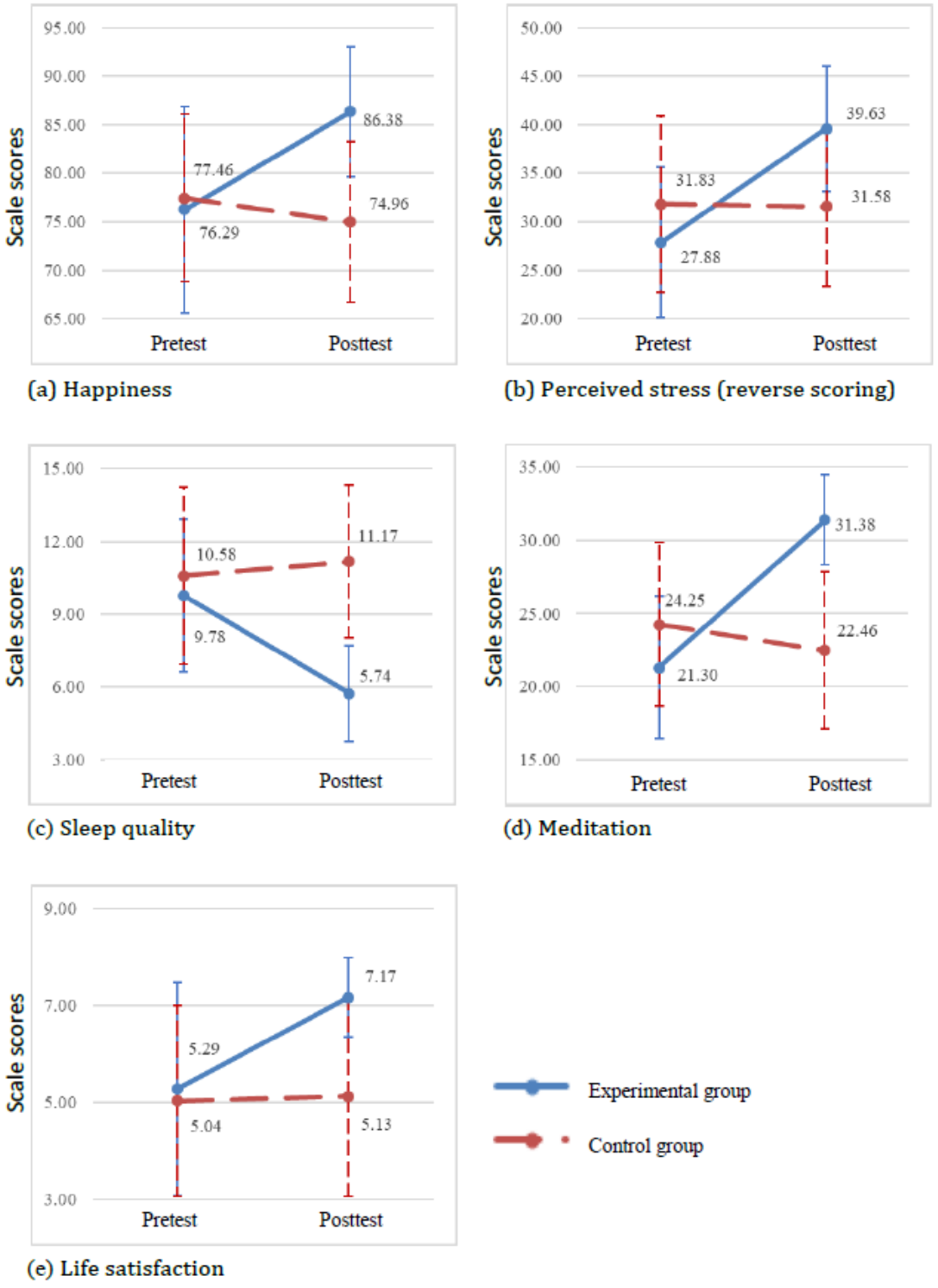
Improvements in Outcome Variables

Group differences in the patterns of change over time are shown in Figure 3. The results of GEE analyses indicated that the experimental group showed significant postintervention improvements in comparison with the control group in terms of the scores for happiness, perceived stress, sleep quality, meditation experience, and life satisfaction, as seen in Table 1. There was a significant group time interaction for the 5 outcome measures. The experimental group showed an improvement in the scores for happiness (GEE coefficient=12.58; P<.001), perceived stress (GEE coefficient=12.00; P<.001), sleep quality (GEE coefficient=–4.72; P<.001), meditation experience (GEE coefficient=11.92; P<.001), and life satisfaction (GEE coefficient=1.79; P<.001).

The majority of participants were adults aged 80 years and older (n=20 and n=15 in the experimental and control group, respectively), and the GEE analysis yielded a significant group time interaction for the 5 outcome variables. The experimental group showed an improvement in the scores for happiness (GEE coefficient=12.43; P<.001), perceived stress (GEE coefficient=13.00; P<.001), sleep quality (GEE coefficient=–4.46; P<.001), meditation experience (GEE coefficient=12.15; P<.001), and life satisfaction (GEE coefficient=1.55; P<.001).

XSL•FO
RenderX

Figure 3. Changes in happiness, perceived stress, sleep quality, meditation, and life satisfaction between experimental and control groups.



XSL•FO
RenderX

Table 1. Results of generalized estimating equation analyses.

	GEE[a] coefficient	SE	95% Wald CI		Wald chi-square (*df*)	*P* value
			Lower	Upper		
Happiness						
Group (experimental group)[b]	–1.17	2.74	–6.53	4.20	0.2 (1)	.67
Time (posttest)[c]	–2.50	0.79	–4.05	–0.95	9.9 (1)	.002
Group (experimental group) × time (posttest)[d]	12.58	1.84	8.98	16.19	46.8 (1)	<.001
Perceived stress (reverse scoring)						
Group (experimental group)[b]	–3.96	2.38	–8.63	0.72	2.8 (1)	.10
Time (posttest)[c]	–0.25	0.73	–1.68	1.18	0.1 (1)	.73
Group (experimental group) × time (posttest)[d]	12.00	1.72	8.64	15.36	48.9 (1)	<.001
Sleep quality						
Group (experimental group)[b]	–0.72	0.97	–2.63	1.18	.6 (1)	.46
Time (posttest)[c]	0.64	0.39	–0.12	1.40	2.7 (1)	.10
Group (experimental group) × time (posttest)[d]	–4.72	0.65	–5.99	–3.45	53.2 (1)	<.001
Meditation						
Group (experimental group)[b]	–3.01	1.50	–5.94	–0.07	4.0 (1)	.04
Time (posttest)[c]	–1.79	0.59	–2.95	–0.63	9.2 (1)	.002
Group (experimental group) × time (posttest)[d]	11.92	1.33	9.32	14.52	80.7 (1)	<.001
Life satisfaction						
Group (experimental group)[b]	0.25	0.59	–0.90	1.40	0.2 (1)	.67
Time (posttest)[c]	0.08	0.17	–0.24	0.41	0.3 (1)	.62
Group (experimental group) × time (posttest)[d]	1.79	0.42	0.97	2.61	18.4 (1)	<.001

[a]GEE: generalized estimating equation.

[b]Reference group (group): control group.

[c]Reference group (time): pretest.

[d]Reference group (group time): control group pretest.

Discussion

Principal Findings

To our best knowledge, this is the first interventional study to use 3D VR and hands-on aromatherapy in a combined program and verify its effects on the psychological health of institutionalized older adults through an appropriate research method. A unique feature of this study is that the participating older adults engaged in hands-on preparation of aromatherapy products that they were then encouraged to use daily over the subsequent week. In addition, a research team member tracked the participants' 7-day use of each aromatherapy product over 8 weeks. We believe that these strategies successfully increased the intervention intensity compared with previous studies, which often lacked a supporting strategy and tracking design for the daily use of hands-on aromatherapy products [33]. These results hold promise for the promotion of this combined aromatherapy program in nursing homes. In addition, the subgroup analyses targeting very old adults also demonstrated similar effects, which suggests that combined aromatherapy programs could be successfully adopted by decision-makers in nursing homes, given the recent sharp increase in the percentage of very old adults in these facilities. The number of very old adults is projected to increase more than threefold from 137 million to 425 million between 2017 and 2050 [2].

We used the outcome variable of happiness instead of depression, which was used in another study [34]. Therefore, we selected the Oxford Happiness Inventory, which is derived

XSL•FO
RenderX

from the Beck Depression Inventory but contains more concepts. Congruent with previous studies, the current findings support the positive effect of short-term interventions on happiness among older adults. Tang and Tse [12] also found short-term improvements with hands-on aromatherapy for patients with depression. However, our posttest scores indicated that the intervention effect was observed after a 9-week intervention, which is different from a previous study [35] that only reported the short-term effect of aromatherapy for older adults with depression.

Moreover, we found an alleviating effect of aromatherapy on perceived stress, which is in line with a previous study [12] in which participants verbally reported stress relief after the first week of the 2-week aromatherapy intervention for stress management. The intervention effect observed in the posttest scores in our study indicated that the use of aromatherapy to decrease perceived stress in older adults is a promising approach. Aromatherapy could successfully reduce the levels of stress hormones and stimulate the production of β-endorphins [36]. For example, a previous study indicated that inhaling lavender and rosemary essential oils can increase free radical–scavenging ability and reduce cortisol secretion, relaxing muscles, relieving stress, and producing calmness [37].

Regarding sleep, a meta-analysis of 12 studies showed that the use of aromatherapy was effective in improving sleep quality [22], suggesting that readily available aromatherapy treatments appear to be effective in promoting sleep. Another systematic review and meta-analysis [13] including 31 trials with a randomized controlled design also indicated a significant effect of aromatherapy on sleep quality. In line with this, our participants reported a significant improvement in sleep quality (P<.001) in terms of posttest scores.

Smith and Kyle [38] indicated that the effects of aromatherapy are mediated by the stimulation of the limbic system by the chemical components or molecules of essential oils that are detected by the olfactory system after inhalation, which activates the hypothalamus and pituitary gland. Olfactory nerves then send signals to the limbic system to trigger memories and emotional responses and thus relieve mental stress.

However, in a systematic review and meta-analysis [13], 11 of 31 trials did not demonstrate any significant effects on sleep quality associated with the use of aromatherapy delivery modes such as inhalation, massage, and oral ingestion; the use of a blend of essential oils or a single essential oil; or the delivery with a mixed method (eg, acupressure massage). Thus, sleep quality interventions using aromatherapy should be administered on the basis of specific guidelines to ensure an efficient use of aromatherapy. We also propose that guidelines should be developed via a systematic approach by conducting appropriate research.

Regarding meditation, a previous study examining 20 adults' meditation processes demonstrated electroencephalogram changes with lavender inhalation, which presented as an increase in fast theta and slow alpha activities in the frontal area during meditation [39]. Redstone [14] reported that meditation and aromatherapy caused a patient to say, "This is the first time I can sit for more than 5 minutes." In line with the findings of previous research, our findings supported the significant effect of the combination of 3D VR and hands-on aromatherapy on the meditation experience. The effects of the combined program on the experience during meditation is an interesting topic that researchers can further explore in the future.

Aromatherapy has been found to be effective in improving psychological symptoms as well as overall quality of life, especially among patients with cancer [40]. We focused on a different study population, namely residents of nursing homes, which will have a wide application in the future. However, some previous studies have also indicated that aromatherapy did not improve life satisfaction [41] or quality of life [42] in the experimental group compared with the control group. These discrepancies might result from differences in the essential oils and targeted psychological variables in the studies. For example, Soden et al [42] used lavender essential oil and an inert carrier oil in their aromatherapy group, and their results were unable to demonstrate any significant long-term benefits of aromatherapy or massage in terms of improving quality of life.

A Cochrane systematic review [43] indicated that aromatherapy is commonly delivered with massage, which serves as complementary therapy that can reduce symptoms and improve the quality of life of patients with cancer. However, although the review found that there was some indication of benefit in the aromatherapy - massage group, this benefit might not be clinically significant due to the small sample size of the studies. The contribution of our study is that it delivered a combination of 3D VR and hands-on aromatherapy to residents of nursing homes, which successfully enhanced their psychological health. Our findings are promising and could encourage researchers and practitioners to provide interventions combining technology and aromatherapy. Our findings suggest that aromatherapy confers benefits for psychological health. However, the methodologies employed in previous studies were heterogeneous, lacked replicable assessments and long-term follow-up, and employed small sample sizes. Replication, longer follow-up periods, and larger trials are critical in future research to accrue the necessary evidence for the research and development of aromatherapy.

Limitations

First, because the program integrated 3D VR and hands-on aromatherapy, the contribution of either approach cannot be easily isolated using the current study design, as they are linked to each other. To validate the effectiveness of the 2 approaches separately, additional studies should be conducted using controlled trials with enough power. Second, we cannot comment on the longer-term effectiveness (ie, 12 months postintervention) of our program. Additional follow-up is needed to determine how the described intervention program affects older adults' psychological health beyond 12 months after completion of the intervention. Third, the intervention might not apply to frail older adults, but it indicated that if more support is provided, such as the involvement of family members and caregivers, frail older adults could also participate in the program.

XSL•FO
RenderX

Conclusion

The results are important for supporting similar future programs for institutionalized older adults. Our program adopted an innovative approach to improve psychological health among institutionalized older adults. A combination of 3D VR and hands-on aromatherapy activities provides more learning opportunities compared with other aromatherapy interventions. In addition, tracking the participants' use of each aromatherapy product for 7-day periods over 8 weeks was a successful approach, as it extended the connection with participants outside the classroom and contributed to the significant improvements in psychological health among participants. In the future, the effectiveness of this approach may be scientifically verified by comparing 2 groups: those who receive only the aromatherapy program and those who receive the aromatherapy program and a monitoring approach such as usage logs.

To ensure early interventions for institutionalized older adults who experience psychological distress and to prevent the development of multiple psychological disorders, researchers and staff in nursing homes should target the older adults who are still in a state of psychological subhealth. Providing one-on-one interventions in nursing homes may not always be feasible or affordable. To reduce the burden of instruction, learning contents could be delivered through a 3D VR program, which can provide a safe and supportive learning environment and empower researchers and practitioners to play a vital role in solving problems, leading discussions, facilitating older adults' learning, and providing feedback. During the intervention period, 2 staff members at the experimental nursing home were interested in participating in the delivery of the intervention and wished to act as facilitators to encourage participants to engage in the interventional activities. This indicates that our program can easily be implemented in nursing homes. Since the implementation is not complicated, nursing home staff receiving short-term training can deliver the program successfully to promote residents' psychological health.

Acknowledgments

We thank all the study participants for their time and commitment. This article was subsidized by the National Taiwan Normal University, Taiwan, Republic of China.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Combination of 3D VR and hands-on aromatherapy: Program components.
[DOCX File , 21 KB - jmir_v22i7e17096_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1510 KB - jmir_v22i7e17096_app2.pdf]

References

1. Guseh JS. Aging of the World's Population. Wiley Online Library. Hoboken, NJ: John Wiley & Sons Inc; 2016. URL: https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781119085621.wbefs352 [accessed 2020-06-10]
2. United Nations, Department of Economic and Social Affairs, Population Division. 2017. World Population Ageing - Highlights. New York, NY: United Nations; 2017. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf [accessed 2019-10-15]
3. Lan Y, Chao S, Chuang C, Shin M, Lin C. A Study of the relationships among depression, hope and attitudes toward death in institutionalized elders. Hungkuang Acad Rev 2008;53:1-13. [doi: 10.6615/HAR.200808.53.01]
4. Tsai Y, Wong TK, Ku Y. Self-care management of sleep disturbances and risk factors for poor sleep among older residents of Taiwanese nursing homes. J Clin Nurs 2008 May;17(9):1219-1226. [doi: 10.1111/j.1365-2702.2007.02020.x] [Medline: 18266847]
5. Tu Y, Lai Y, Shin S, Chang H, Li L. Factors Associated with Depressive Mood in the Elderly Residing at the Long-Term Care Facilities. International Journal of Gerontology 2012 Mar;6(1):5-10 [FREE Full text] [doi: 10.1016/j.ijge.2011.12.004]
6. Hsieh C, Hsu C, Wu H, Liu Y, Chen M, Hsiao Y. The impact of neuropsychiatric symptoms on satisfaction of sleep in institutionalized elderly residents: A pilot study. J Nurs Healthcare Res 2015;11(4):E. [doi: 10.6225/JNHR.11.4.298]
7. Gueldner SH, Loeb S, Morris D, Penrod J, Bramlett M, Johnston L, et al. A comparison of life satisfaction and mood in nursing home residents and community-dwelling elders. Arch Psychiatr Nurs 2001 Oct;15(5):232-240. [doi: 10.1053/apnu.2001.27020] [Medline: 11584352]
8. Lehrner J, Marwinski G, Lehr S, Johren P, Deecke L. Ambient odors of orange and lavender reduce anxiety and improve mood in a dental office. Physiol Behav 2005 Sep 15;86(1-2):92-95. [doi: 10.1016/j.physbeh.2005.06.031] [Medline: 16095639]

XSL•FO
RenderX

9. Perry N, Perry E. Aromatherapy in the management of psychiatric disorders: clinical and neuropharmacological perspectives. CNS Drugs 2006;20(4):257-280. [doi: 10.2165/00023210-200620040-00001] [Medline: 16599645]
10. Snow LA, Hovanec L, Brandt J. A controlled trial of aromatherapy for agitation in nursing home patients with dementia. J Altern Complement Med 2004 Jun;10(3):431-437. [doi: 10.1089/1075553041323696] [Medline: 15253846]
11. Kim S, Song JA, Kim ME, Hur MH. [Effects of Aromatherapy on Menopausal Symptoms, Perceived Stress and Depression in Middle-aged Women: A Systematic Review]. J Korean Acad Nurs 2016 Oct;46(5):619-629. [doi: 10.4040/jkan.2016.46.5.619] [Medline: 27857006]
12. Tang SK, Tse MYM. Aromatherapy: does it help to relieve pain, depression, anxiety, and stress in community-dwelling older persons? Biomed Res Int 2014;2014:430195 [FREE Full text] [doi: 10.1155/2014/430195] [Medline: 25114901]
13. Lin P, Lee P, Tseng S, Lin Y, Chen S, Hou W. Effects of aromatherapy on sleep quality: A systematic review and meta-analysis. Complement Ther Med 2019 Aug;45:156-166. [doi: 10.1016/j.ctim.2019.06.006] [Medline: 31331554]
14. Redstone L. Mindfulness meditation and aromatherapy to reduce stress and anxiety. Arch Psychiatr Nurs 2015 Jun;29(3):192-193. [doi: 10.1016/j.apnu.2015.03.001] [Medline: 26001720]
15. Lua PL, Salihah N, Mazlan N. Effects of inhaled ginger aromatherapy on chemotherapy-induced nausea and vomiting and health-related quality of life in women with breast cancer. Complement Ther Med 2015 Jun;23(3):396-404. [doi: 10.1016/j.ctim.2015.03.009] [Medline: 26051575]
16. Hsu M, Creedy D, Moyle W, Venturato L, Tsay S, Ouyang W. Use of Complementary and Alternative Medicine among adult patients for depression in Taiwan. J Affect Disord 2008 Dec;111(2-3):360-365. [doi: 10.1016/j.jad.2008.03.010] [Medline: 18442859]
17. Yang Y, Lee F, Chao H, Hsu F, Wang J. Comparing the Effects of Cognitive Stimulation, Reminiscence, and Aroma-Massage on Agitation and Depressive Mood in People With Dementia. J Am Med Dir Assoc 2016 Aug 01;17(8):719-724. [doi: 10.1016/j.jamda.2016.03.021] [Medline: 27168052]
18. Burdea GC, Coiffet P. Virtual Reality Technology, 2nd ed. Hoboken, NJ: John Wiley & Sons; 2003:1-464.
19. Yilmaz RM, Baydas O, Karakus T, Goktas Y. An examination of interactions in a three-dimensional virtual world. Computers & Education 2015 Oct;88:256-267. [doi: 10.1016/j.compedu.2015.06.002]
20. Gibbons MC, eHealth SFHD. Märsche, RV 8-RV 652. Berlin, Germany: Strauss Edition Wien; 2018:8147.
21. Mantovani F. VR Learning: Potential and Challenges for the Use of 3D Environments in Education and Training. In: Towards Cyberpsychology: Mind, Cognition and Society in the Internet Age. Amsterdam, Netherlands: IOS Press; 2001:208-225.
22. Hwang E, Shin S. The effects of aromatherapy on sleep improvement: a systematic literature review and meta-analysis. J Altern Complement Med 2015 Feb;21(2):61-68. [doi: 10.1089/acm.2014.0113] [Medline: 25584799]
23. Kirk RE. Statistics: an introduction. In: Effect size measures. Boston: Cengage Learning; 2005.
24. Lee M, Lim S, Song J, Kim M, Hur M. The effects of aromatherapy essential oil inhalation on stress, sleep quality and immunity in healthy adults: Randomized controlled trial. European Journal of Integrative Medicine 2017 Jun;12:79-86. [doi: 10.1016/j.eujim.2017.04.009]
25. Reavley N, Pallant JF. Development of a scale to assess the meditation experience. Personality and Individual Differences 2009 Oct;47(6):547-552. [doi: 10.1016/j.paid.2009.05.007]
26. Hills P, Argyle M. The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Personality and Individual Differences 2002 Nov;33(7):1073-1082. [doi: 10.1016/s0191-8869(01)00213-6]
27. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
28. Tsai P, Wang S, Wang M, Su C, Yang T, Huang C, et al. Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CPSQI) in primary insomnia and control subjects. Qual Life Res 2005 Oct;14(8):1943-1952. [doi: 10.1007/s11136-005-4346-x] [Medline: 16155782]
29. Health Promotion Administration at the Ministry of Health and Welfare in Taiwan. Taiwan Longitudinal Study on Aging (TLSA). HPA. 2018. URL: https://www.hpa.gov.tw/EngPages/Detail.aspx?nodeid=1077&pid=6197 [accessed 2018-10-03]
30. Neugarten BL, Havighurst RJ, Tobin SS. The measurement of life satisfaction. J Gerontol 1961 Apr;16:134-143. [doi: 10.1093/geronj/16.2.134] [Medline: 13728508]
31. Stevens J. Applied multivariate statistics for the social sciences. 5th ed. Mahwah, NJ: Lawrence Erlbaum; 2009.
32. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
33. Babakhanian M, Ghazanfarpour M, Kargarfard L, Roozbeh N, Darvish L, Khadivzadeh T, et al. Effect of Aromatherapy on the Treatment of Psychological Symptoms in Postmenopausal and Elderly Women: A Systematic Review and Meta-analysis. J Menopausal Med 2018 Aug;24(2):127-132 [FREE Full text] [doi: 10.6118/jmm.2018.24.2.127] [Medline: 30202763]
34. Chang SY. [Effects of aroma hand massage on pain, state anxiety and depression in hospice patients with terminal cancer]. Taehan Kanho Hakhoe Chi 2008 Aug;38(4):493-502. [doi: 10.4040/jkan.2008.38.4.493] [Medline: 18753801]

XSL•FO
RenderX

35. Wilkinson SM, Love SB, Westcombe AM, Gambles MA, Burgess CC, Cargill A, et al. Effectiveness of aromatherapy massage in the management of anxiety and depression in patients with cancer: a multicenter randomized controlled trial. J Clin Oncol 2007 Feb 10;25(5):532-539. [doi: 10.1200/JCO.2006.08.9987] [Medline: 17290062]
36. Tiran D, Chummun H. Complementary therapies to reduce physiological stress in pregnancy. Complement Ther Nurs Midwifery 2004 Aug;10(3):162-167. [doi: 10.1016/j.ctnm.2004.03.006] [Medline: 15279857]
37. Atsumi T, Tonosaki K. Smelling lavender and rosemary increases free radical scavenging activity and decreases cortisol level in saliva. Psychiatry Res 2007 Feb 28;150(1):89-96. [doi: 10.1016/j.psychres.2005.12.012] [Medline: 17291597]
38. Smith MC, Kyle L. Holistic foundations of aromatherapy for nursing. Holist Nurs Pract 2008;22(1):3-9. [doi: 10.1097/01.HNP.0000306321.03590.32] [Medline: 18172401]
39. Sayorwan W, Siripornpanich V, Piriyapunyaporn T, Hongratanaworakit T, Kotchabhakdi N, Ruangrungsi N. The effects of lavender oil inhalation on emotional states, autonomic nervous system, and brain electrical activity. J Med Assoc Thai 2012 Apr;95(4):598-606. [Medline: 22612017]
40. Bouya S, Ahmadidarehsima S, Badakhsh M, Balouchi A, Koochakzai M. Effect of aromatherapy interventions on hemodialysis complications: A systematic review. Complement Ther Clin Pract 2018 Aug;32:130-138. [doi: 10.1016/j.ctcp.2018.06.008] [Medline: 30057040]
41. Kim M, Nam E, Paik S. [The effects of aromatherapy on pain, depression, and life satisfaction of arthritis patients]. Taehan Kanho Hakhoe Chi 2005 Feb;35(1):186-194. [doi: 10.4040/jkan.2005.35.1.186] [Medline: 15778570]
42. Soden K, Vincent K, Craske S, Lucas C, Ashley S. A randomized controlled trial of aromatherapy massage in a hospice setting. Palliat Med 2004 Mar;18(2):87-92. [doi: 10.1191/0269216304pm874oa] [Medline: 15046404]
43. Shin E, Seo K, Lee S, Jang J, Jung Y, Kim M, et al. Massage with or without aromatherapy for symptom relief in people with cancer. Cochrane Database Syst Rev 2016 Jun 03(6):CD009873. [doi: 10.1002/14651858.CD009873.pub3] [Medline: 27258432]

Abbreviations

3D: 3-dimensional
CAM: complementary and alternative medicine
EOM DM: Experiences During Meditation
GEE: generalized estimating equation
VR: virtual reality

Edited by G Eysenbach; submitted 18.11.19; peer-reviewed by CY Lin, M Stein; comments to author 13.02.20; revised version received 08.04.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Cheng VYW, Huang CM, Liao JY, Hsu HP, Wang SW, Huang SF, Guo JL
Combination of 3-Dimensional Virtual Reality and Hands-On Aromatherapy in Improving Institutionalized Older Adults' Psychological Health: Quasi-Experimental Study
J Med Internet Res 2020;22(7):e17096
URL: http://www.jmir.org/2020/7/e17096/
doi:10.2196/17096
PMID:32706660

©Vivian Ya-Wen Cheng, Chiu-Mieh Huang, Jung-Yu Liao, Hsiao-Pei Hsu, Shih-Wen Wang, Su-Fei Huang, Jong-Long Guo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality Smartphone-Based Intervention for Smoking Cessation: Pilot Randomized Controlled Trial on Initial Clinical Efficacy and Adherence

Emilio Goldenhersch[1], BA; Johannes Thrul[2], PhD; Joaquín Ungaretti[3,4], PhD; Nicolas Rosencovich[5], BSc; Cristian Waitman[6], BA; Marcelo Rodriguez Ceberio[1,7,8], PhD

[1]Laboratorio de Investigación en Neurociencia y Ciencias Sociales, Universidad de Flores, Ciudad Autónoma de Buenos Aires, Argentina

[2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]Facultad de Psicología, Universidad de Buenos Aires, Buenos Aires, Argentina

[4]Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

[5]Escuela de Ingeniería Biomédica, Universidad Nacional de Córdoba, Córdoba, Argentina

[6]Universidad Siglo XXI, Córdoba, Argentina

[7]Departamento de Psicología, Universidad de Flores, Buenos Aires, Argentina

[8]Escuela Sistemica de Psicología, Buenos Aires, Argentina

Corresponding Author:
Emilio Goldenhersch, BA
Laboratorio de Investigación en Neurociencia y Ciencias Sociales
Universidad de Flores
Avenida Rivadavia 5741
Ciudad Autónoma de Buenos Aires, CP 1414
Argentina
Phone: 1 4808860102
Email: emiliogolden@gmail.com

Abstract

Background: Obstacles to current tobacco cessation programs include limited access and adherence to effective interventions. Digital interventions offer a great opportunity to overcome these difficulties, yet virtual reality has not been used as a remote and self-administered tool to help increase adherence and effectiveness of digital interventions for tobacco cessation.

Objective: This study aimed to evaluate participant adherence and smoking cessation outcomes in a pilot randomized controlled trial of the digital intervention Mindcotine (MindCotine Inc) using a self-administered treatment of virtual reality combined with mindfulness.

Methods: A sample of 120 participants was recruited in the city of Buenos Aires, Argentina (mean age 43.20 years, SD 9.50; 57/120, 47.5% female). Participants were randomly assigned to a treatment group (TG), which received a self-assisted 21-day program based on virtual reality mindful exposure therapy (VR-MET) sessions, daily surveys, and online peer-to-peer support moderated by psychologists, or a control group (CG), which received the online version of the smoking cessation manual from the Argentine Ministry of Health. Follow-up assessments were conducted by online surveys at postintervention and 90-day follow-up. The primary outcome was self-reported abstinence at postintervention, with missing data assumed as still smoking. Secondary outcomes included sustained abstinence at 90-day follow-up, adherence to the program, and readiness to quit.

Results: Follow-up rates at day 1 were 93% (56/60) for the TG and 100% (60/60) for the CG. At postintervention, the TG reported 23% (14/60) abstinence on that day compared with 5% (3/60) in the CG. This difference was statistically significant (χ^2_1=8.3; *P*=.004). The TG reported sustained abstinence of 33% (20/60) at 90 days. Since only 20% (12/60) of participants in the CG completed the 90-day follow-up, we did not conduct a statistical comparison between groups at this follow-up time point. Among participants still smoking at postintervention, the TG was significantly more ready to quit compared to the CG (TG: mean 7.71, SD 0.13; CG: mean 7.16, SD 0.13; *P*=.005). A total of 41% (23/56) of participants completed the treatment in the time frame recommended by the program.

XSL•FO
RenderX

Conclusions: Results provide initial support for participant adherence to and efficacy of Mindcotine and warrant testing the intervention in a fully powered randomized trial. However, feasibility of trial follow-up assessment procedures for control group participants needs to be improved. Further research is needed on the impact of VR-MET on long-term outcomes.

Trial Registration: ISRCTN Registry ISRCTN50586181; http://www.isrctn.com/ISRCTN50586181

(J Med Internet Res 2020;22(7):e17571) doi:10.2196/17571

KEYWORDS

smoking cessation; nicotine dependence; craving; virtual reality; mindfulness; digital therapy; mHealth; mobile phone

Introduction

Background

Tobacco use remains one of the biggest threats to public health and the leading preventable cause of mortality and morbidity worldwide [1]. Interventions on web-based platforms and mobile apps, among them smoking cessation apps, can deliver effective interventions for various diseases and behavioral disorders [2-16]. Smartphone-based smoking cessation apps can provide an important channel for offering interventions to the entire population [17]. However, participant adherence remains a challenge, and many of these apps struggle to maintain high levels of adherence, thus limiting their potential effectiveness [18].

In addition to smartphone-based interventions for smoking cessation, there have been promising developments in the use of virtual reality (VR). VR is defined as "real-time interactive graphics with [3-dimensional] models, combined with a display technology that gives the user the immersion in the model world and direct manipulation" [19]. The use of VR in digital medicine has been applied in consultancy and hospitals and accompanied by a health professional [20]. It has been demonstrated that the introduction of VR into treatment can improve patient engagement in a range of chronic disease interventions [21]. However, such technology has never been tested for tobacco use disorder as a self-assisted and remote solution to take advantage of the knowledge on how virtual environments can be used to elicit and reduce cravings and support smoking cessation.

Cravings are defined as intense urges or impulses for substance use [22] and are an essential component of substance use disorders, as they provoke drug-seeking behavior [23]. External cues can provoke cue-induced cravings [24] and thus provoke relapse [25,26]. Taken together, these findings suggest that craving is a critical target in the development of novel therapeutics for tobacco use disorder treatment [27,28].

Cue exposure therapy for substance use disorders involves controlled and repeated exposure to drug-related cues in order to extinguish cue-induced cravings [29], and VR technology is a potential mode of cue presentation [30,31]. Virtual reality exposure therapy (VRET) has already been used for smoking cessation [32,33], and studies have applied rigorous systematizations in VRET and cognitive behavioral therapy (CBT) [34-37], providing further evidence that this technique can reduce craving and smoking behavior, with similar effectiveness as stand-alone CBT.

Mindfulness is usually thought of as the awareness to attend to any thought, feeling, or sensation that occurs by simply acknowledging it, without attempting to regulate emotions [38]. Mindfulness training is already used in mobile health (mHealth) smoking cessation interventions [39] by teaching individuals to pay attention to the present moment, understand affective states and cravings to smoke as they appear, and consciously choose to let them pass without impulsively reacting to them. Preliminary evidence suggests that mindfulness-based interventions are associated with increased efficacy compared with other smoking cessation treatments [40,41]. Despite these promising results, in-person mindfulness training continues to present certain challenges. It requires experienced psychotherapists, increases time demands, limits access, and elevates costs [39].

An Innovative Intervention: Mindcotine

The aim of the current study was to develop an accessible and cost-effective digital intervention for smoking cessation that uses the latest technology adapted for large-scale use and to evaluate participant adherence and smoking cessation outcomes. Our novel intervention combines exposure to smoking-related cues in ecological situations using virtual reality and mindfulness as a tool to cope with in situ cravings, bodily sensations, affective states, and automatic reactions. The merge of these psychological frameworks are expected to work synergistically by reducing cravings. On the one hand, cue exposure treatment involves repeated exposure to smoking-related stimuli in order to elicit and, over time, extinguish cue-induced cravings. In addition, mindfulness techniques work on top of the immersive experience by helping the user to focus on cravings in the present moment. Mindfulness practice includes elements of acceptance and compassion through a perspective of curiosity towards the emerging sensations. Taken together, this intervention can help the user stay grounded in the present moment while acknowledging internal and external triggers. Moreover, these virtual reality and mindfulness sessions are part of a classic cognitive behavioral smoking cessation program that provides information on relevant topics daily through CBT notifications and community support among users moderated by psychologists and mindfulness facilitators. We hypothesized that this novel intervention approach would increase both adherence and abstinence rates among participating smokers compared to a treatment-as-usual control group.

XSL•FO
RenderX

Methods

Research Design

The design of the study follows the recommendations for clinical trials in health using virtual reality, in particular, tier VR2 [42], focusing on acceptability, feasibility, tolerability, and initial clinical efficacy. We conducted a clinical trial with a control group, including baseline and follow-up assessments at days 1 and 90 posttreatment.

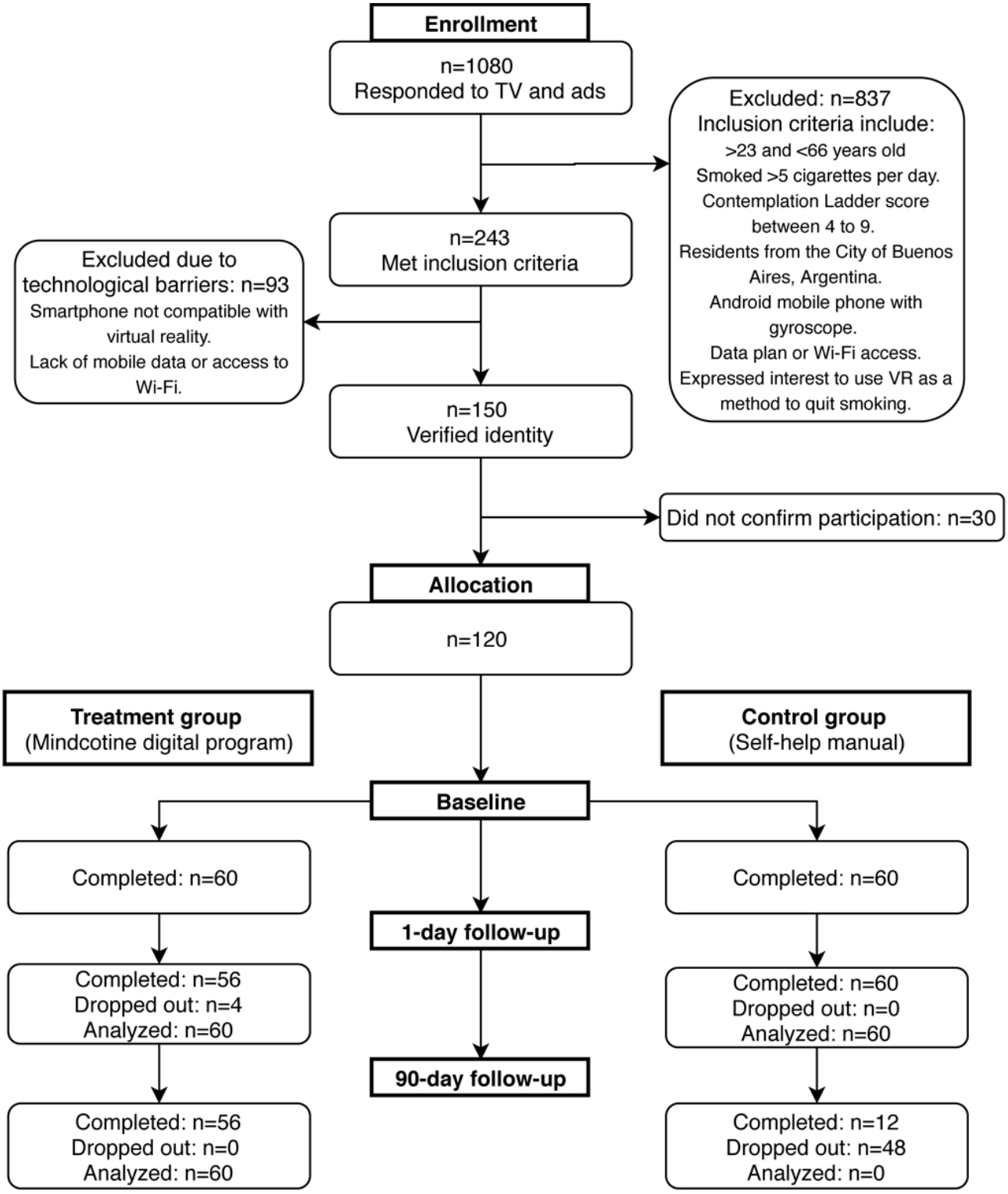
Procedure

The study was conducted from February to April of 2018 with a group of smokers from Buenos Aires, Argentina. The administration to the treatment group was carried out through an onboarding process that established the framework of the experience and the basic principles of the use of VR. The treatment was remotely self-administered through the use of a mobile app. The control group received a smoking cessation manual of the Argentine Ministry of Health.

The intervention group completed the onboarding process, which included digital informed consent, and received an intervention kit, which included guidelines that explained the use of the program, a unique ID to access the app, one cardboard headset, three stickers to announce smoke-free areas, and two wrist bracelets as a behavior replacement method that smokers could use to snap on their wrists when faced with a craving to smoke. Participants were also instructed to download the mobile app from the Google Play Store (Google Corp). The participants were trained in assembling the cardboard headset (see Multimedia Appendix 1) and app usability, and they completed the baseline assessment and received an explanation of the activities to be carried out during the 21-day intervention phase. The control group signed a digital informed consent sent to their emails and received the online version of the smoking cessation manual developed by the Office of the President of the Argentine Nation [43]. Intervention group participants received the first follow-up 1 day after they completed all 21 days of intervention content, which for some participants took longer than 21 days (postintervention follow-up) and another follow-up assessment at day 90 of posttreatment. Control group participants were assessed at days 30 and 90 after they had received the smoking cessation manual. All assessment invitations were sent via email and assessments were completed online using the Typeform (Typeform SL) platform. Figure 1 shows the participant flow diagram of the study. Participants did not receive any incentives. All study procedures were approved by the Institutional Review Board of the University of Flores, Buenos Aires, Argentina. The trial was registered in the International Standard Randomized Controlled Trial Number Registry (50586181).

XSL•FO
RenderX

Figure 1. Participant flow through the Mindcotine method smoking cessation trial. TV: television; VR: virtual reality.



Participants

In order to be included in the study, the selected participants had to meet the following criteria: (1) be aged between 24 and 65 years, representing smokers with a high prevalence of daily smoking [44]; (2) consume a minimum of 5 cigarettes per day, with a score of 4 to 9 on the Contemplation Ladder [45]; (3) be residents in the city of Buenos Aires; (4) own an Android mobile phone with gyroscope; (5) have a data plan or Wi-Fi access; and (6) have an interest in using VR as a method to quit smoking. Each of these criteria was based on previous work with mobile apps for smoking cessation [46]. The one difference in our inclusion criteria compared with this existing research was that we also included smokers that scored below 7 on the Contemplation Ladder in order to investigate intervention effects for smokers not motivated to quit. Participants were excluded if they were diagnosed with a current psychiatric disorder.

For the recruitment of treatment and control group participants, unpaid advertisements were posted for 75 days on the

XSL•FO
RenderX

Mindcotine page on Facebook, and a 13-minute segment was aired on Argentine national public television (channel C5N). The social network advertisements linked invitations directly to a screening questionnaire on the Typeform platform [47], which was completed by 1080 potentially eligible volunteers from all over Argentina. Of these initial applicants, 234 subjects lived in the city of Buenos Aires and fulfilled all other necessary inclusion criteria, and telephone contact was established to verify the data. A total of 150 participants were accepted after establishing contact and verifying their responses. However, 30 of these individuals did not confirm their participation, which resulted in a final sample of 120 volunteers. Participants were randomized into the treatment group (TG) (n=60) or control group (CG) (n=60) 1:1 using a blocked random assignment sequence. The treatment group was invited to the research site (University of Flores) and all participants signed an informed consent form, underwent the face-to-face onboarding process with one member of the study team, and were given the Mindcotine Kit to begin the study. The control group was contacted through email, provided digital informed consent, and received the online version of the smoking cessation manual by the Argentine Ministry of Health.

Intervention

The app consists of a 21-day treatment that includes 2 main activities each day, which become available after completing the activities of the previous day (see Figure 2). Intervention development was conducted iteratively using face-to-face testing and focus groups, with feedback from 250 smokers who tested the virtual reality environment of the intervention. Moreover, we reviewed existing smoking cessation apps and features related to trigger identification and reflective questions.

Figure 2. Mindcotine procedure flow diagram. 2D: 2-dimensional; CL: Contemplation Ladder; FFMQ: Five Facet Mindfulness Questionnaire; FTND: Fagerström Test of Nicotine Dependence; FU-90: 90-day follow-up; PI: postintervention; QSU: Questionnaire for Smoking Urges.

Mindcotine´s Procedure		Days																					
Instrument	Format	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	PI	FU-90
CL	Questionnaire	✓																					✓	
QSU	Questionnaire	✓						✓							✓								✓	
FTND	Questionnaire	✓						✓							✓								✓	✓
FFMQ	Questionnaire	✓																					✓	✓
Daily self-registration	Daily questionnaire	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mindfulness sessions	Audio	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Formal mindfulness	2D session	✓	✓	✓	✓	✓	✓																	
Informal mindfulness	Virtual reality session							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Community chat	Chat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Mindcotine support	SMS and phone calls			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

The elements of the program were (1) practice sessions in formal mindfulness, (2) practice sessions in informal mindfulness using virtual reality mindful exposure therapy (VR-MET), (3) daily self-reports, (4) peer-to-peer support, and (5) Mindcotine support.

First, practice sessions in formal mindfulness included 6 sessions of mindfulness in video format of up to 10 minutes each and 7 sessions in audio format of 3 to 10 minutes each. These were based on the works of Kabat-Zinn [48] and Bowen and Marlatt [40], where the user is given an initial introduction to mindfulness involving the recognition of bodily sensations and the ability to practice nonreactivity to emotions and thoughts related to smoking, from a compassionate and nonjudgmental position.

Second, practice sessions in informal mindfulness using VR included 2 sessions of VR-MET, each lasting 10 minutes. These sessions included a selection of 2 virtual environments that combined the awareness of the act of smoking and the recognition of craving from a perspective of acceptance and commitment. This selection is the result of an adaptation of the previously conducted work synthesized by Brewer [49] and included virtual environments used in cue exposure therapy. Regarding VR-MET design, each of the virtual environments was created based on previous research that proved the environment to elicit craving [50] and was recorded with an Insta360Pro camera (Insta360) [51] in locations in Mexico City, Mexico, and Buenos Aires, Argentina. The animated parts of the environments were created in Unreal Engine 4 (Epic Games) and based on previous research to induce emotional states of tranquility and relaxation [52] (see Multimedia Appendix 2). The mindfulness audio from both environments was chosen to work consciously with craving-related acts (“RAIN”: Recognize, Accept, Investigate and Nourish, and “Act of Smoking”: consciously review each moment of the act itself). Each video was repeated over a period of 14 consecutive days, based on previous research regarding the time of exposure in virtual environments and the time between the exposures [30,34,37,53].

Third, daily self-reports were included. At the end of the day, each user reported on the app their total daily number of cigarettes, reasons that any cravings were triggered, and a written answer to the question “What do you think has changed in your relationship to smoking as of today?” (nightly reflections). This data was collected on the Typeform platform embedded in the app.

Fourth, peer-to-peer support was provided via the app’s group chat feature for interacting with all other participants. The group

XSL•FO
RenderX

chat was moderated by a psychologist and a mindfulness facilitator to promote engagement and respond to participant questions.

Fifth, Mindcotine support could be prompted. If participants were inactive for a certain amount of time, they received a text message (after 2 days) and a phone call (after 4 days) to encourage engagement within the program. Participants could contact Mindcotine for technical support by email anytime as well.

These components can be seen in Figure 3 (mindfulness exercises), Figure 4 (nightly reflection), Figure 5 (dashboard), and Figure 6 (VR-MET).

Figure 3. App screenshot of the home screen showing mindfulness activities.

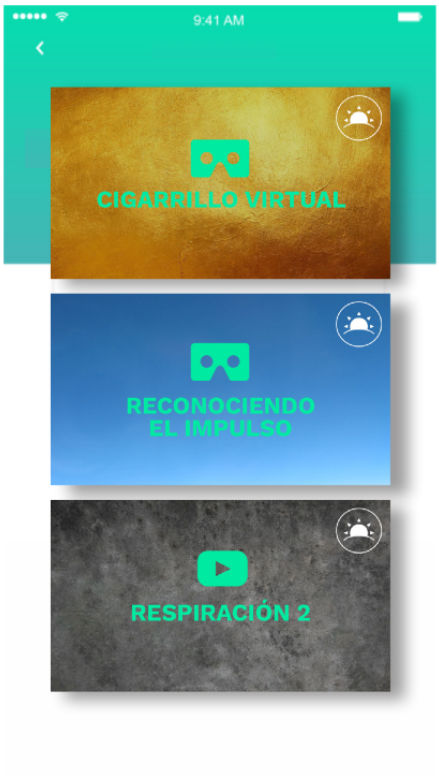


Figure 4. App screenshot of the nightly reflection.



XSL•FO
RenderX

Figure 5. App screenshot on the dashboard showing statistics about money and the number of smoked cigarettes.

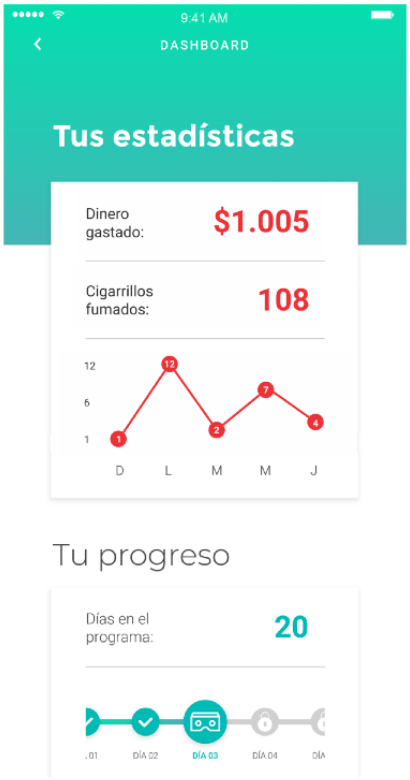


Figure 6. Virtual reality mindful exposure therapy screenshot: Act of Smoking.



Measures

Baseline Assessment

The baseline assessment was conducted on the app and included demographic data (age, gender, education), smoking behavior (number of cigarettes per day, years of smoking, previous cessation attempts, readiness to quit), characteristics of the environment and family context (marital status, children, living with smokers), and other health-related characteristics (body mass index, physical activity, previous meditation experience).

Primary Outcome

The primary outcome was self-reported abstinence at postintervention, assessed 1 day after the end of the program. The question used for assessment was “Did you smoke tobacco in the last 24 hours?”

XSL•FO
RenderX

Secondary Outcomes

Sustained Abstinence

Sustained abstinence was self-reported abstinence at 90 days after the end of the program in the TG. The question used for assessment was “Did you smoke tobacco in the last 90 days?”

Adherence

Adherence was assessed using log file data collected by employing the platform Amplitude (Amplitude Inc) [54], which measured total use of the app for each participant. Adherence to treatment was categorized as the percentage of participants who completed all daily activities and all daily diary questionnaires throughout the intervention with intervals of discontinuity no greater than 4 consecutive days. We operationalized full adherence as to whether or not the participant completed the treatment in the suggested time (ie, 21 days without any breaks), regular adherence as whether participants completed all daily activities in up to 60 days, and depth of adherence as the total number of mindfulness training minutes [55]. Chat room activity beyond the initial onboarding post was measured by the number of participant comments in the chat room and dichotomized to any activity versus no activity.

Cigarette Consumption

Participants were instructed to report the number of cigarettes they smoked each day during treatment as part of their daily self-reports, at postintervention, and at 90-day follow-up. Over the treatment period, the daily number of cigarettes was stored and averaged per week.

Craving

Craving was assessed using the Questionnaire for Smoking Urges (QSU) [56], which consists of a 7-point Likert scale. These data were collected through an online survey at baseline, at the end of days 7 and 14, and at postintervention. The internal consistency for the overall scale was adequate (α=.87).

Mindfulness

Mindfulness was assessed using the Five Facet Mindfulness Questionnaire (FFMQ) [57]. The 5 dimensions are observing, describing, acting with awareness, not judging internal experience, and not reacting to internal experience. These data were collected through an online survey at baseline, at postintervention, and at 90-day follow-up.

Readiness to Quit

Readiness to quit was assessed using the Contemplation Ladder [45], which consists of 11 rungs and 5 anchor statements reflecting the stages of change and was designed to measure readiness to quit smoking. It was assessed through an online survey at baseline and at postintervention.

Nicotine Dependence

Nicotine dependence was assessed using the Fagerström Test of nicotine dependence [58], which consists of a 6-item self-report scale and observes responses suggestive of physiological dependence on nicotine. These data were collected through an online survey at the beginning of the program, at the end of day 7, day 14, postintervention, and 90-day follow-up. The internal consistency for the overall scale was adequate (α=.81).

Data Analysis

Statistical analysis was performed with SPSS for Windows (version 22; IBM Corp). Abstinence rates between groups were compared using a chi-square test. Participants with missing data at follow-up were assumed to be smoking. Repeated measure analyses of variance (ANOVAs) were used to determine between-group differences and intervention effects for continuous variables. The group factor had 2 levels, corresponding to intervention and control group, and the time factor had 3 levels, corresponding to the 3 assessment points (baseline, postintervention, and 90-day follow-up). Post hoc analyses were conducted to determine significant pairwise comparisons. ANOVAs and post hoc tests were conducted to determine changes within the treatment group over time. All *t* tests in the study were 2-tailed.

Results

Recruitment and Participation in the Intervention

Participant baseline characteristics are displayed in Table 1. On average, participants in the sample were 43.25 years old (SD 10.06) and 47.5% (57/120) were female. The largest group in terms of completed educational level was high school (45/120, 37.5%), followed by university (43/120, 35.8%). On average, the sample started smoking 19.15 years ago (SD 12.35) and consumed an average of 10.77 (SD 5.47) cigarettes per day. Of all participants, 14.2% (17/120) lived with other smokers in their homes. The entire sample reported a moderated nicotine dependence index, based on the mean scores obtained in the Fagerström Nicotine Dependence Test [58], of 4.48 (SD 1.55). Only 40.8% (49/120) had practiced meditation at least once in their life. No statistically significant differences were observed between treatment and control group on any variables at baseline.

XSL•FO
RenderX

Table 1. Participant baseline characteristics (N=120).

Characteristics	Full sample (N=120)	Treatment group (n=60)	Control group (n=60)
Age (years), mean (SD)	43.25 (10.06)	44.06 (9.89)	42.44 (10.23)
Gender (female), n (%)	57 (48)	33 (55)	24 (41)
Education, n (%)			
PhD or higher	9 (8)	7 (11)	2 (5)
University	43 (36)	25 (38)	18 (30)
Undergraduate	10 (8)	3 (3)	7 (6)
Technical	6 (5)	3 (3)	3 (8)
High school	45 (38)	19 (28)	26 (44)
Primary school	7 (6)	3 (5)	4 (6)
Years since started smoking, mean (SD)	19.15 (12.35)	19.40 (12.28)	18.91 (12.42)
Current cigarette consumption (cigarettes per day), mean (SD)	10.77 (5.47)	11.09 (5.27)	10.45 (5.67)
At least 1 attempt to quit in the last year, n (%)	65 (54)	32 (53)	33 (55)
Live with other smokers, n (%)	17 (14)	10 (17)	7 (12)
Nicotine dependence (FTND[a]), mean (SD)	4.48 (1.55)	4.22 (1.51)	4.75 (1.56)
Craving (QSU[b]), mean (SD)	29.5 (9.67)	30.28 (11.72)	28.71 (7.08)
Readiness to quit (Contemplation Ladder), mean (SD)	6.68 (1.17)	6.53 (1.14)	6.83 (1.18)
Precontemplation, n (%)	5 (4)	2 (3)	3 (5)
Contemplation, n (%)	45 (38)	26 (43)	19 (32)
Preparation, n (%)	67 (56)	31 (52)	36 (60)
Action, n (%)	3 (3)	1 (2)	2 (3)
No experience in meditation, n (%)	71 (59)	33 (54)	38 (64)
Five Facet Mindfulness Questionnaire, mean (SD)	123.16 (8.73)	123.94 (9.55)	122.43 (7.83)

[a]FTND: Fagerström Test of Nicotine Dependence.

[b]QSU: Questionnaire for Smoking Urges.

Primary Outcomes

At postintervention, the TG reported 23% (14/60) abstinence on that day compared with 5% (3/60) of the CG. This difference was statistically significant (χ^2_1=8.3; *P*=.004).

Secondary Outcomes

Sustained Abstinence

The TG reported 33% (20/60) sustained abstinence on the 90-day follow-up, compared with 5% (3/60) of participants in the CG. Since only 20% (12/60) of participants in the CG completed the 90-day follow-up, we did not conduct a statistical comparison between groups at this follow-up time point.

Adherence Rates

Intervention adherence was analyzed only in the TG, and 93% (56/60) of participants finished the 21-day program. Of those who finished, 41% (23/56) were fully adherent to the program (ie, completed all daily sessions and nightly reflections 21 days in a row) and 59% (33/56) were regularly adherent (ie, completed the program in >21 days), completing the program in 28.56 days on average. At the postintervention, 30% (7/23) of fully adherent participants reported smoking abstinence, while only 21% (7/33) of regularly adherent participants were abstinent. At the 90-day follow-up, smoking abstinence rates were 39% (9/23) among fully adherent participants and 33% (11/33) among regularly adherent participants. A statistically significant difference in readiness to quit was found between the fully and the regularly adherent group at baseline (t_{55}=3.092; *P*=.003), with fully adherent participants reporting greater readiness to quit. No differences were found regarding nicotine dependence (t_{55}=1.206; *P*=.23) at baseline, nor regarding abstinence rates between the 2 adherence groups at postintervention (χ^2_1=6.1; *P*=.43) and at 90-day follow-up (χ^2_1=1.9; *P*=.66)

If participants had not engaged with the program for 2 subsequent days, they were contacted in order to improve engagement. A total of 92% (54/60) of participants were contacted through SMS text messaging once; 65% (39/60), twice; 53% (32/60), thrice; and 34% (20/60) were contacted through both SMS text messaging and phone call reminders. Depth of adherence for all participants, measured by the number of minutes of mindfulness training during the program (including virtual reality, video format, and audio format), was

XSL•FO
RenderX

259.05 minutes on average, with a maximum of 386 minutes and a minimum of 216 minutes.

For those who reported abstinence at postintervention, the average mindfulness minutes trained was 250.43 minutes, while the average was 261.93 minutes among those who continued smoking. Statistically significant differences were not found between groups (t_{55}=–1.291; *P*=.20).

Participant activity in the chat room beyond an initial introduction message suggested in the onboarding process was low and only 13 participants commented in the chat room beyond the onboarding. Chat room activity was not associated with smoking cessation outcomes, having only 1 participant that successfully quit smoking while being active in the chat.

Cigarette Consumption

Statistically significant differences in cigarettes per day over time were found between the TG and the CG ($F_{5,114}$=95.73; *P*<.001) in the third week of the intervention and at postintervention. At intervention week 3, the TG consumed significantly fewer cigarettes than the CG (TG: mean 6.92, SD 5.26; CG: mean 9.03, SD 5.42; *P*=.03). Significant differences in cigarettes per day between groups were also found at postintervention (TG: mean 5.07, SD 5.65; CG: mean 9.53, SD 0.56; *P*<.001).

Readiness to Quit

The means comparison in the Contemplation Ladder for the TG and the CG showed statistically significant differences between groups ($F_{2,113}$=4.55; *P*=.01). Post hoc comparisons revealed that there were no differences at baseline (TG: mean 6.53, SD 0.14; CG: mean 6.83, SD 0.18; *P*=.32), but there were at postintervention (TG: mean 7.71, SD 0.13; CG: mean 7.16, SD 0.13; *P*=.005).

Mean Differences Within the Treatment Group Over Time

Cigarette Consumption

Post hoc tests after ANOVA were conducted for cigarette consumption, and statistically significant differences over time were found within the TG ($F_{4,52}$=13.79; *P*<.001). Participants significantly reduced their cigarettes per day from baseline (mean 11.09, SD 5.27) to postintervention (mean 6.05, SD 5.67; *P*<.001), as well as from baseline to 90-day follow-up (mean 5.07, SD 5.65; *P*<.001).

Mindfulness

Moreover, according to paired *t* test comparisons for FFMQ scores, no statistically significant difference was found between baseline (mean 123.95, SD 9.55) and postintervention (mean 122.61, SD 6.31; t_{55}=0.772; *P*=.44). However, FFMQ at 90-day follow-up (mean 115.89, SD 12.30) was significantly lower than scores at baseline (t_{55}=5.594; *P*<.001) and postintervention (t_{55}=3.234; *P*=.002).

Craving

A significant reduction in self-reported craving over time was observed in the treatment group (baseline: mean 30.28, SD 11.72; intervention week 1: mean 30.01, SD 11.75; intervention week 2: mean 28.00, SD 13.00; postintervention: mean 26.00, SD 11.37). These differences over time were statistically significant ($F_{1,24}$=3.725; *P*=.005).

Discussion

Principal Findings

This is the first study to report results of a pilot trial testing a VR smartphone-based smoking cessation program using remote and self-assisted delivery. At postintervention, the intervention group had a significantly higher abstinence rate (14/60, 23% abstinence) compared with the control group, which received a smoking cessation manual (3/60, 5% abstinence). Moreover, the intervention resulted in a 33% (20/60) abstinence rate at 90-day follow-up and high levels of adherence and engagement. Findings suggest the potential efficacy of a smartphone-based VR intervention that combines exposure therapy and mindfulness for smoking cessation.

When compared with other smoking cessation app studies, abstinence rates observed in the current study were in a similar range. For instance, an 8-week single-arm trial of a smartphone app that delivered essential features of US clinical practice guidelines with personalization resulted in a 26% abstinence rate at 30-day follow-up [59]. Another single-arm trial based on acceptance and commitment therapy had abstinence rates of 33% at 7-day follow-up and 28% at 30-day follow-up [60]. These rates are similar to the 23% (14/60) and 33% (20/60) observed in our study and provide confidence for further development and testing.

With 93% (56/60) of participants finishing the 21-day program, the current study had a high completion and low dropout rate. In line with other studies, we used several strategies to enhance adherence, including SMS text messaging and phone call reminders [55]. Consistent with our findings, adherence rates were also above 80% in the above-mentioned 8-week single-arm clinical trial [59] and in a randomized controlled trial of a text messaging program [61]. Of all 56 TG participants who finished the program, 23 completed without any gaps in treatment adherence and were classified as fully adherent, while the other 33 participants took an average of 28.56 days to complete the program and were classified as regularly adherent. The fully adherent group reported greater readiness to quit at baseline. Abstinence at postintervention and 90-day follow-up was higher among fully adherent participants compared with regularly adherent participants. These findings suggest that repeated exposure to smoking-related cues in virtual environments alongside mindfulness practice on a consecutive daily basis may increase abstinence outcomes compared with sessions that are more spaced out over time. Thus, a more consistent and disciplined training using virtual reality mindfulness-based exposure therapy to both internal and external triggers over a shorter period of time may result in better outcomes. Based on other studies [62], the chat room activity was low probably due to the lack of promotion to increase engagement through tailored material by the moderators. Overall, the methods used to strengthen engagement have been shown to have value and can be further developed.

XSL•FO
RenderX

Our results also indicate that mindfulness scores measured by the FFMQ significantly decreased within the TG from baseline to 90-day follow-up. It is unclear what may have caused this decrease despite the promising smoking cessation outcomes of our intervention. Given that the current study is the first to combine virtual reality exposure therapy and mindfulness-based relapse prevention for smoking cessation, replication of these findings is needed. Moreover, future studies should further explore the mechanisms of action of this novel intervention approach.

Other studies using VR and exposure therapy in smoking cessation showed similar results in decreasing cue-induced craving, such as a randomized clinical trial that combined CBT and virtual reality cue exposure therapy (VR-CET), which found a significant reduction of cue-induced cravings after group-based sessions over the course of 6 months with 5 individual sessions of VR-CET [37]. In the current study, the virtual reality environments exposed participants to smoking-related cues and at the same time presented a mindfulness narrative based on relapse prevention. Therefore, the intervention simultaneously elicits cravings by means of VRET and provides the user with tools for reducing these cravings by means of mindfulness. Our results of QSU craving scores over time show that scores did not decrease during the first week of treatment, in which there was no exposure therapy. However, self-reported cravings decreased during the following 14 days of the program, in which VR-MET was delivered.

Feasibility of conducting the trial was demonstrated by the fact that we were able to recruit 120 participants in the current pilot trial. However, feasibility of collecting follow-up data from a control group needs to be improved. The delivery of the intervention was feasible and acceptable to participants; 41% (23/56) of intervention participants completed the entire intervention content in 21 days, as recommended, and the average adherence to mindfulness training was 259.05 minutes per participant. Feasibility of the VR component as a central element of the program was high, with all participants practicing at least 15 informal mindfulness practice sessions lasting 10 minutes each. Even though this intervention introduced VR and mindfulness training to a population that had almost never tried any of these approaches before, the intervention had high engagement rates, suggesting remote and self-administered VR can be used as a strategy for improving adherence to mHealth interventions. These findings are in line with those of existing studies, which have shown that VR can enhance treatment fidelity by having behavioral interventions delivered by a programmed avatar [21] and that VR environments offer existing opportunities to enhance a patient´s involvement in treatments [63].

Given that VR in mental health to date has predominantly been used in inpatient hospital environments [20], the recruitment method used in the current study to test the potential of VR as a self-administered and remote smoking cessation intervention on an outpatient population is worth mentioning. After appearing on national television, more than 1000 volunteers from all over Argentina registered within a few hours. The average age of these volunteers was 41 years, which was considerably older than we had anticipated. The fact that we were able to recruit this population for a smartphone-based smoking cessation study using a VR cardboard headset demonstrates that this intervention can be accessible to populations that do not need to be exceptionally tech savvy. In light of existing challenges to recruit participants for mHealth interventions [64], this particular recruitment strategy may prove valuable for future research.

The creation of Mindcotine involved a group of psychologists, psychiatrists, and other physicians, as well as developers, actors, and mindfulness facilitators, and took over 6 months. While initial development costs were high, the fact that this program can be delivered remotely and self-administered makes it a low-cost and accessible intervention to promote smoking cessation.

Limitations

The current study has several limitations. The follow-up time was relatively short, given that only between 3% and 5% of smokers remain abstinent within the first year of quitting [65]. Follow-up assessment time points were not identical across intervention and control groups. Before the current pilot trial, we did not know how long participants would take to complete the entire program, thus we selected a 30-day follow-up time point for the control group. Future research will adhere to a consistent follow-up assessment time point across intervention and control groups. Completion of the 90-day follow-up in the control group was low. It is possible that control group participants may not have been motivated to complete follow-up surveys due to not seeing changes in their behavior, not receiving a digital intervention, or not receiving any monetary incentive. Control group follow-up rates need to be improved in future trials of this intervention. Other limitations include that at baseline, participants were not regular users of VR, the virtual intervention content was not interactive, only an Android version of the intervention app was available, and the version of the intervention tested in the current study did not include features for the users to track cravings on the app and in virtual reality environments. These features are currently in development for future versions of the intervention. Moreover, abstinence at follow-up was self-reported and subsequent investigations should include biochemical verification of outcomes.

Finally, the face-to-face meeting and the onboarding process at the beginning of the program could have impacted the high adherence rates observed in this trial and online onboarding may work differently. Future studies should administer the entire program remotely.

Conclusion

Overall, our VR smartphone-based Mindcotine intervention to support smoking cessation showed great potential with regard to participant adherence and initial efficacy. These findings warrant testing the intervention in a fully powered randomized trial including longer follow-up intervals to investigate relapse prevention and biochemical verification of abstinence.

XSL•FO
RenderX

Acknowledgments

This research was supported financially by the University of Flores (internal protocol number 17EX04), by a grant received from the Argentine Ministry of Production (Seed Fund project number 3291022226), and through sponsorship by MindCotine Inc.

Authors' Contributions

EG and MRC participated in the study design. NR and CW participated in the data collection process. JU participated in the data analysis. MRC, EG, and JT participated in the interpretation of data. EG and JT wrote the first draft of the manuscript. All authors participated in revising the manuscript.

Conflicts of Interest

JT's contribution to this publication was as a member of the Advisory Board of MindCotine Inc. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflicts of interest policies. EG, NR, and CW have shares in MindCotine Inc, the company that developed the VR environments and the app for this study. JU and MRC have no conflicts of interest.

Multimedia Appendix 1
Instructional video. How to assemble the virtual reality headset.
[MP4 File (MP4 Video), 7676 KB - jmir_v22i7e17571_app1.mp4]

Multimedia Appendix 2
Brief VR-MET environment.
[MP4 File (MP4 Video), 26023 KB - jmir_v22i7e17571_app2.mp4]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2557 KB - jmir_v22i7e17571_app3.pdf]

References

1. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2017: monitoring tobacco use and prevention policies. World Health Organization. Geneva, Switzerland: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255874/9789241512824-eng.pdf [accessed 2020-06-29]
2. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. Addiction 2005 May;100(5):682-688. [doi: 10.1111/j.1360-0443.2005.01093.x] [Medline: 15847626]
3. Tufano JT, Karras BT. Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends. J Med Internet Res 2005 Dec 20;7(5):e58 [FREE Full text] [doi: 10.2196/jmir.7.5.e58] [Medline: 16403722]
4. Smith D, Griffiths E, Poole R, di Florio A, Barnes E, Kelly M, et al. Beating Bipolar: exploratory trial of a novel Internet-based psychoeducational treatment for bipolar disorder. Bipolar Disord 2011;13(5-6):571-577. [doi: 10.1111/j.1399-5618.2011.00949.x] [Medline: 22017225]
5. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796 [FREE Full text] [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
6. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
7. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective Behavioral Changes through a Digital mHealth App: Exploring the Impact of Hedonic Well-Being, Psychological Empowerment and Inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
8. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
9. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-84; discussion 485. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
10. Riley W, Obermayer J, Jean-Mary J. Internet and mobile phone text messaging intervention for college smokers. J Am Coll Health 2008;57(2):245-248. [doi: 10.3200/JACH.57.2.245-248] [Medline: 18809542]
11. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]

XSL•FO
RenderX

12. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
13. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
14. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
15. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 02;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
16. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
17. Zeng EY, Vilardaga R, Heffner JL, Mull KE, Bricker JB. Predictors of Utilization of a Novel Smoking Cessation Smartphone App. Telemed J E Health 2015 Dec;21(12):998-1004 [FREE Full text] [doi: 10.1089/tmj.2014.0232] [Medline: 26171733]
18. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
19. Bishop G, Fuchs H. Research directions in virtual environments. SIGGRAPH Comput Graph 1992 Aug;26(3):153-177. [doi: 10.1145/142413.142416]
20. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual Reality and Medical Inpatients: A Systematic Review of Randomized, Controlled Trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
21. Coons MJ, Roehrig M, Spring B. The potential of virtual reality technologies to improve adherence to weight loss behaviors. J Diabetes Sci Technol 2011 Mar 01;5(2):340-344 [FREE Full text] [doi: 10.1177/193229681100500221] [Medline: 21527103]
22. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). Washington, DC: American Psychiatric Publishing; 2013.
23. Baker TB, Piper ME, McCarthy DE, Majeskie MR, Fiore MC. Addiction motivation reformulated: an affective processing model of negative reinforcement. Psychol Rev 2004 Jan;111(1):33-51. [doi: 10.1037/0033-295X.111.1.33] [Medline: 14756584]
24. Ferguson SG, Shiffman S. The relevance and treatment of cue-induced cravings in tobacco dependence. J Subst Abuse Treat 2009 Apr;36(3):235-243. [doi: 10.1016/j.jsat.2008.06.005] [Medline: 18715743]
25. Conklin CA, Vella EJ, Joyce CJ, Salkeld RP, Perkins KA, Parzynski CS. Examining the relationship between cue-induced craving and actual smoking. Exp Clin Psychopharmacol 2015 Apr;23(2):90-96 [FREE Full text] [doi: 10.1037/a0038826] [Medline: 25730416]
26. Stevenson JG, Oliver JA, Hallyburton MB, Sweitzer MM, Conklin CA, McClernon FJ. Smoking environment cues reduce ability to resist smoking as measured by a delay to smoking task. Addict Behav 2017 Apr;67:49-52 [FREE Full text] [doi: 10.1016/j.addbeh.2016.12.007] [Medline: 28012965]
27. Bowen S, Chawla N, Collins SE, Witkiewitz K, Hsu S, Grow J, et al. Mindfulness-based relapse prevention for substance use disorders: a pilot efficacy trial. Subst Abus 2009;30(4):295-305 [FREE Full text] [doi: 10.1080/08897070903250084] [Medline: 19904665]
28. Brewer JA, Elwafi HM, Davis JH. Craving to quit: psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. Psychol Addict Behav 2013 Jun;27(2):366-379 [FREE Full text] [doi: 10.1037/a0028490] [Medline: 22642859]
29. Conklin C, Tiffany S. Applying extinction research and theory to cue-exposure addiction treatments. Addiction 2002 Mar;97(2):155-167. [doi: 10.1046/j.1360-0443.2002.00014.x] [Medline: 11860387]
30. Culbertson CS, Shulenberger S, De La Garza R, Newton TF, Brody AL. Virtual Reality Cue Exposure Therapy For The Treatment Of Tobacco Dependence. J Cyber Ther Rehabil 2012;5(1):57-64 [FREE Full text] [Medline: 25342999]
31. Hone-Blanchet A, Wensing T, Fecteau S. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. Front Hum Neurosci 2014;8:844 [FREE Full text] [doi: 10.3389/fnhum.2014.00844] [Medline: 25368571]
32. Powers MB, Emmelkamp PM. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. J Anxiety Disord 2008;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
33. Malbos E, Rapee RM, Kavakli M. A controlled study of agoraphobia and the independent effect of virtual reality exposure therapy. Aust N Z J Psychiatry 2013 Feb;47(2):160-168. [doi: 10.1177/0004867412453626] [Medline: 22790176]
34. Choi J, Park S, Lee J, Jung H, Lee H, Jin C, et al. The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: a preliminary study. Psychiatry Investig 2011 Jun;8(2):155-160 [FREE Full text] [doi: 10.4306/pi.2011.8.2.155] [Medline: 21852993]

XSL•FO
RenderX

35. Moon J, Lee J. Cue exposure treatment in a virtual environment to reduce nicotine craving: a functional MRI study. Cyberpsychol Behav 2009 Feb;12(1):43-45. [doi: 10.1089/cpb.2008.0032] [Medline: 19025463]
36. Pericot-Valverde I, García-Rodríguez O, Ferrer-García M, Secades-Villa R, Gutiérrez-Maldonado J. Virtual reality for smoking cessation: a case report. Stud Health Technol Inform 2012;181:292-296. [Medline: 22954874]
37. Pericot-Valverde I, Secades-Villa R, Gutiérrez-Maldonado J. A randomized clinical trial of cue exposure treatment through virtual reality for smoking cessation. J Subst Abuse Treat 2019 Jan;96:26-32. [doi: 10.1016/j.jsat.2018.10.003] [Medline: 30466545]
38. Westbrook C, Creswell JD, Tabibnia G, Julson E, Kober H, Tindle HA. Mindful attention reduces neural and self-reported cue-induced craving in smokers. Soc Cogn Affect Neurosci 2013 Jan;8(1):73-84 [FREE Full text] [doi: 10.1093/scan/nsr076] [Medline: 22114078]
39. Garrison KA, Pal P, Rojiani R, Dallery J, O'Malley SS, Brewer JA. A randomized controlled trial of smartphone-based mindfulness training for smoking cessation: a study protocol. BMC Psychiatry 2015 Apr 14;15:83 [FREE Full text] [doi: 10.1186/s12888-015-0468-z] [Medline: 25884648]
40. Bowen S, Marlatt A. Surfing the urge: brief mindfulness-based intervention for college student smokers. Psychol Addict Behav 2009 Dec;23(4):666-671. [doi: 10.1037/a0017127] [Medline: 20025372]
41. Brewer JA, Mallik S, Babuscio TA, Nich C, Johnson HE, Deleone CM, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. Drug Alcohol Depend 2011 Dec 01;119(1-2):72-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.05.027] [Medline: 21723049]
42. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
43. Cobertura Universal de Salud. Como le ayudara este manual a dejar de fumar? Ministerio de Salud Argentina. Argentina: Ministerio de Salud; 2016. URL: http://www.msal.gob.ar/images/stories/ryc/graficos/0000000584cnt-2017-05_manual-autoayuda-dejar-de-fumar.pdf [accessed 2017-05-01]
44. Secretaría de Políticas Integrales sobre Drogas de la Nación Argentina (Sedronar). Tabaco-Alcohol: Intensidad del consumo. Estudio Nacional en población de 12 a 65 años sobre consumo de sustancias psicoactivas. Sedronar: Secretaria de Políticas Integrales sobre Drogas de la Nación Argentina. Argentina: Observatorio Argentino de Drogas; 2017. URL: http://www.observatorio.gov.ar/media/k2/attachments/InformeZconsumoZdeZTabaco_1.pdf [accessed 2017-05-01]
45. Biener L, Abrams DB. The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. Health Psychol 1991;10(5):360-365. [doi: 10.1037//0278-6133.10.5.360] [Medline: 1935872]
46. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
47. TypeForm.: Typeform URL: http://www.typeform.com [accessed 2017-11-01]
48. Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. Gen Hosp Psychiatry 1982 Apr;4(1):33-47. [doi: 10.1016/0163-8343(82)90026-3] [Medline: 7042457]
49. Brewer J, Kabat-Zinn J. The Craving Mind: From Cigarettes to Smartphones to Love--Why We Get Hooked and How We Can Break Bad Habits. New Haven, CT: Yale Univerity Press; 2017:1-256.
50. Ferrer-García M, García-Rodríguez O, Gutiérrez-Maldonado J, Pericot-Valverde I, Secades-Villa R. Efficacy of virtual reality in triggering the craving to smoke: its relation to level of presence and nicotine dependence. Stud Health Technol Inform 2010;154:123-127. [Medline: 20543283]
51. Insta360.: Insta360 URL: http://www.insta360.com [accessed 2017-06-15]
52. Miralles I, Castilla D, Botella C. Desarrollo de un Entorno Virtual para la Inducción Emocional. In: I Congreso de la Sociedad Española para las Ciencias del Videojuego. 2014 Presented at: CEUR Workshop Proc; June 24, 2014; Barcelona, Spain.
53. Lee J, Lim Y, Graham SJ, Kim G, Wiederhold BK, Wiederhold MD, et al. Nicotine craving and cue exposure therapy by using virtual environments. Cyberpsychol Behav 2004 Dec;7(6):705-713. [doi: 10.1089/cpb.2004.7.705] [Medline: 15687806]
54. Amplitude.: Amplitude URL: http://www.amplitude.com [accessed 2017-11-15]
55. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: Adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
56. Cox LS, Tiffany ST, Christen AG. Evaluation of the brief questionnaire of smoking urges (QSU-brief) in laboratory and clinical settings. Nicotine Tob Res 2001 Feb;3(1):7-16. [doi: 10.1080/14622200020032051] [Medline: 11260806]
57. Anchorena M, Ghiglione F, Nader M. Psychometric Properties of the Five Facet Mindfulness Questionnaire in an Argentine Sample. Archives of Assessment Psychology 2017 Feb 23;7(1):29-41 [FREE Full text]
58. Heatherton T, Kozlowski L, Frecker R, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]

XSL•FO
RenderX

59. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, A Personalized Smartphone App for Smoking Cessation: Initial Evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
60. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 01;170:37-42 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
61. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
62. Thrul J, Klein AB, Ramo DE. Smoking Cessation Intervention on Facebook: Which Content Generates the Best Engagement? J Med Internet Res 2015 Nov 11;17(11):e244 [FREE Full text] [doi: 10.2196/jmir.4575] [Medline: 26561529]
63. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Aug;5(8):625-632 [FREE Full text] [doi: 10.1016/S2215-0366(18)30226-8] [Medline: 30007519]
64. Ahmed S, Sharmin M. Why do mobile phone-based smoking cessation interventions struggle and how can we make them more effective? iConference 2015 Proceedings 2015 Mar 15:2015 [FREE Full text]
65. Hughes J, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
CG: control group
FFMQ: Five Facet Mindfulness Questionnaire
mHealth: mobile health
QSU: Questionnaire for Smoking Urges
TG: treatment group
VR: virtual reality
VR-CET: virtual reality cue exposure therapy
VRET: virtual reality exposure therapy
VR-MET: virtual reality mindful exposure therapy

Edited by G Eysenbach; submitted 21.12.19; peer-reviewed by A Miloff, O Perski; comments to author 19.03.20; revised version received 08.04.20; accepted 03.06.20; published 29.07.20.

Please cite as:
Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR
Virtual Reality Smartphone-Based Intervention for Smoking Cessation: Pilot Randomized Controlled Trial on Initial Clinical Efficacy and Adherence
J Med Internet Res 2020;22(7):e17571
URL: http://www.jmir.org/2020/7/e17571/
doi:10.2196/17571
PMID:32723722

©Emilio Goldenhersch, Johannes Thrul, Joaquín Ungaretti, Nicolas Rosencovich, Cristian Waitman, Marcelo Rodriguez Ceberio. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Biased Sampling and Causal Estimation of Health-Related Information: Laboratory-Based Experimental Research

María Manuela Moreno-Fernández[1], PhD; Helena Matute[1], PhD

Departamento de Fundamentos y Métodos de la Psicología, Faculty of Psychology and Education, University of Deusto, Bilbao, Spain

Corresponding Author:
María Manuela Moreno-Fernández, PhD
Departamento de Fundamentos y Métodos de la Psicología
Faculty of Psychology and Education
University of Deusto
Avenida de las Universidades, 24
Bilbao, 48007
Spain
Phone: 34 944 139 000 ext 3229
Email: manuela.moreno@deusto.es

Abstract

Background: The internet is a relevant source of health-related information. The huge amount of information available on the internet forces users to engage in an active process of information selection. Previous research conducted in the field of experimental psychology showed that information selection itself may promote the development of erroneous beliefs, even if the information collected does not.

Objective: The aim of this study was to assess the relationship between information searching strategy (ie, which cues are used to guide information retrieval) and causal inferences about health while controlling for the effect of additional information features.

Methods: We adapted a standard laboratory task that has previously been used in research on contingency learning to mimic an information searching situation. Participants (N=193) were asked to gather information to determine whether a fictitious drug caused an allergic reaction. They collected individual pieces of evidence in order to support or reject the causal relationship between the two events by inspecting individual cases in which the drug was or was not used or in which the allergic reaction appeared or not. Thus, one group (cause group, n=105) was allowed to sample information based on the potential cause, whereas a second group (effect group, n=88) was allowed to sample information based on the effect. Although participants could select which medical records they wanted to check—cases in which the medicine was used or not (in the cause group) or cases in which the effect appeared or not (in the effect group)—they all received similar evidence that indicated the absence of a causal link between the drug and the reaction. After observing 40 cases, they estimated the drug–allergic reaction causal relationship.

Results: Participants used different strategies for collecting information. In some cases, participants displayed a biased sampling strategy compatible with positive testing, that is, they required a high proportion of evidence in which the drug was administered (in the cause group) or in which the allergic reaction appeared (in the effect group). Biased strategies produced an overrepresentation of certain pieces of evidence at the detriment of the representation of others, which was associated with the accuracy of causal inferences. Thus, how the information was collected (sampling strategy) demonstrated a significant effect on causal inferences ($F_{1,185}$=32.53, P<.001, η^{2p}=0.15) suggesting that inferences of the causal relationship between events are related to how the information is gathered.

Conclusions: Mistaken beliefs about health may arise from accurate pieces of information partially because of the way in which information is collected. Patient or person autonomy in gathering health information through the internet, for instance, may contribute to the development of false beliefs from accurate pieces of information because search strategies can be biased.

(J Med Internet Res 2020;22(7):e17502) doi:10.2196/17502

KEYWORDS

information sampling; causal illusion; causal bias; health information; health communication

XSL•FO
RenderX

Introduction

Background

In recent years, we have witnessed an increase in diseases, such as measles, that had almost been eradicated in the Western world [1]. One of the factors underlying this outbreak is the antivaccination movement, led by individuals who do not adhere to recommendations for vaccinations (for themselves or their children). Reluctance with respect to vaccination led to the World Health Organization listing vaccine hesitancy as 1 of 10 threats to global health in 2019 [2]. Although vaccine hesitancy is a complex phenomenon [3], common reasons for refusing vaccination are the underestimation of its benefits or the overestimation of its negative side effects; a high proportion of these concerns are based on information disseminated by the media or received from acquaintances [4]. Moreover, previous research [5] has shown that people use the internet as an information source about vaccines, and that side effects and possible negative outcomes of vaccination are one of the most searched topics.

In fact, the internet is an important source of health-related information [6-11]. Newly diagnosed cancer patients perceive the internet as a tool for acquiring information and for making informed decisions [12-14]; patients with diabetes use the internet to seek general information about the disease or about treatment options [15]. The relevant role of the internet, in this context, is not restricted to pathological states. Women who are pregnant use the internet to get informed about topics such as fetal development or to make pregnancy-related decisions [6,16], and after pregnancy, parents use the internet to retrieve health information regarding their infants [8,17].

The internet has accessibility, anonymity, and interactivity as advantages, but these advantages do not come without risks [18]. Some of these risks, such as information overload or lack of credibility, can be considered intrinsic limitations of the internet as a source for health information; however, there are other risks that are based on human skills and cognitive abilities. For example, the huge amount of information available on the internet forces users to engage in an active process of information selection to filter content. As we will discuss later, this information sampling process may play a crucial part in establishing and maintaining mistaken beliefs.

Evaluating the Risks and Benefits of Medical Treatments

When people judge the risks and benefits of a treatment option, they infer the causal relationship between two events—the treatment and its effect. Unfortunately, causal inferences of this kind are highly difficult because, among other reasons, causality is not directly observable; rather, it must be inferred from cues such as contingency [19,20]. The principle of contingency posits that, unless hidden factors are at play, all causes correlate statistically with their effects.

Consider the simplest case in which a person may try to judge the effect of only one factor (cause) on one given outcome (effect). For the sake of simplicity, it can be assumed that both cause, *C*, and effect, *E*, are binary variables—they either occur or do not occur. In this situation, the person can collect information that fits into one of the following four possibilities (Figure 1, panel A)—*type a*, in which both the cause and the effect occur; *type b*, in which only the cause occurs; *type c*, in which only the effect but not the cause occurs; and *type d*, in which neither the cause nor the effect occurs—which define the cause and effect contingencies. Although different indices have been proposed to represent contingency [21,22], the most popular is, perhaps, the Δp index [23] which is computed as the difference between the probability of the effect conditional on occurrence of the cause, $E|C$, and the probability of the effect conditional on absence of the cause, $E|{\sim}C$, as shown in the equation,

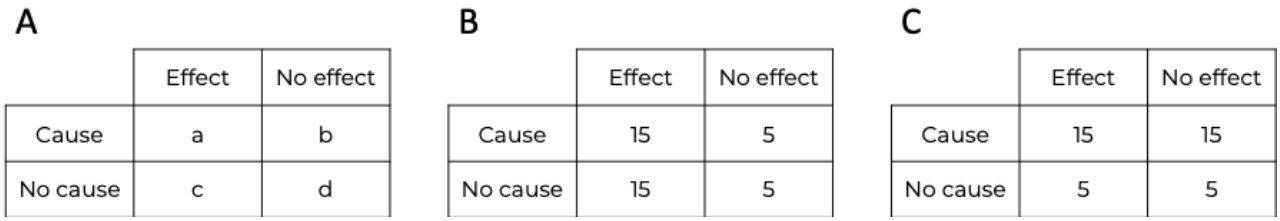
$$\Delta p = p(E|C) - p(E|{\sim}C) = a/(a + b) - c/(c + d)$$

where Δp can take on values between –1 and 1. Positive values indicate a generative relationship, and negative values indicate a preventive relationship. When the potential cause and the effect are not related to each other, the index equals zero and contingency is null.

Previous research [23-28] has shown that people are sensitive to contingency between events, and that contingency is used as a cue to make causal inferences; however, under some circumstances, people systematically deviate from the normative standard. Researchers have described two systematic deviations: the influence of the probability of effect occurrence [29-34], when the effect occurs frequently, the causal relationship tends to be overestimated (Figure 1, panel B); and the influence of the probability of occurrence of the cause [24,35,36], when the probability of the cause is high, the contingency perceived between cause and effect is also high (Figure 1, panel C). These biases can be detected even if the contingency between the cause and the effect is null, leading to causal illusions [37].

XSL•FO
RenderX

Figure 1. Contingency matrices where (A) shows the four information types as a function of whether the cause and the effect are present, (B) shows an example with a high probability of the effect with null contingency, and (C) shows an example with a high probability of the cause with null contingency.



A

	Effect	No effect
Cause	a	b
No cause	c	d

B

	Effect	No effect
Cause	15	5
No cause	15	5

C

	Effect	No effect
Cause	15	15
No cause	5	5

Health-Related Information Seeking and Causal Illusions

Imagine someone who is worried about the potential relationship between a vaccine and autism. Usually, it is not possible to evaluate the effects of the vaccine by administering and not administering the vaccine and observing the outcomes since people rarely have the opportunity to conduct a randomized controlled trial. Instead, people will search for information about the relationship between the vaccine and the side effect by consulting an expert, by consulting a friend, or by searching the internet.

Perhaps the most obvious concern about internet-sourced information is the lack of quality control. Internet users may come across and trust information that is not supported by evidence; however, the act of information seeking may entail additional and specific concerns. For example, people who worry about the safety of vaccination and its relationship with autism may look for information about the vaccine (exploring its side effects and the probability of experiencing those side effects, etc). If they are concerned about autism, they may focus their search on autism (exploring which factors have been related with the development of autism or what the proportion of vaccinated children is among those who were diagnosed with autism). In the former, cause (ie, the vaccine) is the cue that guides the search, while in the latter, effect (ie, autism) is the cue that guides the search.

The sampling strategy (how people search for information) will affect their final inference about the relationship between vaccination and autism. If people search for information about the vaccine (ie, the potential cause of autism), they may introduce the name of the vaccine on the search engine, and they will mostly retrieve instances of type a and type b information. This information will allow them to make a general estimation of the probability of the effect (ie, autism) when the potential cause (ie, the vaccine) has been presented; however, in this case, their sampling strategy is biased toward the cause, and therefore, no information about the effect in the absence of the cause, that of either type c or type d, is collected. This will eventually bias their judgements. Indeed, even when sampling is not completely biased toward the cause and some type c or type d instances are collected, it has been repeatedly shown that the higher the tendency to sample information about the cause, the higher the probability of overestimating the cause–effect relationship [37-39].

In our example about the effects of vaccination, this strategy may, nevertheless, be considered as not particularly dangerous. The prevalence of autism spectrum disorders is actually low (1 out of 160 children) [40]; therefore, this sampling strategy will retrieve more type b information than that of type a. In this example, the low base rate of the effect may protect people from developing a causal illusion [41], but in other cases, this protection does not exist (imagine, for example, potential effects such as nausea, high temperature, headache, or any other common effect).

As previously noted, people may also gather information using the effect as their cue for sampling; they may search using terms related to the effect rather than those related to the cause. If people use this sampling strategy for collecting information, they may learn which factors have been associated with autism and will discover that many children among those who developed autism spectrum disorders had been vaccinated. Thus, information sampling will be biased, overrepresenting information in which the effect (ie, autism) is present. In the long run, this strategy will increase the proportion of type a and type c information (relative to that of type b and type d) and will favor a sampling-induced overestimation of the relationship (Figure 1, panel B). This sampling-induced illusion may explain how concerned and educated parents end up overestimating the potential risks of vaccination [42]. Since many countries usually have systematic immunization programs, when information sampling is biased toward effect, the probability of collecting type a information (ie, cases in which autism and vaccination coincide) is even higher than the probability of collecting type c information (ie, cases in which autism occurs in the absence of vaccination). This increases the probability of overestimating the link between vaccination and autism—ie, the probability of experiencing a causal illusion.

The vaccination and autism example illustrates quite well how information sampling may become a crucial element for establishing and maintaining mistaken beliefs; however, the biases in sampling strategies can be extended to a wide range of health issues; people interested in assessing the causal relationship between any common behavior and an infrequent disease will find a high proportion of information where the behavior and the disease coincide if they use the effect as a cue in their internet search. Correspondingly, people using the cause to guide their internet search may end up neglecting the base rate of the effect and end up overestimating the causal relationship when the effect is frequent [43]. For example, a recent study [44] which tracked internet-browsing behavior in a controlled setting showed that when women were required to consult the internet for health information after the hypothetical onset of an unfamiliar breast change (eg, nipple rash), most

XSL•FO
RenderX

participants used rash-related search terms (a cue-guided sampling strategy), and the majority accessed websites containing breast cancer information with National Health Service Paget disease of the nipple being the most visited site. In this situation, even when information is accurate, the potential relationship between both events could be overestimated at a substantial emotional cost. Note that, if considering no other information, a nipple rash may be produced by other skin conditions with high incidence rates (such as eczema) rather than by Paget disease, which is a rare type of breast cancer [45].

Information sampling biases may also affect inferences about treatment effectiveness; it is also possible for a biased sampling strategy to induce a perception that underestimates a treatment that is actually effective, or a perception that overestimates the effectiveness of alternative practices proven by clinical trial to perform no better than placebos [39,46].

Study Goals

As previously described, research on contingency learning has demonstrated how people may use different pieces of information to infer causal relationships [19,20,47]. We have also mentioned that individual behavior may bias information sampling, and consequently, causal inferences; however, these behavior-induced causal illusions have only been explored in situations in which the potential cause was used to guide information sampling. As far as we know, the influence of an equivalent effect-driven sampling has not yet been explored. In addition, these causal illusions have been explored using procedures that usually include motivational components or additional goals which may affect information gathering and causal inferences. For example, participants may be required to evaluate the effects of a fictitious medicine while at the same time trying to heal as many patients as possible [48,49]. It is not clear whether a causal illusion can be detected in an information sampling setting when motivational components or secondary goals are removed, or when sampling strategies are guided by the effect rather than by the cause. The research reported herein explores these two possibilities.

Methods

Participants

A sample of adults (N=193) with a mean age of 34.07 (SD 11.41) years consisting of women (83/193, 43.0%), men (109/193, 56.5%), and one (1/193, 0.5%) nonbinary participant were recruited via Prolific Academic internet platform [50]. They were compensated £0.75 (US $0.93) for their participation, which worked out to approximately £5.01 ($6.19) per hour. Enrollment was only offered to individuals in Prolific Academic's pool whose first language was English (to ensure that instructions were correctly understood) and to individuals who had not taken part in previous studies carried out by our research team. We did not use any exclusion criteria (all participants were included in reporting). Participants were randomly assigned to experimental groups—a cause group (n=105) and an effect group (n=88).

Instruments

Because of ethical considerations, we avoided using a real-world example in this experiment and instead used a simplified fictitious scenario that is often used in causal learning research; we adapted the allergy task [20,51] for presentation as a web app based on World Wide Web Consortium standards [52]. A demonstration of the program can be downloaded from the Open Science Framework [53].

Experimental Design and Procedure

We adapted the allergy task [20,51] to make it akin to an information gathering situation with no goal other than that of assessing the causal link between two events. This procedure has been widely used in causal learning research and allows for the assessment of causal illusions while controlling other relevant parameters, which ultimately ensures a high degree of internal validity.

The procedure was set to allow for cause-driven and effect-driven sampling by including a reduced number of changes. Participants were required to gather information to discover whether a fictitious drug (ie, "Batatrim") caused a fictitious allergic reaction (ie, "Lindsay syndrome"). Participants in the cause group were allowed to sample information based on the potential cause, therefore, could choose to retrieve patient medical records based on patient treatment (whether the patient was treated with Batatrim or not treated with Batatrim), whereas participants in the effect group were allowed to sample information based on the effect, therefore, could choose to retrieve the records based on the development of the syndrome (whether the patient developed Lindsay syndrome while hospitalized or did not). Detailed instructions for the task can be found in Multimedia Appendix 1.

The probability of the cause was under participant control in the cause group whereas the probability of the effect was under participant control in the effect group. Information presented objectively reflected the absence of a causal link between the two events. The probability of the effect in cause group and the probability of the cause in effect group were fixed to 0.75. This design allowed for the evaluation of sampling-dependent causal illusion in a null contingency situation with a high probability of the outcome (cause group) and with a high probability of the cause (effect group).

We expected causal estimations to vary depending on sampling strategy. For participants in the cause group, we expected estimations to increase as the participants increased the probability of the cause—the more biased the participant behavior toward cause-present events, the higher the probability of experiencing causal illusion. Analogously, for participants in the effect group, we expected estimations to increase as participants increased the probability of the effect, also resulting in an increased likelihood of experiencing causal illusion. The procedure used in this study was approved by the ethical review board of the University of Deusto.

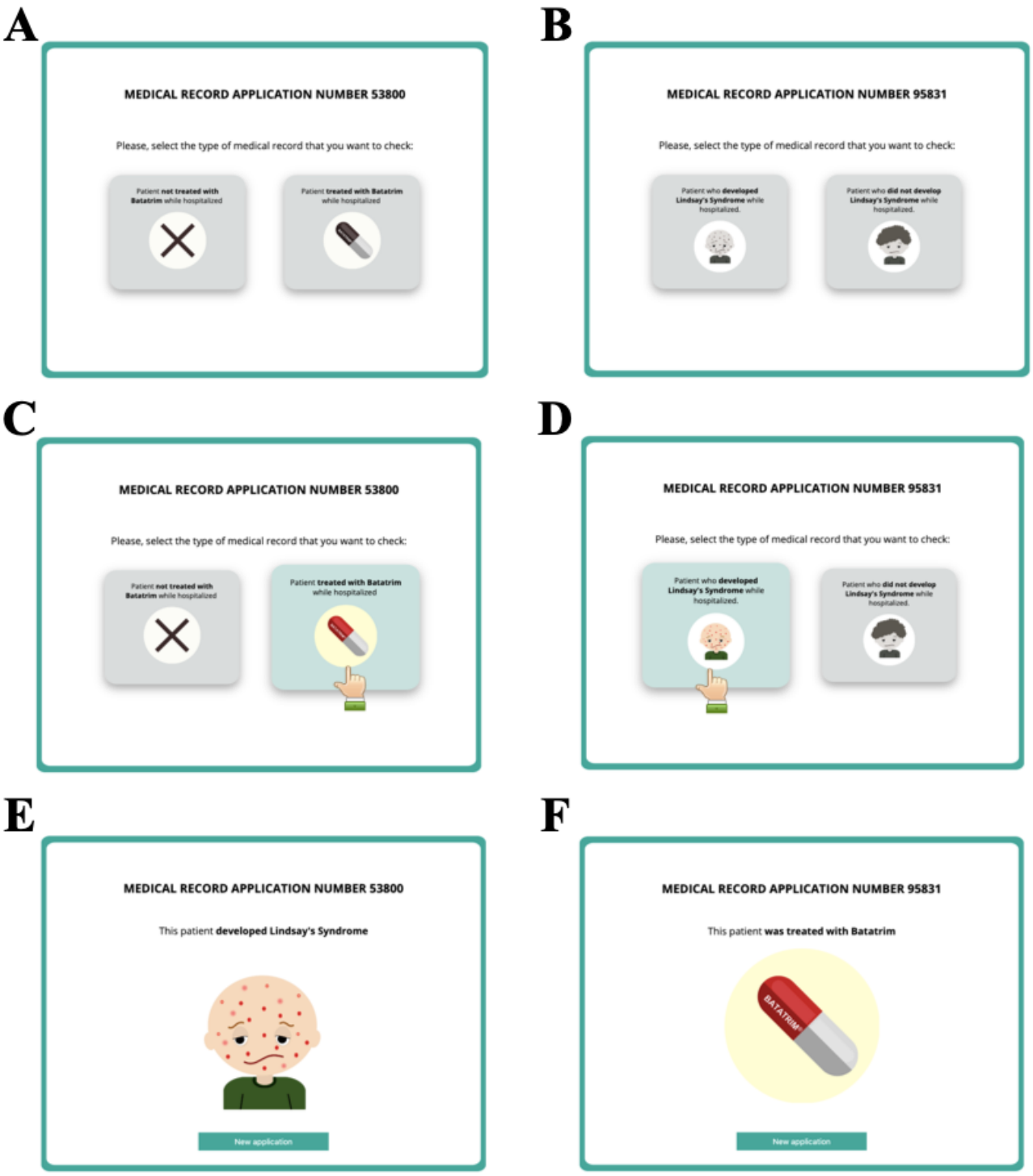
Learning Task

A series of 40 patient records was presented, each in a separate trial. Each trial started with a screen on which participants were

XSL•FO
RenderX

required to indicate which type of medical record they wanted to check by pressing one of two buttons (Figure 2, panels A and B). The location of each button (left or right) was randomized for each participant. When participants moved the cursor over the buttons, the button colored and zoomed, and a hand pointer appeared to indicate that a response could be made (Figure 2, panels C and D). Once a button was clicked, all the information presented on the screen was removed (with the exception of the sentence stating the medical record application number which remained for esthetic purposes, and for which a random number was used). The information that had been removed was replaced with information about the syndrome when viewed by those in the cause group, or about the treatment when viewed by those in the effect group (Figure 2, panels E and F). One second later, a green button with the words "New application" was presented. After clicking this button, all elements were removed, and after one second, a new trial (a new patient record) appeared. There was no time limit for progressing through the task.

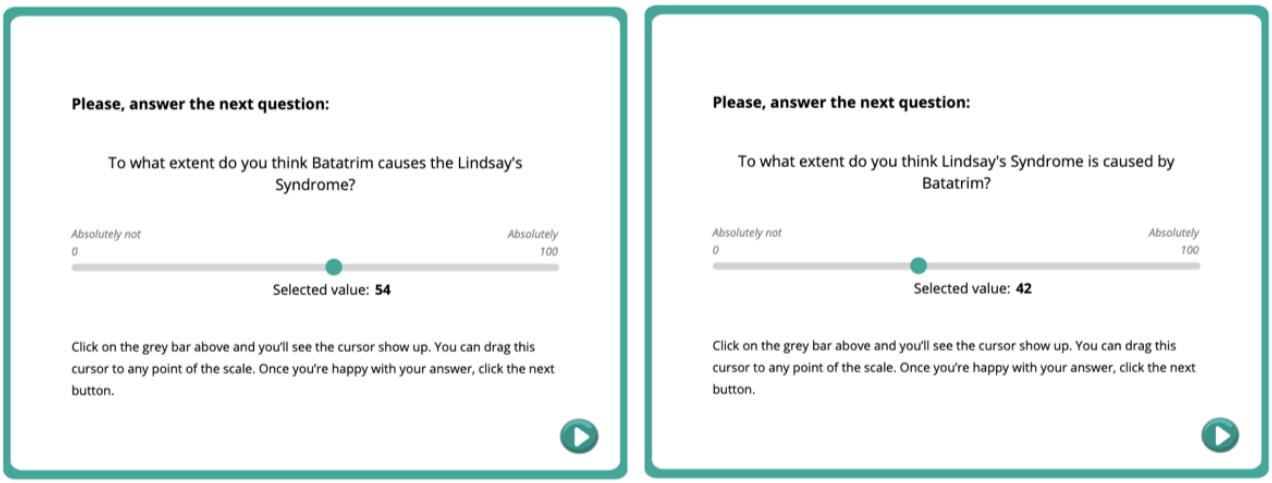
Figure 2. The sequence of events within a trial presented to the cause group (panels on the left) and to the effect group (panels on the right).



XSL•FO
RenderX

When participants completed the training stage, they were required to use a 100-point scale to make a global estimation about the causal relationship between Batatrim and Lindsay Syndrome. The question was formulated in the direction of either cause-to-effect "To what extent do you think Batatrim causes the Lindsay syndrome?" or effect-to-cause "To what extent do you think Lindsay syndrome is caused by Batatrim?" The format of the question was randomized among participants (Figure 3). In both cases, the response scale was from 0 (absolutely not) to 100 (absolutely).

Figure 3. The final screen where the causal relation between Batatrim and Lindsay syndrome is assessed. The question is shown worded as cause-to-effect (left) or effect-to-cause (right).



Measures

Causal estimations at the end of the learning task were used as a measure of causal inference [39]. For each participant, we also calculated a sampling strategy index and a measure of experienced contingency.

As previously described, participants in the cause group could choose records based on patient treatment (patients treated or not treated with Batatrim), while participants in the effect group could choose records based on the development of the syndrome (patients who developed or who did not develop the syndrome). Participants could display an unbiased information gathering strategy, asking for a similar proportion of records in both categories; however, it was also possible for participants to display a biased sampling strategy, ie, to preferentially ask for one type of medical record more often than for the other. To measure bias in the information sampling strategy, we calculated a sampling strategy index from training responses as the probability of choosing records of patients treated with Batatrim (in the cause group) or records of patients who developed Lindsay syndrome (in effect group); therefore, the sampling strategy index could range between 0 and 1. Values near 1 indicated a strong preference for checking the medical records of patients treated with Batatrim or patients who developed the syndrome (depending on the group). Values near 0 indicated the opposite strategy, that is, a preference for checking medical records of patients who were not treated with Batatrim or who did not develop the syndrome. A value of 0.5 indicated an unbiased strategy with no preference for either of the two strategies. The higher the index, the higher the probability of retrieving a medical record in which the potential cause and the consequence coincide (type a information), and consequently, the higher the probability of developing a causal illusion.

Additionally, and given that participants could decide which type of medical record they wanted to check, experienced contingency could depart from the programmed value (Δp=0) and also affect their causal estimations [35,54]; therefore, a measure of experienced contingency was calculated (Δp using the actual number of type a, b, c, and d trials to which each participant was exposed).

Statistical Analysis

Unless noted otherwise, P<.05 was deemed as statistically significant. Two-tailed independent t tests were used to determine if sampling strategy indices were significantly different from 0.50 (neutral strategy) in either group.

A 2×2 analysis of variance (ANOVA) was performed to assess the effect of group (cause versus effect) and button position (left versus right) on information sampling strategy (sampling strategy index).

An analysis of covariance (ANCOVA) was performed using group (cause or effect) and directionality in which the causal estimation was required (cause-to-effect or effect-to-cause) as fixed factors and information sampling strategy (sampling strategy index) as a covariate to determine the effect on causal estimation. We expected causal estimation to vary as a function of sampling strategy index in both groups—the higher the index, the stronger the causal overestimation of the relationship between the cause and the effect. Additionally, and since the probability of the outcome in the cause group and the probability of the cause in the effect group were fixed at the same high rate, $p(C)=p(E)=0.75$, we explored whether the effect of sampling strategy on causal estimations was equivalent in both groups. We also explored if causal estimations were affected by directionality.

A t test was used to compare experienced contingencies with the programmed value (Δp=0). In participants who exhibited an extremely biased strategy by checking only records of patients treated with Batatrim (ie, a sampling strategy index

XSL•FO
RenderX

equal to 1), no trials without the cause were sampled, and the probability of the effect in the absence of the cause $p(E|\sim C)=0$, and consequently experienced contingency, could not be computed since $c/(c + d)=0/0$.

To explore sampling strategies, learning phase data were split into 8 blocks of 5 trials, and a sampling strategy index was calculated for each block. A repeated measures ANOVA was used to explore the effect of block (from 1 to 8) and group (cause and effect) on sampling strategy index. Posthoc analyses (28 comparisons) were performed using Bonferroni correction.

Results

Sampling strategy indices were significantly different from 0.50 (neutral strategy) in both groups. Participants preferentially checked the medical records of patients who, in the case of the cause group, were treated with Batatrim (mean 0.54, SD 0.17; $t_{104}=2.18$, $P=.03$, Cohen $d=0.21$), or who, in the case of the effect group, developed the syndrome (mean 0.57, SD 0.16; $t_{87}=4.07$, $P<.001$, Cohen $d=0.43$).

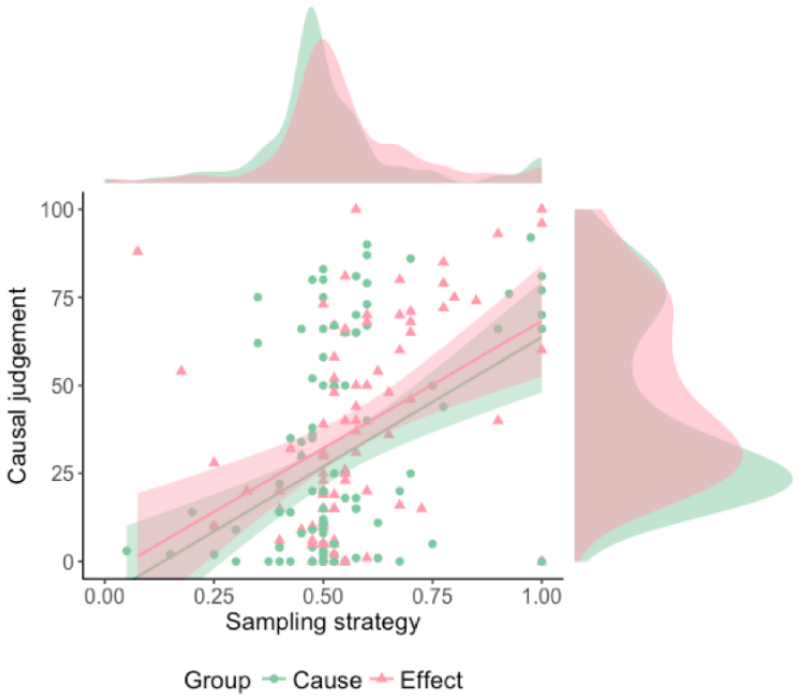
The 2×2 ANOVA demonstrated that the sampling strategy index did not differ between groups ($F_{1,189}=2.30$, $P=.13$, $\eta^2_p=0.01$), and it was not affected by the position in which buttons were presented ($F_{1,189}=1.31$, $P=.26$, $\eta^2_p=0.01$). The group × button position interaction also was not significant ($F_{1,189}=0.14$, $P=.71$, $\eta^2_p=0$); therefore, both groups showed a similar sampling strategy, selecting the medical records of patients who were exposed to the potential cause or who suffered the effect more often than the medical records of patients who were not exposed to the potential cause or who did not suffer the effect.

Only a significant effect of sampling strategy index ($F_{1,185}=32.53$, $P<.001$, $\eta^2_p=.15$) was demonstrated by the ANCOVA, suggesting that the relationship between information searching strategy and causal estimation was independent of group and directionality (Table 1 and Figure 4).

Table 1. Summary of ANCOVA analysis for variables predicting causal estimations.

Effect	*F* test (df_1,df_2)	*P* value	Partial eta square
Sampling strategy index	32.53 (1,185)	<.001	0.15
Directionality	0.61 (1,185)	.44	0
Group	0.20 (1,185)	.66	0
Sampling strategy index × directionality	0.22 (1,185)	.64	0
Sampling strategy index × group	0.01 (1,185)	.93	0
Directionality × group	1.63 (1,185)	.20	0.01
Sampling strategy index × directionality × group	1.60 (1,185)	.21	0.01

Figure 4. Causal estimations as a function of sampling strategy index and group.



XSL•FO
RenderX

Note that any analysis that included experienced contingency did not take into account 5 participants who exhibited an extremely biased strategy (sampling strategy index=1).

No differences were found between experienced contingency and the programmed value (p=0) either in the cause group (t_{99}=1.49, P=.14, Cohen d=0.15) or in the effect group (t_{87}=0.54, P=.59, Cohen d=0.06) meaning that most participants experienced a near zero contingency. Once the 5 participants for whom it was not possible to calculate Δp were discarded, no relationship between sampling strategy and experienced contingency was detected (r=0.06, P=.41); therefore, the effect of sampling strategy on causal estimation could not be attributed to experienced contingency.

The repeated measures ANOVA showed a significant effect of block ($F_{7,1337}$=5.24, P<.001, η^2_p =.027). The sampling strategy index was significantly higher in block 1 than in the other seven blocks, while no other significant differences were found (Table 2 and Figure 5). In block 1, 93% (98/105) of participants in the cause group selected the medical record of a patient treated with Batatrim as their first choice, and similarly, 86% (76/88) of participants in the effect group selected the medical record of a patient who developed the syndrome.

Table 2. Posthoc comparisons.

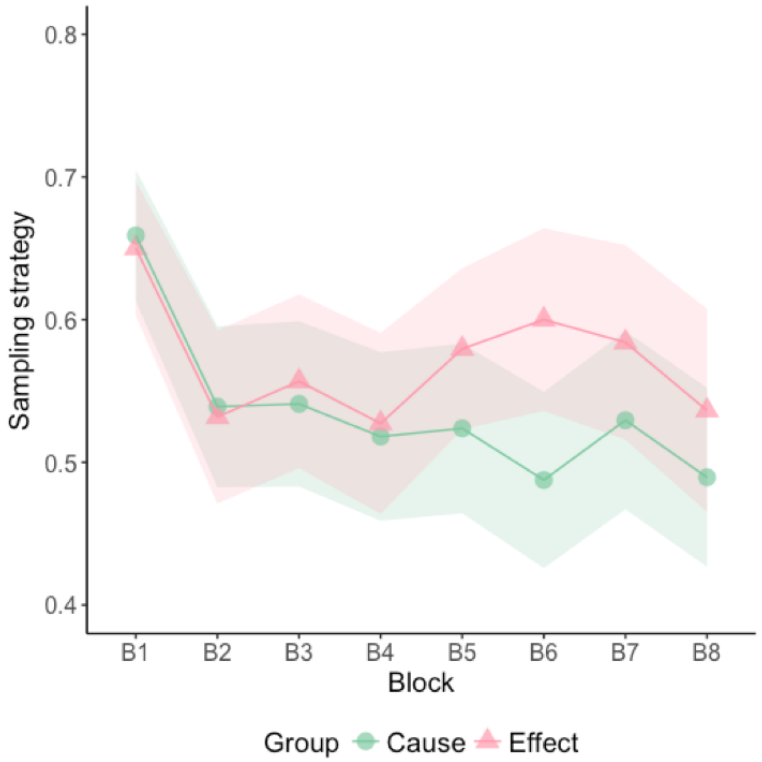
Comparison	Mean difference	*t* test (*df*)	*P* value (Bonferroni[a])	*P* value (uncorrected[b])
Block 1 - block 2	0.12	4.43 (1337)	<.001	<.001
Block 1 - block 3	0.11	3.93 (1337)	.003	<.001
Block 1 - block 4	0.13	4.90 (1337)	<.001	<.001
Block 1 - block 5	0.10	3.82 (1337)	.004	<.001
Block 1 - block 6	0.11	4.12 (1337)	.001	<.001
Block 1 - block 7	0.10	3.63 (1337)	.008	<.001
Block 1 - block 8	0.14	5.27 (1337)	<.001	<.001
Block 2 - block 3	–0.01	–0.50 (1337)	>.999	.62
Block 2 - block 4	0.01	0.47 (1337)	>.999	.64
Block 2 - block 5	–0.02	–0.60 (1337)	>.999	.55
Block 2 - block 6	–0.01	–0.31 (1337)	>.999	.76
Block 2 - block 7	–0.02	–0.79 (1337)	>.999	.43
Block 2 - block 8	0.02	0.84 (1337)	>.999	.40
Block 3 - block 4	0.03	0.97 (1337)	>.999	.33
Block 3 - block 5	–0.00	–0.10 (1337)	>.999	.92
Block 3 - block 6	0.01	0.19 (1337)	>.999	.85
Block 3 - block 7	–0.01	–0.29 (1337)	>.999	.77
Block 3 - block 8	0.04	1.34 (1337)	>.999	.18
Block 4 - block 5	–0.03	–1.08 (1337)	>.999	.28
Block 4 - block 6	–0.02	–0.79 (1337)	>.999	.43
Block 4 - block 7	–0.03	–1.27 (1337)	>.999	.21
Block 4 - block 8	0.01	0.36 (1337)	>.999	.72
Block 5 - block 6	0.01	0.29 (1337)	>.999	.77
Block 5 - block 7	–0.01	–0.19 (1337)	>.999	.85
Block 5 - block 8	0.04	1.44 (1337)	>.999	.15
Block 6 - block 7	–0.01	–0.48 (1337)	>.999	.63
Block 6 - block 8	0.03	1.15 (1337)	>.999	.25
Block 7 - block 8	0.04	1.63 (1337)	>.999	.10

[a]Bonferroni corrected values; statistically significant when P<.05.

[b]Uncorrected values; statistically significant when P<.002.

XSL•FO
RenderX

Figure 5. Mean sampling strategy index for each block of 5 trials for the cause group and for the effect group. Ribbons depict 95% CI.



Discussion

Principal Results

The main goal of this experiment was to assess the potential relationship between information gathering biases and causal inferences using an experimental procedure. Thus, we adapted a standard laboratory task which has previously been used in research on causal illusions in order to imitate an information searching situation. Results showed a significant a relationship between causal illusion and information sampling strategy. When the potential cause was used to collect information, the causal link can be overestimated when cause-absent information is undersampled. Similarly, when the effect is the cue that drives information gathering, causal estimations can be overestimated when effect-absent information is insufficiently sampled.

Although we did not explicitly include any manipulation aimed at biasing sampling strategy, we found a general preference for checking the medical records of patients treated with Batatrim or patients who developed the syndrome (depending on the group). We may explain this preference as the result of a positive testing strategy driven by a confirmation bias [55]. Instructions presented the treatment with Batatrim as a potential cause for the allergic reaction: "You suspect that Lindsay syndrome may be caused by a medical treatment called Batatrim..." Consequently, we provided participants with the initial hypothesis that Batatrim caused Lindsay syndrome. People using a positive testing strategy will search for information that confirms their hypothesis. Under the initial hypothesis that Batatrim causes Lindsay syndrome, a positive testing strategy involved searching information to obtain coinciding events. When a search is based on the cause, the strategy that allows for retrieving cause-effect coincidences is to select cases in which the cause is present (ie, medical records of patients treated with Batatrim) whereas the way to obtain these coincidences when searching is driven by the effect is by selecting cases in which the effect occurred (ie, medical records of patients who developed the syndrome). These two biases resemble the sampling strategy bias detected in our experiment. Our results about the relationship between information searching strategy and causal estimation (a significant effect of sampling strategy index on causal estimation) also showed the danger associated with this testing strategy—information collected using a positive testing strategy will led to an overrepresentation of cue-present trials and will increase the likelihood of a causal illusion.

Since a positive testing strategy has been claimed to be a general default heuristic that is often used in the absence of specific information identifying some tests as more relevant than others [55], it is not necessary to assume that a confirmation bias support our results; however, previous beliefs should be taken into account when the information sampling strategies are tested in real contexts given that personal interest and motivation may exert a heavy influence boosting the effect of a default-biased strategy. The role of confirmation bias has already been explored in health information sampling research suggesting that it may significantly affect how information is collected. In recent research, Meppelink et al [56] investigated the role of confirmation bias in information seeking with respect to early-childhood vaccination and found that a priori vaccination beliefs biased selection of online health information—people predominantly selected information that was consistent with their existing beliefs (ie, selective exposure) [57]. The significant effect of sampling strategy index on causal estimations showed that, in addition, a partial selection of information in which belief-supporting evidence is overrepresented may be related

XSL•FO
RenderX

to damaged causal estimation (note that the correlational nature of our design does not allow us to discriminate whether the biased searching caused the estimates to be biased, or whether a stronger initial belief about the causal relationship might have biased the sampling strategy, strengthening the initial belief).

Limitations and Strengths

In order to ensure strong experimental control of the variables involved and a high internal validity, we decided to use a standard and very simple procedure that is often used to study how people make causal inferences in laboratory settings. Our use of this procedure in the current situation, however, resulted in limitations related to its ecological validity. Our procedure does not exactly mirror how internet users search for information. Most internet users do not sequentially select information about individual people in the same way as was done in the experimental task; however, the process of collecting information and the subsequent processes of integration, combination, and interpretation of the information are, fundamentally, the same.

Participants searched for and collected pieces of evidence that ultimately were used to shape their estimation about the relationship between the events. Similarly, internet users may use web search engines which provide them with discrete bits of information that are used as evidence to support or reject the causal relationship between the events under assessment. In our experiment, these pieces of information were less enriched than those collected in real-world settings, but they did contain the core information needed for causal inference. Reducing the ecological validity of our procedure ensured a high level of internal validity—an advantage that made our procedure a better option than other naturalistic paradigms. The most relevant advantage was that it allowed us to explore information sampling biases while controlling for the effect of additional information features.

Real situations contain a high degree of ambiguity and subtle information nuances that may limit research inferences by weakening internal validity. For example, website design or the perceived authority of the author have been shown to influence the trust and credibility of web-based health information [58], which may affect how specific pieces of information are weighted and integrated to make causal inferences. Instead, the experimental approach allowed for the isolation of searching strategy from other factors. Another advantage of the experimental approach was that we were able to control which information was presented, and consequently, whether the information objectively supported any relationship between the cause and the alleged effect in order to detect causal illusions. Finally, by using a fictitious scenario instead of a real-world example, we avoided the potential consequences of experimentally induced causal beliefs on real-life decisions and controlled for any contribution of a priori beliefs.

Future Work

Now that the contribution of sampling strategies on causal inferences has been documented in a laboratory setting, future research may extend our results to real-world situations to assess the generalizability of our findings when information collection is more complex. This research may be considered a first step in building interventions aimed at protecting people when using the internet to search health information.

Conclusions

The internet has become a relevant source of health-related information [6-8]. Despite its advantages, using the internet to gather information requires several considerations such as the lack of quality control of the information and the subsequent possibility of misinformation dissemination. A relevant example is misinformation concerning scientific strategies that are aimed at protecting and promoting public health such as vaccination. Although, the determinants of vaccine uptake are complex, online misinformation has been claimed to contribute to the phenomenon of individuals foregoing vaccinations [59,60] and major search engines and social media organizations have been recently called to actively support fact-based communication programs that positively contribute to restoring confidence in vaccinations [61]. Using the internet to gather health information may cause additional concerns beyond those of information quality. How people search for information may determine which information is retrieved [62,63], shaping their beliefs about health, and eventually, their health-related behavior, such as vaccination refusal [64,65]. Results from our study have shown that sampling biases are related to causal perceptions. Thus, partial selection of information may induce an uneven representation of information that may produce and perpetuate causal illusions.

Laboratory-based research on contingency learning has been shown to be a successful approach to real-life problems because of its ability to detect relevant factors that may contribute to causal inferences, but also because it has been the foundation for designing and testing evidence-based interventions that have proven to be effective in improving critical thinking skills, and therefore, at reducing potentially harmful causal misconceptions in real contexts [48,49]. Future research may extend our results to real-world contexts in order to design interventions aimed at protecting users when using the internet.

Acknowledgments

Support for this research was provided by grants RTI2018-096700-J-I00 and PSI2016-78818-R from *Agencia Estatal de Investigación* of the Spanish government awarded to MMMF and to HM, respectively, and Grant IT955-16 from the Basque Government awarded to HM.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Detailed instructions used in the allergy task.
[DOCX File , 39 KB - jmir_v22i7e17502_app1.docx]

References

1. New measles surveillance data for 2019. World Health Organization. 2019. URL: https://www.who.int/immunization/newsroom/measles-data-2019/en/ [accessed 2020-06-23]
2. Ten threats to global health in 2019 internet. World Health Organization. 2019. URL: https://www.who.int/emergencies/ten-threats-to-global-health-in-2019 [accessed 2020-06-23]
3. Hickler B, Guirguis S, Obregon R. Vaccine special issue on vaccine hesitancy. Vaccine 2015 Aug 14;33(34):4155-4156 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.034] [Medline: 25896381]
4. McKee C, Bohannon K. Exploring the reasons behind parental refusal of vaccines. J Pediatr Pharmacol Ther 2016;21(2):104-109 [FREE Full text] [doi: 10.5863/1551-6776-21.2.104] [Medline: 27199617]
5. Harmsen IA, Doorman GG, Mollema L, Ruiter RAC, Kok G, de Melker HE. Parental information-seeking behaviour in childhood vaccinations. BMC Public Health 2013 Dec 21;13:1219 [FREE Full text] [doi: 10.1186/1471-2458-13-1219] [Medline: 24358990]
6. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
7. Betsch C. Innovations in communication: the Internet and the psychology of vaccination decisions. Euro Surveill 2011 Apr 28;16(17):19849 [FREE Full text] [Medline: 21543043]
8. Bianco A, Zucco R, Nobile CGA, Pileggi C, Pavia M. Parents seeking health-related information on the internet: cross-sectional study. J Med Internet Res 2013 Sep 18;15(9):e204 [FREE Full text] [doi: 10.2196/jmir.2752] [Medline: 24047937]
9. Haluza D, Naszay M, Stockinger A, Jungwirth D. Digital natives versus digital immigrants: influence of online health information seeking on the doctor-patient relationship. Health Commun 2017 Nov;32(11):1342-1349. [doi: 10.1080/10410236.2016.1220044] [Medline: 27710132]
10. Medlock S, Eslami S, Askari M, Arts DL, Sent D, de RSE, et al. Health information-seeking behavior of seniors who use the internet: a survey. J Med Internet Res 2015 Jan;17(1):e10 [FREE Full text] [doi: 10.2196/jmir.3749] [Medline: 25574815]
11. Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, et al. Online health information–seeking among older women with chronic illness: analysis of the women's health initiative. J Med Internet Res 2020 Apr 9;22(4):e15906. [doi: 10.2196/15906] [Medline: 32271152]
12. Bass SB, Ruzek SB, Gordon TF, Fleisher L, McKeown-Conn N, Moore D. Relationship of internet health information use with patient behavior and self-efficacy: experiences of newly diagnosed cancer patients who contact the National Cancer Institute's Cancer Information Service. J Health Commun 2006 Mar;11(2):219-236. [doi: 10.1080/10810730500526794] [Medline: 16537289]
13. Nguyen S, Ingledew PA. Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. J Cancer Educ 2013;28(4):662-668. [doi: 10.1007/s13187-013-0509-6] [Medline: 23832398]
14. Zhu C, Zeng R, Zhang W, Evans R, He R. Pregnancy-related information seeking and sharing in the social media era among expectant mothers: qualitative study. J Med Internet Res 2019 Dec 4;21(12):e13694. [doi: 10.2196/13694] [Medline: 31799939]
15. Shaw RJ, Johnson C. Health information seeking and social media use on the internet among people with diabetes. Online J Public Health Inform 2011;3(1):ojphi.v3i1.3561 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
16. Lagan B, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
17. Moon RY, Mathews A, Oden R, Carlin R. Mothers' perceptions of the internet and social media as sources of parenting and health information: qualitative study. J Med Internet Res 2019 Jul 09;21(7):e14289 [FREE Full text] [doi: 10.2196/14289] [Medline: 31290403]
18. Cline R, Haynes K. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
19. Shanks DR, Dickinson A. Associative accounts of causality judgment. Psychol Learn Motiv 1988;21:229-261. [doi: 10.1016/S0079-7421(08)60030-4]
20. Wasserman EA. Detecting response-outcome relations: toward an understanding of the causal texture of the environment. Psychol Learn Motiv 1990;26:27-82. [doi: 10.1016/s0079-7421(08)60051-1]
21. Perales J, Catena A, Cándido A, Maldonado A. Rules of causal judgment: Mapping statistical information onto causal beliefs. In: The Oxford Handbook of Causal Reasoning. Oxford: University Press; 2017:29-51.
22. Perales JC, Shanks DR. Models of covariation-based causal judgment: a review and synthesis. Psychon Bull Rev 2007 Aug;14(4):577-596. [doi: 10.3758/bf03196807] [Medline: 17972719]
23. Allan LG. A note on measurement of contingency between two binary variables in judgment tasks. Bull. Psychon. Soc 2013 Nov 7;15(3):147-149. [doi: 10.3758/bf03334492]

XSL•FO
RenderX

24. Allan LG, Jenkins H. The effect of representations of binary variables on judgment of influence. Learn Motiv 1983 Nov;14(4):381-405. [doi: 10.1016/0023-9690(83)90024-3]
25. Chatlosh D, Neunaber D, Wasserman E. Response-outcome contingency: behavioral and judgmental effects of appetitive and aversive outcomes with college students. Learn Motiv 1985 Feb;16(1):1-34. [doi: 10.1016/0023-9690(85)90002-5]
26. Dickinson A, Shanks D, Evenden J. Judgement of act-outcome contingency: the role of selective attribution. Q J Exp Psychol A 2007 May 29;36(1):29-50. [doi: 10.1080/14640748408401502]
27. Shanks DR. Selective attribution and the judgment of causality. Learn Motiv 1986 Nov;17(4):311-334. [doi: 10.1016/0023-9690(86)90001-9]
28. Ward WC, Jenkins HM. The display of information and the judgment of contingency. Can J Psychol 1965 Sep;19(3):231-241. [doi: 10.1037/h0082908] [Medline: 5824962]
29. Alloy L, Abramson L. Judgment of contingency in depressed and nondepressed students: sadder but wiser? J Exp Psychol Gen 1979;108(4):441-485. [doi: 10.1037/0096-3445.108.4.441]
30. Murphy RA, Schmeer S, Vallée-Tourangeau F, Mondragón E, Hilton D. Making the illusory correlation effect appear and then disappear: the effects of increased learning. Q J Exp Psychol (Hove) 2011 Jan;64(1):24-40. [doi: 10.1080/17470218.2010.493615] [Medline: 20623441]
31. Wasserman E, Elek S, Chatlosh D, Baker A. Rating causal relations: Role of probability in judgments of response-outcome contingency. J Exp Psychol Learn Mem Cogn 1993;19(1):174-188 [FREE Full text] [doi: 10.1037/0278-7393.19.1.174]
32. Moreno-Fernández MM, Blanco F, Matute H. Causal illusions in children when the outcome is frequent. PLoS One 2017;12(9):e0184707 [FREE Full text] [doi: 10.1371/journal.pone.0184707] [Medline: 28898294]
33. Chow J, Colagiuri B, Livesey E. Bridging the divide between causal illusions in the laboratory and the real world: the effects of outcome density with a variable continuous outcome. Cogn Res Princ Implic 2019 Jan 28;4(1):1 [FREE Full text] [doi: 10.1186/s41235-018-0149-9] [Medline: 30693393]
34. Blanco F, Matute H. Diseases that resolve spontaneously can increase the belief that ineffective treatments work. Soc Sci Med 2020;255:113012 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113012] [Medline: 32387871]
35. Blanco F, Matute H, Vadillo MA. Making the uncontrollable seem controllable: the role of action in the illusion of control. Q J Exp Psychol (Hove) 2011 Jul;64(7):1290-1304. [doi: 10.1080/17470218.2011.552727] [Medline: 21432736]
36. Perales JC, Catena A, Shanks DR, González JA. Dissociation between judgments and outcome-expectancy measures in covariation learning: a signal detection theory approach. J Exp Psychol Learn Mem Cogn 2005 Sep;31(5):1105-1120. [doi: 10.1037/0278-7393.31.5.1105] [Medline: 16248753]
37. Matute H, Blanco F, Díaz-Lago M. Learning mechanisms underlying accurate and biased contingency judgments. J Exp Psychol Anim Learn Cogn 2019 Oct;45(4):373-389. [doi: 10.1037/xan0000222] [Medline: 31380677]
38. Blanco F, Matute H, Vadillo MA. Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. Learn Behav 2013 Dec;41(4):333-340. [doi: 10.3758/s13420-013-0108-8] [Medline: 23529636]
39. Matute H, Blanco F, Yarritu I, Díaz-Lago M, Vadillo MA, Barberia I. Illusions of causality: how they bias our everyday thinking and how they could be reduced. Front Psychol 2015;6:888 [FREE Full text] [doi: 10.3389/fpsyg.2015.00888] [Medline: 26191014]
40. Autism spectrum disorders. World Health Organization. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders [accessed 2020-06-23]
41. Blanco F, Matute H. Base-rate expectations modulate the causal illusion. PLoS One 2019;14(3):e0212615 [FREE Full text] [doi: 10.1371/journal.pone.0212615] [Medline: 30835775]
42. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J. Internet exposure associated with canadian parents' perception of risk on childhood immunization: cross-sectional study. JMIR Public Health Surveill 2018 Jan 19;4(1):e7. [doi: 10.2196/publichealth.8921] [Medline: 29351896]
43. Matute H, Yarritu I, Vadillo MA. Illusions of causality at the heart of pseudoscience. Br J Psychol 2011 Aug;102(3):392-405. [doi: 10.1348/000712610X532210] [Medline: 21751996]
44. Marcu A, Muller C, Ream E, Whitaker KL. Online information-seeking about potential breast cancer symptoms: capturing online behavior with an internet browsing tracking tool. J Med Internet Res 2019 Feb 06;21(2):e12400 [FREE Full text] [doi: 10.2196/12400] [Medline: 30724741]
45. Karakas C. Paget's disease of the breast. J Carcinog 2011;10:31 [FREE Full text] [doi: 10.4103/1477-3163.90676] [Medline: 22279416]
46. Lilienfeld SO, Ritschel LA, Lynn SJ, Cautin RL, Latzman RD. Why ineffective psychotherapies appear to work: a taxonomy of causes of spurious therapeutic effectiveness. Perspect Psychol Sci 2014 Jul;9(4):355-387. [doi: 10.1177/1745691614535216] [Medline: 26173271]
47. Johansen M, Osman M. Coincidences: a fundamental consequence of rational cognition. New Ideas Psychol 2015 Oct;39:34-44 [FREE Full text] [doi: 10.1016/j.newideapsych.2015.07.001]
48. Barberia I, Blanco F, Cubillas CP, Matute H. Implementation and assessment of an intervention to debias adolescents against causal illusions. PLoS One 2013;8(8):e71303 [FREE Full text] [doi: 10.1371/journal.pone.0071303] [Medline: 23967189]

XSL•FO
RenderX

49. Barberia I, Tubau E, Matute H, Rodríguez-Ferreiro J. A short educational intervention diminishes causal illusions and specific paranormal beliefs in undergraduates. PLoS One 2018;13(1):e0191907 [FREE Full text] [doi: 10.1371/journal.pone.0191907] [Medline: 29385184]
50. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. J Behav Exp Finance 2018;17:22-27.
51. Blanco F, Barberia I, Matute H. Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. PLoS One 2015;10(7):e0131378 [FREE Full text] [doi: 10.1371/journal.pone.0131378] [Medline: 26177025]
52. About W3C. World Wide Web Consortium. 2020. URL: https://www.w3.org/Consortium/ [accessed 2020-06-23]
53. Moreno-Fernández MM, Matute H. data and materials:. Open Science Framework. 2020. URL: https://osf.io/bcfeq/ [accessed 2020-06-23]
54. Hannah SD, Beneteau JL. Just tell me what to do: bringing back experimenter control in active contingency tasks with the command-performance procedure and finding cue density effects along the way. Can J Exp Psychol 2009 Mar;63(1):59-73. [doi: 10.1037/a0013403] [Medline: 19271817]
55. Klayman J, Ha Y. Confirmation, disconfirmation, and information in hypothesis testing. Psychol Rev 1987;94(2):211-228. [doi: 10.1037/0033-295x.94.2.211]
56. Meppelink C, Smit E, Fransen M, Diviani N. "I was right about vaccination": confirmation bias and health literacy in online health information seeking. J Health Commun 2019;24(2):129-140. [doi: 10.1080/10810730.2019.1583701] [Medline: 30895889]
57. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
58. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
59. Larson HJ. The state of vaccine confidence. The Lancet 2018 Nov;392(10161):2244-2246. [doi: 10.1016/s0140-6736(18)32608-4] [Medline: 30496111]
60. Larson HJ. Negotiating vaccine acceptance in an era of reluctance. Hum Vaccin Immunother 2013 Aug;9(8):1779-1781 [FREE Full text] [doi: 10.4161/hv.25932] [Medline: 23896582]
61. Ratzan SC, Bloom BR, El-Mohandes A, Fielding J, Gostin LO, Hodge JG, et al. The salzburg statement on vaccination acceptance. J Health Commun 2019;24(5):581-583. [doi: 10.1080/10810730.2019.1622611] [Medline: 31262227]
62. Ruiz JB, Bell RA. Understanding vaccination resistance: vaccine search term selection bias and the valence of retrieved information. Vaccine 2014 Oct 07;32(44):5776-5780. [doi: 10.1016/j.vaccine.2014.08.042] [Medline: 25176640]
63. Wolfe RM, Sharp LK. Vaccination or immunization? the impact of search terms on the internet. J Health Commun 2005 Sep;10(6):537-551. [doi: 10.1080/10810730500228847] [Medline: 16203632]
64. Olive J, Hotez P, Damania A, Nolan M. The state of the antivaccine movement in the United States: a focused examination of nonmedical exemptions in states and counties. PLoS Med 2018 Jun 12;15(6):e1002578. [doi: 10.1371/journal.pmed.1002578]
65. Rottman B, Marcum Z, Thorpe C, Gellad W. Medication adherence as a learning process: insights from cognitive psychology. Health Psychol Rev 2017 Mar;11(1):17-32. [doi: 10.1080/17437199.2016.1240624] [Medline: 27707099]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by B Rottman, B Bie, D Bowen; comments to author 10.03.20; revised version received 03.05.20; accepted 14.05.20; published 24.07.20.

Please cite as:
Moreno-Fernández MM, Matute H
Biased Sampling and Causal Estimation of Health-Related Information: Laboratory-Based Experimental Research
J Med Internet Res 2020;22(7):e17502
URL: http://www.jmir.org/2020/7/e17502/
doi:10.2196/17502
PMID:32706735

©María Manuela Moreno-Fernández, Helena Matute. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any

XSL•FO
RenderX

medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Crawling the German Health Web: Exploratory Study and Graph Analysis

Richard Zowalla[1,2], MSc; Thomas Wetter[3,4,5], Dr rer nat, Dipl Math; Daniel Pfeifer[1,2], Dr Ing

[1]Department of Medical Informatics, Heilbronn University, Heilbronn, Germany

[2]Center for Machine Learning, Heilbronn University, Heilbronn, Germany

[3]Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany

[4]Institute of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel

[5]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

Corresponding Author:
Richard Zowalla, MSc
Department of Medical Informatics
Heilbronn University
Max-Planck-Str 39
Heilbronn,
Germany
Phone: 49 713 150 46791
Email: richard.zowalla@hs-heilbronn.de

Abstract

Background: The internet has become an increasingly important resource for health information. However, with a growing amount of web pages, it is nearly impossible for humans to manually keep track of evolving and continuously changing content in the health domain. To better understand the nature of all web-based health information as given in a specific language, it is important to identify (1) information hubs for the health domain, (2) content providers of high prestige, and (3) important topics and trends in the health-related web. In this context, an automatic web crawling approach can provide the necessary data for a computational and statistical analysis to answer (1) to (3).

Objective: This study demonstrates the suitability of a focused crawler for the acquisition of the German Health Web (GHW) which includes all health-related web content of the three mostly German speaking countries Germany, Austria and Switzerland. Based on the gathered data, we provide a preliminary analysis of the GHW's graph structure covering its size, most important content providers and a ratio of public to private stakeholders. In addition, we provide our experiences in building and operating such a highly scalable crawler.

Methods: A support vector machine classifier was trained on a large data set acquired from various German content providers to distinguish between health-related and non–health-related web pages. The classifier was evaluated using accuracy, recall and precision on an 80/20 training/test split (TD1) and against a crowd-validated data set (TD2). To implement the crawler, we extended the open-source framework StormCrawler. The actual crawl was conducted for 227 days. The crawler was evaluated by using harvest rate and its recall was estimated using a seed-target approach.

Results: In total, n=22,405 seed URLs with country-code top level domains .de: 85.36% (19,126/22,405), .at: 6.83% (1530/22,405), .ch: 7.81% (1749/22,405), were collected from Curlie and a previous crawl. The text classifier achieved an accuracy on TD1 of 0.937 (TD2=0.966), a precision on TD1 of 0.934 (TD2=0.954) and a recall on TD1 of 0.944 (TD2=0.989). The crawl yields 13.5 million presumably relevant and 119.5 million nonrelevant web pages. The average harvest rate was 19.76%; recall was 0.821 (4105/5000 targets found). The resulting host-aggregated graph contains 215,372 nodes and 403,175 edges (network diameter=25; average path length=6.466; average degree=1.872; average in-degree=1.892; average out-degree=1.845; modularity=0.723). Among the 25 top-ranked pages for each country (according to PageRank), 40% (30/75) were web sites published by public institutions. 25% (19/75) were published by nonprofit organizations and 35% (26/75) by private organizations or individuals.

Conclusions: The results indicate, that the presented crawler is a suitable method for acquiring a large fraction of the GHW. As desired, the computed statistical data allows for determining major information hubs and important content providers on the GHW. In the future, the acquired data may be used to assess important topics and trends but also to build health-specific search engines.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e17853) doi:10.2196/17853

KEYWORDS

health information; internet; web crawling; distributed system

Introduction

Overview

The internet has become an increasingly important resource for health information, especially for laypeople [1-10]. Web users perform online searches to obtain health information regarding diseases, diagnoses, and different treatments [1]. However, with a growing amount of web pages, it is nearly impossible for humans to manually keep track of evolving and continuously changing content in the health domain. According to the (German) Good Practice Guidelines for Health Information, "evidence-based health information is...a trustworthy state of the medical knowledge" [11]. Even if health information is found via well-known search engines, it does not necessarily meet with this definition and may be influenced by commercial interests [12].

Therefore, it is important to identify health content providers and assess their relevance [13]. In this context, an automatic web crawling approach can help to understand the structure of the health-related web (ie all web pages offering health-related information). By focusing only on such content, it is possible to (1) identify information hubs for the health domain, (2) find content providers of high prestige, and (3) identify important topics and trends within the health-related web. In future work, the identified content providers of high prestige could be analyzed for their respective trustworthiness and their compliance with the criteria of evidence-based health information [11].

According to Van der Bosch et al [14], in 2015 the (indexed) web was estimated to consist of roughly 47 billion web pages. However, only a fraction of those web pages contain health-related information. So, in order to determine the structure of the health-related web, it is crucial to determine for each web page's content whether it is health-related or not.

A related filter method can be used within a web crawler to filter out irrelevant web pages, therefore reducing the total number of web pages that need to be crawled. This saves time and financial resources for the crawling task. Nevertheless, analyzing such an amount of data requires high performance hardware and parallelization approaches.

Yet, to the best of the authors' knowledge, no study has been previously conducted and published about the health-related web. This study provides a first analysis of the health-related web limited to web pages in German, the so-called German health web (GHW). In this regard, we restrict our study to the three mostly German-speaking countries Germany, Austria and Switzerland (D-A-CH).

A distributed focused crawler for the GHW is outlined and evaluated as part of this study. Using the acquired data it is possible to extract the graph structure of the GHW for the goals listed above and provide access to health-related text material for linguistic analysis and further research purposes.

Related Work

Importance of Health Information on the Web

The World Wide Web and its graph structure have been a subject of study for many years [15-17]. However, domain-specific and/or country-dependent analysis of graph properties have not been the primary scope of research in the recent years [18,19]. Moreover, a review by Kumar et al [20] shows that research related to focused crawling was popular in the late 1990s and mid 2000s but seems to have lost attention in the last decade. As the internet is an important resource for health information [21], finding relevant content remains an important task [8].

Web Crawling of Health Information

In 2005, Tang et al [22] investigated the use of focused crawling techniques to assess the topic relevance and quality of medical information. For this purpose, n=160 seeds from the category depression of the Open Directory Project (now Curlie) were selected. They found that such an approach fetches twice as many pages as a crawler without topic focus. In another study, Pirkola et al [23] described the use of focused crawlers to acquire text from the genomics domain. They found, that "the source of seed URLs and the issues related to the multilinguality of the web" are major challenges in this area.

Abbasi et al [24] used a focused crawler to collect credible medical sentiments and opinions in the context of drug surveillance. In this context, their crawler was evaluated on "a set of 100 seed URLs pertaining to health and drug-related websites" and achieved a harvest rate of 10.06% (1,243,074/12,362,406). In 2016, Abbasi et al [25] demonstrated the use of a focused crawler to acquire credible online medical content in the context of postmarket drug surveillance. Their method was able to "collect over 80% of all relevant credible content in the first 20% of the crawl."

In Xu et al [26], a user-oriented adaptive focused crawler was implemented and applied in the cancer domain (ie, on breast and lung cancer). The authors found "that the new crawler can substantially accelerate the online user-generated content acquisition efforts for cancer researchers."

Amalia et al [27] presented a focused crawler for the acquisition of health-related articles written in Indonesian. In this study, different crawling strategies and their relative impacts on crawler performance were investigated. They found that crawling larger sites first improves the number of crawled articles.

In 2016, Rheinländer et al [28] studied the scalability of an information extraction framework using a focused crawling approach to collect and analyze "a 1 TB collection of web text from the biomedical domain" written in English. For this purpose, they generated a set of n=485,462 seeds using

XSL•FO
RenderX

commercial search engines with which their focused crawler achieved a harvest rate of 38%.

Aims of the Study

The authors decided to concentrate on health-related web pages available free of charge on the internet in the D-A-CH region that can be found under the respective country-code top-level domains (ccTLDs) .de, .at, and .ch. In this context, the aim of this study was fourfold:

- Demonstrate the suitability of a focused crawler approach for the acquisition of health-related content in the D-A-CH region
- Provide a curated list of seed points for the health domain in the D-A-CH region
- Provide a crowd-validated evaluation data set consisting of health-related and non–health-related URLs that can be used to evaluate other classifiers used in focused crawlers for the health domain in the D-A-CH region
- Give preliminary insights into the graph structure of the GHW

To the best of the authors' knowledge, no similar study has been previously conducted on a large scale. In particular, this has not been done for the GHW.

Besides a statistical analysis of the GHW, this paper shares our experience in building and operating a highly scalable focused crawler. Thus, researchers who want to perform a similar analysis for web pages of the health domain in their country can benefit from the experiences gained.

Methods

Focused Web Crawling

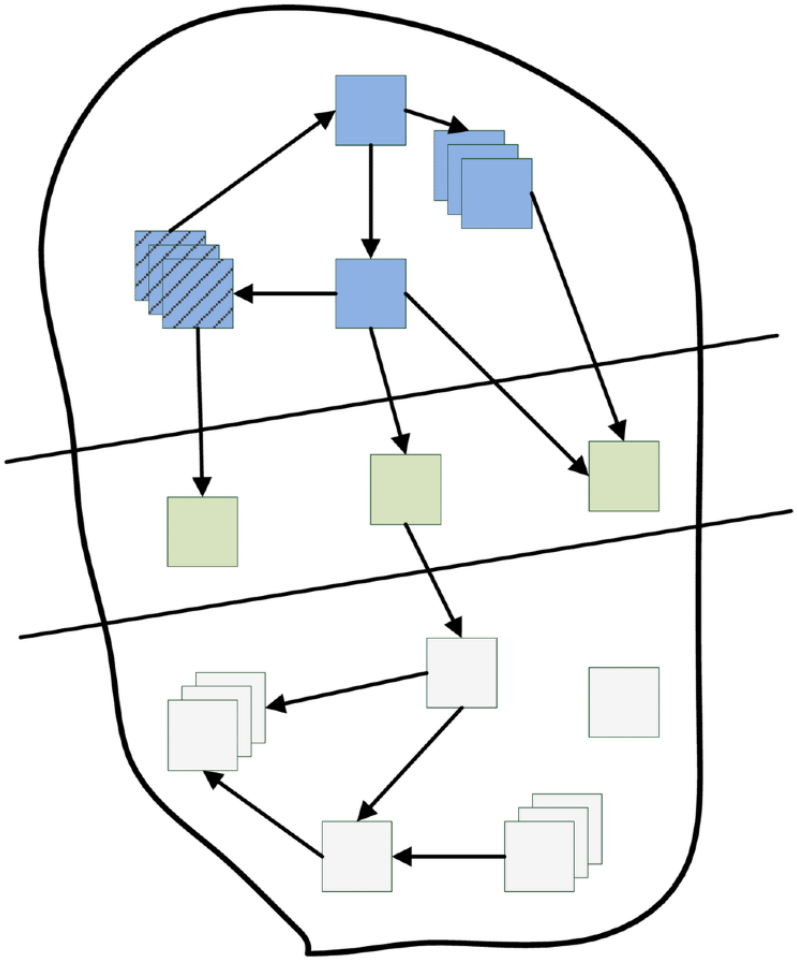
Basic Web Crawling Process

As depicted in Figure 1, a web crawler traverses the directed graph of the web [29,30]. Starting from a given set of seed URLs, the web crawler fetches web pages. After the download is successful, the HTML of a web page is parsed and hyperlinks to other web pages are extracted. These links are then analyzed and added in a priority queue called frontier [30,31]. The web graph is then visited via those URLs kept in the frontier. The crawler repeats this process until the frontier is empty or it is stopped manually.

Due to the enormous size of the web [14], one must focus on a certain domain of interest to speed up the crawl. In this context, a focused crawler only visits those outgoing links of a web site that appear to be relevant for the given topic. To determine whether a link is relevant or not, the assumption is made that web pages of a certain topic are most likely linked to other web pages of the same topic [32]. To assess the relevance of a certain web page, a focused crawler often uses techniques from the field of machine learning [16,31]. Classifiers are then leveraged to filter irrelevant content during the crawl process and assign priority on extracted URLs based on the classification result.

XSL•FO
RenderX

Figure 1. Schematic representation of the web graph traversal by a crawler. Pages colored in blue represent processed pages; in green, pages referenced in the frontier; in gray, undiscovered web content. Pages in dashed blue represent so-called initial seed pages.



System Architecture and Processing Workflow

Given the results of the Van der Bosch study [14], it is obvious that sequential processing of such an amount of data would take a tremendous amount of time and/or financial resources. For this reason, a parallel and distributed system architecture as described by Shkapenyuk [33] is necessary to crawl within a reasonable amount of time: to make results available before the web has notably changed. Therefore, such an architecture must be designed to handle thousands of worker threads to fetch web pages in parallel. Besides efficiency in terms of throughput, a crawler should also respect crawler ethics [29,34] (ie, it should honor the robot exclusion protocol [robots.txt]) [29,35,36]; this protocol allows web site administrators to inform the web crawler which parts of a web site should not be processed. In addition, a crawler should not overwhelm the target web server by sending too many request in a short period of time. For this reason, applying a politeness delay (time between requests to the same server) is mandatory. Furthermore, it must be robust protection against so-called spider traps, or web sites containing programmatic errors or dynamically generated links that cause the crawler to be trapped in an infinite loop [29]. Moreover, the HTML parser must tolerate broken and/or invalid markup [29,37,38]. In addition, text extraction components must handle boilerplate detection in an appropriate way [39,40].

There are several frameworks that realize such distributed crawlers; we built our system on top of the open-source framework StormCrawler [41], a software development kit for building low-latency, scalable crawlers based on the Apache Storm framework [42]. It lacks out-of-the-box components for focused crawling but offers the possibility of adding custom extensions and configuration options. For this reason, we extended it with classifiers and the necessary logic to implement a focused crawler. Figure 2 depicts the architecture of StormCrawler (black) with our focused crawler extension (orange).

The StormCrawler software development kit provides a conventional recursive crawler architecture (upper part of Figure 2); a seed injector is used to read URLs from a text file and adds them to the CrawlDB, which acts as the crawl frontier and content storage. Next, a set of spouts emit yet unseen URLs from the crawl database. To maintain politeness, these URLs are then assigned to cluster nodes (based on their resolved hostname) and directed to the fetchers. The latter will download the respective web pages and forward them to the parsers for link and content extraction; unseen URLs are added to the frontier. Next, the content is sent to the indexers, which store it inside the CrawlDB (in this case an Elasticsearch cluster [43]).

To add focus to StormCrawler, the framework was extended by adding additional bolts and filter components (lower part of

XSL•FO
RenderX

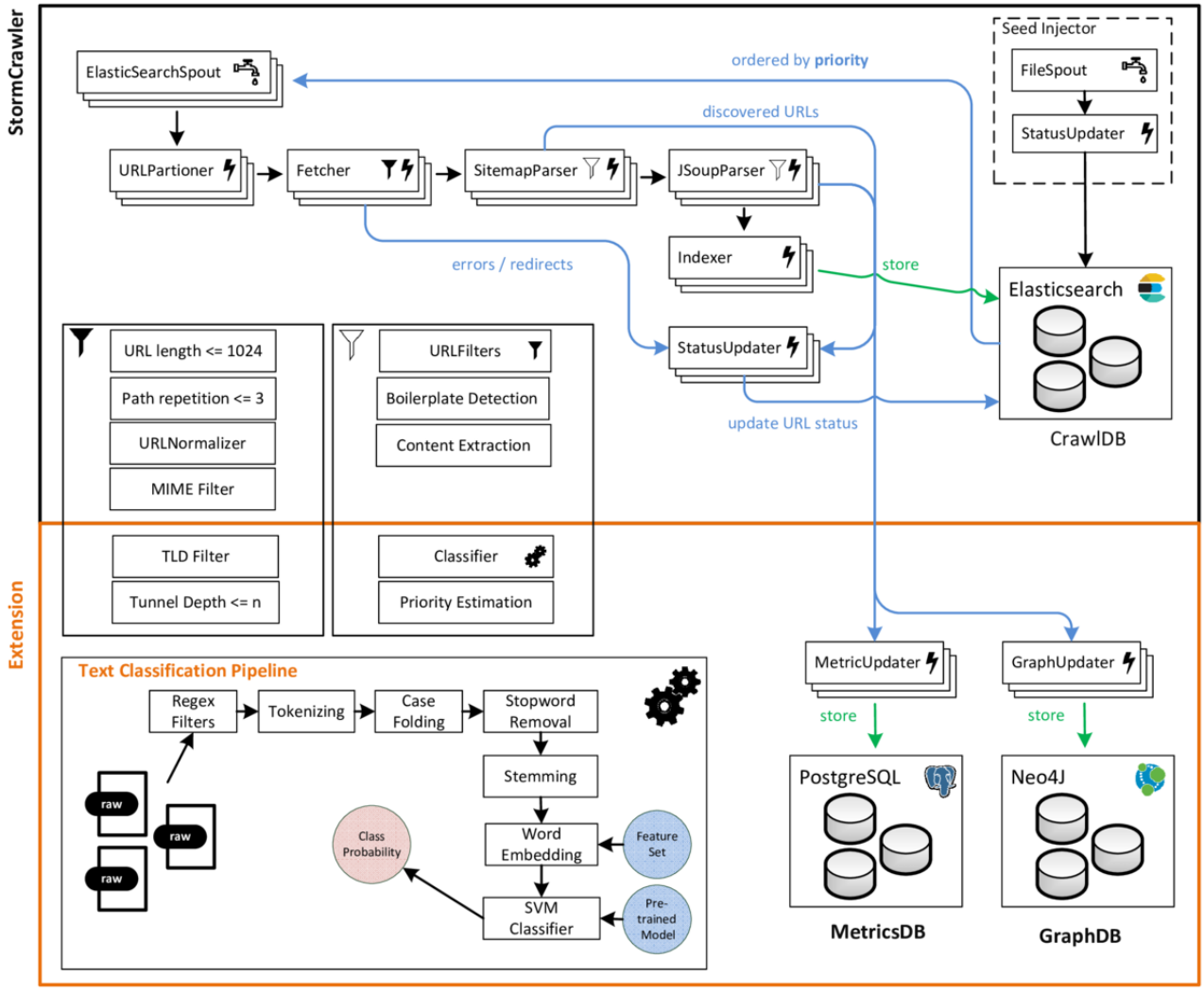
Figure 2). After a web page is parsed, the raw text is extracted by using boilerplate detection and XML path language expressions. It is then processed by a text classification pipeline to compute the relevance to the health domain as described by Joachims [44] and Zowalla et al [45]. If a web page is classified as relevant, it is marked for further processing.

Next, a priority value (in this case a value between 0 and 127) is assigned to every URL contained on the given web page [46]. This is done by using (1) the class probability of the current web page [29,32], (2) a check whether the extracted URLs target the same hostname (a web site covering a certain topic will most likely contain more web pages of that topic) [29,32], (3) the anchor text of that link [47], and (4) the link itself using an n-gram approach [48]. Higher priority values will guarantee earlier processing.

In addition, we implemented a soft focused crawling strategy using tunneling to avoid stopping at the first irrelevant page. For example, many front pages of portals may be classified as irrelevant but link to relevant health-related content [28,31]. To do so, a specific filter component tracks the depth and stops after given n steps (eg, n=2, n=3). Irrelevant web pages are not indexed.

To build the web graph of the health domain, during the crawl process a specific bolt keeps track of the visited and discovered links and adds them to a clustered Neo4J graph database. For statistics and metrics related to the crawl, another bolt continuously updates the crawling progress inside a PostgreSQL database. The crawling and classification process is repeated until the frontier is empty or it is stopped manually by the user.

Figure 2. Architecture of a focused crawler based on the StormCrawler software development kit. Spouts (tap symbol) emit data (here: URLs), bolts (lightning symbol) process data (ie fetch, parse, and store the extracted content). Bolts can be enhanced with URL filters (white filter symbol) or parse filters (black filter symbol). URL filters are used to remove URLs based on predefined criteria. Parse filters include URL filters but are primarily used to clean the parsed content and compute topic relevance and priority.



System Environment and Hardware Setup

In total, 22 virtual machines participate in the computing cluster providing the infrastructure for the crawler. The corresponding services are used to run, manage, and analyze the crawled web pages on the fly. For this setup, two physical servers of a Cisco unified computing system provide the computational resources and run as a virtualization platform to allow shared resource allocation.

Each server offers two physical central processing units (Intel Xeon E5-2689) with 8 cores each and 256 GB of memory. On

XSL•FO
RenderX

the network side, the Cisco unified computing system is attached to two optical 10 gigabit ethernet fibers that provide high bandwidth and ensure scalable throughput. A network attached storage system provides a total disc capacity of 60 TB to persist crawled data and store participating virtual machines via the network file system protocol. This network attached storage is also connected via optical fibers to our university's core router.

Evaluation Measures for Focused Web Crawling

Several studies state that the primary metric in evaluating focused crawler performance is the harvest rate [20,29,31,49,50]. Harvest rate is defined as "the fraction of webpages crawled that satisfy the relevance criteria among all crawled webpages" [20]. Previous studies reported that the harvest rate ranges between 10% and 45% for such systems [24,28,31,51].

In addition, the recall (also known as sensitivity) measure can be estimated by using the seed-target approach [29,52,53]. In this context, the initial set of seed pages is split into two sets of which one can be used as seeds and the other as targets (T). Figure 3 depicts the relationship between relevant (R), crawled (S) and target web pages.

According to Liu [29], the recall may be estimated if T is a representative, unbiased sample of R independent of the crawling process by the equation in Figure 4 at any time t.

Figure 3. Relationship between target, relevant, and crawled web pages. Recall is estimated based on known relevant target pages and underlying independence assumption.

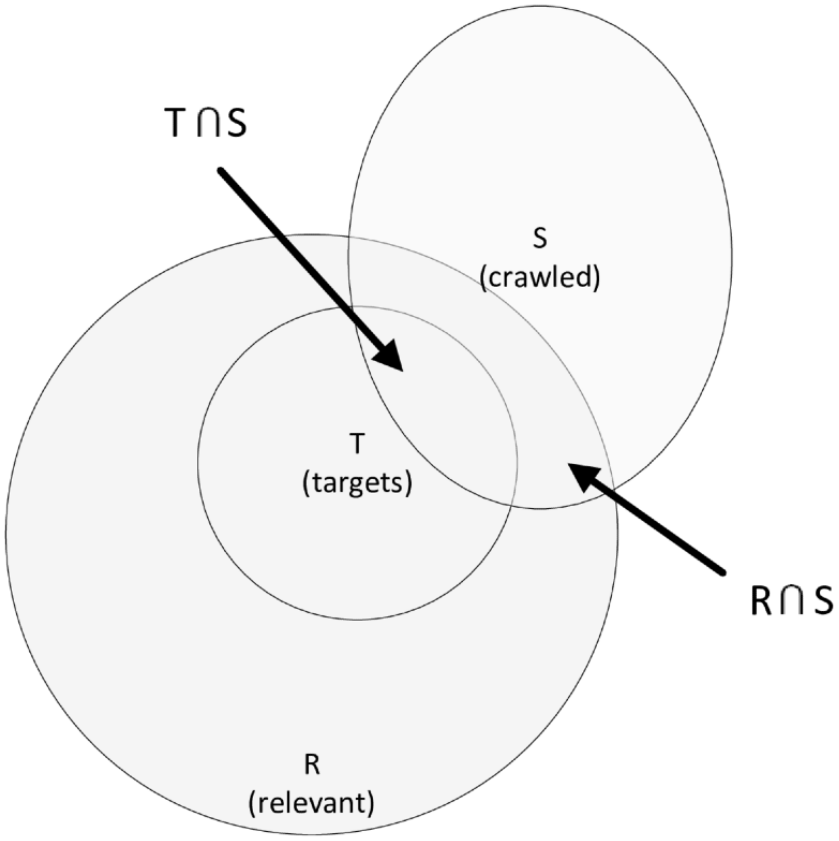


Figure 4. Recall estimate equation.

$$R_T(t) = \frac{|T \cap S_t|}{|T|}$$

Text Classification

Support vector machines (SVMs) originate from the field of machine learning and are known to perform well for text classification tasks [44,45]. For this reason, we relied on an SVM to determine a web page's relevance within the health domain. Figure 5 depicts the system's workflow for the training and classification phase.

To build our SVM-based text classifier, we followed related methods as described by Joachims [44]: as a first step toward a health text classifier, automatically gathered health-related articles (contained in a document collection [D]) were cleaned from syntactic markup (eg, boilerplate code, HTML tags). Each article was then tokenized (ie, split into single word fragments) and each character was converted to lower case (also known as case folding). Stop words (eg, the, and, it) were removed as these kinds of tokens do not carry any relevant information. Next, stemming techniques were applied in order to map tokens to their stem forms and reduce morphological variations of words (eg, goes becomes go). Each article was transformed into a document vector containing all distinct terms. To do so, it is

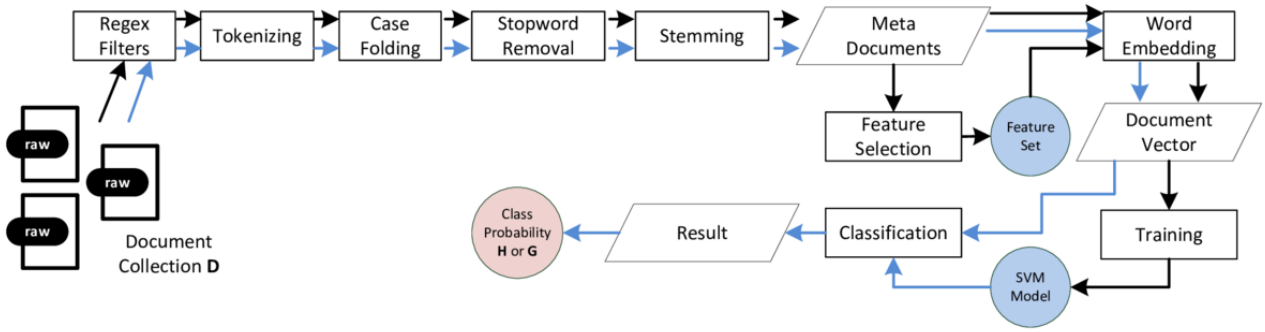
XSL•FO
RenderX

necessary to compute the terms that are representative for every article. A so-called feature selection produces a smaller subset of features (F) which yields the most relevant features for each article, limited by a predetermined threshold [54]. Given D and F, an SVM was trained to distinguish between vectors of health-related (H) and non–health-related (G) articles. The resulting classifier may be applied to previously unclassified web pages in order to predict their health-relatedness. To evaluate the classifier's quality, we used well-established metrics from the field of information retrieval such as accuracy, recall, and precision [29].

LIBSVM [55] and its object-oriented binding zlibsvm [56] were used as an SVM implementation of the text classifier. For building and training the SVM, the process described by Joachims [44] was applied. To reduce dimensionality, the feature selection method *information gain* was used [54]. Word embedding was conducted using *tfc* [57] as a term-weighting approach.

To find an optimal hyperparameter combination for the chosen radial basis function kernel, a grid search using 10-fold cross-validation, as recommended by the LIBSVM authors [55], was conducted. According to the Pareto Principle, training and test data were constructed using an 80:20 split [58]. In addition, the classes inside these data sets were equally balanced according to Wei and Dunbrack [59] as the real-world class distribution of H and G is unknown.

Figure 5. Workflow of an support vector machine–based text classification system: black lines indicate the training process; blue lines indicate the classification process; slanted boxes represent data; rectangular boxes represent computational steps.



Graph Metrics

The graph structure of the web has been extensively analyzed in several studies [15,17,60,61]. In this context, a graph node represents a web page and an edge represents a link between two web pages. We generated a so-called host-aggregated graph from the original web graph in order to reduce its computational complexity and explore its properties [61]. In this process, single web pages are combined and represented by their parent web site (including outgoing and ingoing links). On the resulting host-aggregated graph, we applied the following metrics:

- Average degree is the average number of edges connected to a node [62]. For a (directed) web graph, this is defined as the total number of edges divided by the total number of nodes. The average in-degree denotes the average number of ingoing edges to a node (ie, links to a web site). The average out-degree is defined as the average number of outgoing edges of a node (ie, links targeting other web sites).
- Modularity measures the strength of division of a graph into clusters or groups [62,63]. Graphs with a high modularity have dense connections between the web sites within certain clusters but sparse connection to other web sites, which are contained in different clusters.
- PageRank is a centrality-based metric that allows identification of web sites (nodes) of importance inside a graph [64]. The underlying assumption is that an important or prestigious web site will receive more links from other important web sites (ie, higher in-degree).

Other metrics such as network diameter and the average path length (ie, the average number of clicks which will lead from one web site to another) are frequently used for graph analysis [62,65].

Data Acquisition

Seed Generation

The selection of seed sources is crucial for the performance of a focused crawler [24,28,66-68]. For certain top-level domains (TLDs; eg, .com), the domain name system zone files are available to the public free of charge containing all registered domains for the given TLD. These zone files can then be used to extract seeds. However, due to data protection regulations, accessing and using the domain name system zone files for the ccTLDs .de, .at, and .ch was not possible.

Other studies leverage search engines with specific queries [28,66,69] to obtain high-quality seeds. However, most search engines restrict the amount of queries and limit the returned amount of results. Also, the results might be influenced by commercial interests and crafting high-quality search queries demands time and/or financial resources.

Another widely used seed source is the web taxonomy Curlie [22,29,31,66,70,71], which provides human-maintained precategorized web sites. Seeds can be harvested as dumps and are available free of charge. In addition, it is possible to reuse the results of a previous crawl to generate seeds. For this study, we relied on Curlie and the data of a previous health-related crawl conducted in 2016 [72].

XSL•FO
RenderX

Machine Learning Data Sets

Training and Test Corpus

To obtain a large enough data set for training and testing the SVM text classifier used within the focused crawler, web pages from various German content providers were obtained. First, the web pages were downloaded by specialized web crawlers implemented in Java using the crawler4j framework [73]. Next, boilerplate detection and data cleaning were conducted using regular expression filters. After this step, the cleaned textual content was stored in a relational database for further processing. Regarding each content provider, a random sample was manually inspected by the authors in order to assess data quality.

Each content provider and all related articles were put into one of the two classes: health-related language (H) or general language (G). The coding was based on (1) the organizations providing the content, (2) health-related content certification (eg, Health On the Net Foundation Code of Conduct), and (3) a manual inspection by the authors, in which the topic relevance of a random sample for each content provider was assessed.

Crowd-Validated Test Corpus

As the training and test corpus were generated by using a priori knowledge of each content provider, the authors decided to construct an additional independent human-validated data set to evaluate the classifier's performance.

Recent studies have shown that crowdsourcing can produce comparable results to human experts at a faster pace [74-78]. Thus, crowdsourcing was used to assess the evaluation data set. Figure 6 depicts the process of building this validated data set.

First, web pages were manually selected from a crawl conducted in 2016 [72]. It was ensured that the selected pages were neither included in the training set nor in the test corpus generated in the previous step. Next, each web page was assessed by a group of workers and categorized as H or G. Raters were given clear instructions on how to categorize given web pages (see Multimedia Appendix 1). In addition, each rater successfully completed a quiz-based training before they could participate in the study [79-83]. Precategorized web pages (test questions) were mixed into the rating process as test questions to keep the attention of the raters at a high level.

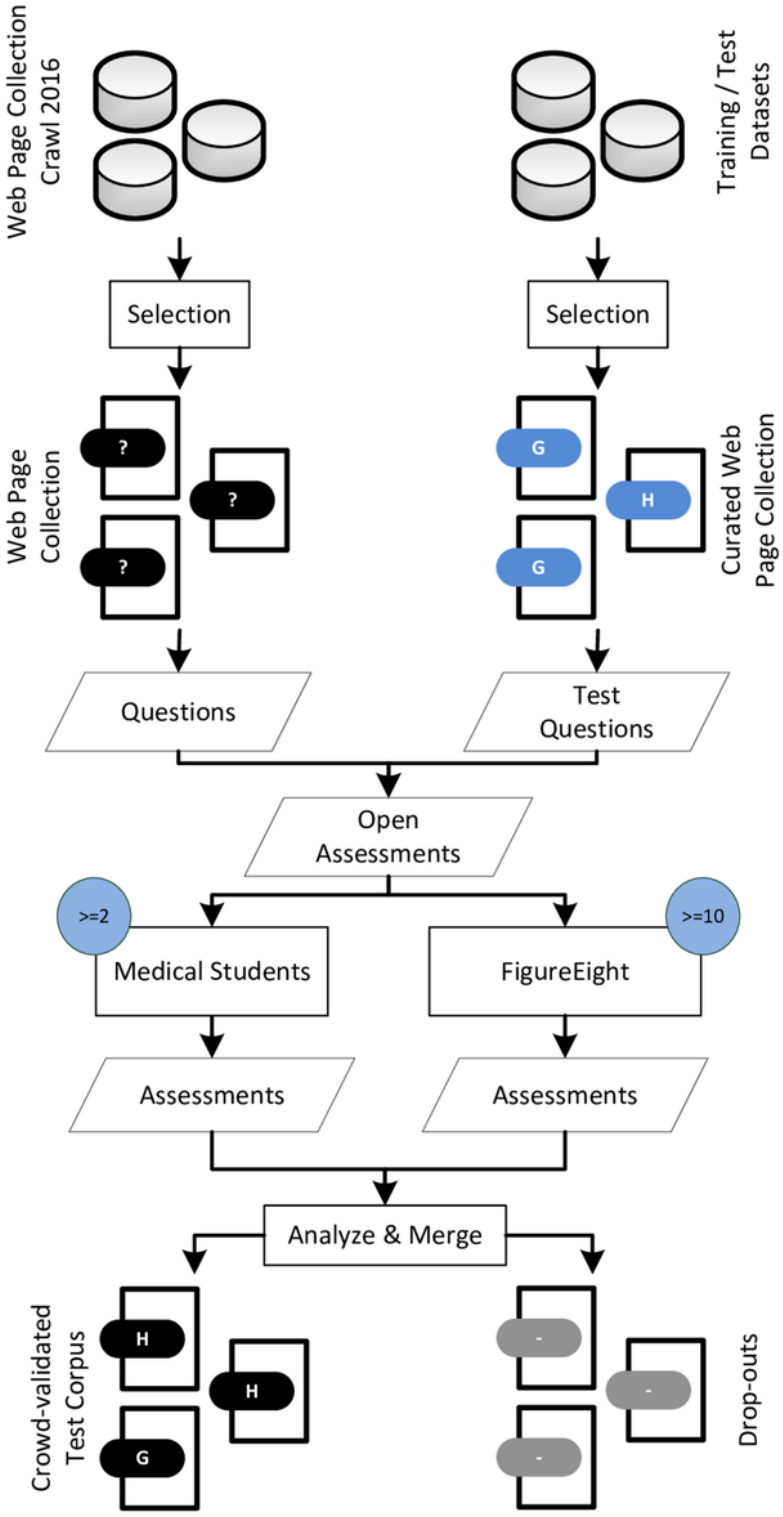
If a rater failed to answer a specific amount of such test questions, the assessments of this rater were considered as dropouts. Following the recommendation by Carvalho et al [84], each web page was assessed by at least 10 crowd-workers on the commercial crowd-working platform FigureEight [85]. In addition, the same web pages were coded by final year medical students (at least two students per web page) at the University of Heidelberg in the context of the lecture Medical Informatics. Study participation was voluntary.

If there was no clear majority vote for a certain class between the crowd-workers, the assessments of the medical students were taken into consideration. If there was still no agreement, the web page was listed as a dropout.

The statistical software R version 3.4.4 (R Foundation for Statistical Computing) on an Ubuntu 18.04 LTS 64-bit computer was used to compute percent agreement [86] and Fleiss κ [87].

XSL•FO
RenderX

Figure 6. Workflow of the crowd-sourcing approach to build a test corpus for the purpose of classifier evaluation. Black lines indicate the assessment process; slanted boxes represent data; rectangular boxes represent processing steps.



Results

Seeds

Seeds were obtained from the health category (German: Gesundheit) of Curlie and a health-related crawl conducted in 2016 targeting health-related web sites in German [72]. In total, n=22,405 seeds with ccTLDs .de (19,126/22,405, 85.36%), .at (1530/22,405, 6.83%), .ch (1749/22,405, 7.81%) were collected and used in this study. The full list of seeds can be found in Multimedia Appendix 2.

Machine Learning Data Sets

Data Set Characteristics

Web pages from various German content providers were collected between April 24, 2018, and August 16, 2018. A detailed list and description of each content provider is shown

XSL•FO
RenderX

in Table 1. In total, 98,442 articles were collected. The average word count for each document was 741; the average sentence count was 44.

For category H, we collected 9638 articles from the categories "medicine" and "medical report" from Deutsches Ärzteblatt (a magazine tailored to physicians) and 1907 from Apotheken Umschau (a magazine freely available in German pharmacies, tailored to lay citizens). In addition, we acquired 235 and 636 articles from the medical content providers Institute for Quality and Efficiency in Healthcare and Onmeda, respectively. Moreover, 2829 documents were obtained from the national health portal of the Republic of Austria. In addition, 28,436 health-related articles from Wikipedia Health were gathered by using the Wikipedia category graph. For category G, 18,364 random articles from Wikipedia General were collected, which were not related to the category Health (German: Gesundheit). In addition, 36,297 German web pages were selected randomly from the Common Crawl Foundation.

Table 1. Total number of acquired articles and respective class labels of various German content providers.

Content provider	Class	Cert[a]	Organization	Articles	Words (mean)	Words (median)	Sentences (mean)
Wikipedia Health	H[b]	no	Wikimedia Foundation	28,436	429	254	31
Wikipedia General	G[c]	no	Wikimedia Foundation	18,364	736	266	26
Common Crawl	G	no	Common Crawl Foundation	36,297	480	429	33
Deutsches Ärzteblatt	H	no	German Medical Association, National Association of Statutory Health Insurance Physicians	9638	1852	520	136
Onmeda	H	yes	Gofeminin.de GmbH	636	6564	6113	439
gesundheitsinformation.de	H	yes	Institute for Quality and Efficiency in Healthcare	235	1923	1799	139
Apotheken Umschau	H	yes	Wort & Bild Verlag	1907	1052	658	73
GESUNDheit.gv.at	H	no	Ministry of Social Affairs (Austria)	2929	295	221	21
Total	—[d]	—	—	98,442	741	339	44

[a]Yes indicates that a provider is certified by the Health On The Net Foundation Code of Conduct or another certification provider.

[b]H: health-related language.

[c]G: general language.

[d]Not applicable.

Training and Test Corpus

For training and evaluation of the SVM classifier, 87,562 articles were used. Table 2 lists the final data sets. In total, 80.00% (70,048/87,562) of articles were used for training the classifier and 20.00% (17,514 of 87,562) were used for testing.

XSL•FO
RenderX

Table 2. Total amount of articles used in the training and test corpus per content provider with corresponding class labels: health-related language (H) and general language (G).

Content provider	Class	Documents		
		Training	Test	Total
Wikipedia	H[a]	22,748	5688	28,436
Wikipedia	G[b]	10,339	2585	12,924
Common Crawl	G	24,685	6172	30,857
Deutsches Ärzteblatt	H	7710	1928	9638
Onmeda	H	509	127	636
gesundheitsinformation.de	H	189	46	235
Apotheken Umschau	H	1525	382	1907
GESUNDheit.gv.at	H	2343	586	2929
Total	_[c]	70,048	17,514	87,562

[a]H: health-related language.

[b]G: general language.

[c]Not applicable.

Crowd-Validated Test Corpus

A total of 432 web pages (216 per class) were manually selected from a health-related crawl conducted in 2016 [72]. The selected web pages were neither contained in the training nor in the test corpus (see Table 2).

Each web page was assessed by 10 crowd-workers between February 2, 2019, and February 16, 2019, on the commercial crowd-working platform, FigureEight [85]. In total, 4367 assessments by 28 crowd-workers were collected at a cost of US $36.06. The overall satisfaction (as measured by FigureEight) was 4.45 out of 5 possible points (instructions clear: 4.5/5; test questions fair: 4.55/5; ease of job: 4.5/5; payment: 3.65/5); 14 out of 28 (50%) workers participated in this voluntary exit survey. Percent agreement was 0.855; Fleiss κ was 0.279.

In addition, the same web pages were coded by medical students (n=40). Study participation was voluntary. Each web page was assessed by at least two students. Percent agreement was 0.719; Fleiss κ was 0.337. According to Landis and Koch [88], these κ values correspond to a fair agreement.

The resulting data set contained n=384 web pages (192 per class). This corresponds to a dropout rate of 11.1% (48/432). The full list of coded web pages is given in Multimedia Appendix 3.

Classifier Performance

The classifier was evaluated against the test and crowd-validated data set, and the results are presented in Table 3. The classifier achieved a precision of 0.934, a recall of 0.940, and an accuracy of 0.937 on its test data set; 5.96% (522/8757) of health-related web pages were falsely classified as nonrelevant by the SVM. On the other hand, 6.57% (575/8757) of the nonrelevant pages were classified as health-related.

On the crowd-validated real-world data set, the classifier achieved an accuracy of 0.966, a precision of 0.954, and a recall of 0.989. Only 1.0% (2/192) of the health-related web pages were falsely classified as nonrelevant, and 5.7% (11/192) of nonrelevant web pages were classified as health-related.

XSL•FO
RenderX

Table 3. Listing of the confusion matrix and related evaluation metrics for the test and crowd-validated data set.

Evaluation data sets	Baseline					
	Health	General	Sum	Accuracy	Precision	Recall
Test data set				0.937	0.934	0.94
SVM[a]	—[b]	—	—	—	—	—
Health	8182	575	8757	—	—	—
General	522	8235	8757	—	—	—
Sum	8704	8810	17,514	—	—	—
Crowd-validated data set				0.966	0.954	0.989
SVM	—	—	—	—	—	—
Health	181	11	192	—	—	—
General	2	190	192	—	—	—
Sum	183	211	384	—	—	—

[a]SVM: support vector machine.

[b]Not applicable.

Crawler Performance

Our system achieved a download rate of 7 to 10 documents per second. This sums up to 227 days of pure crawling and classification of approximately 133 million web pages.

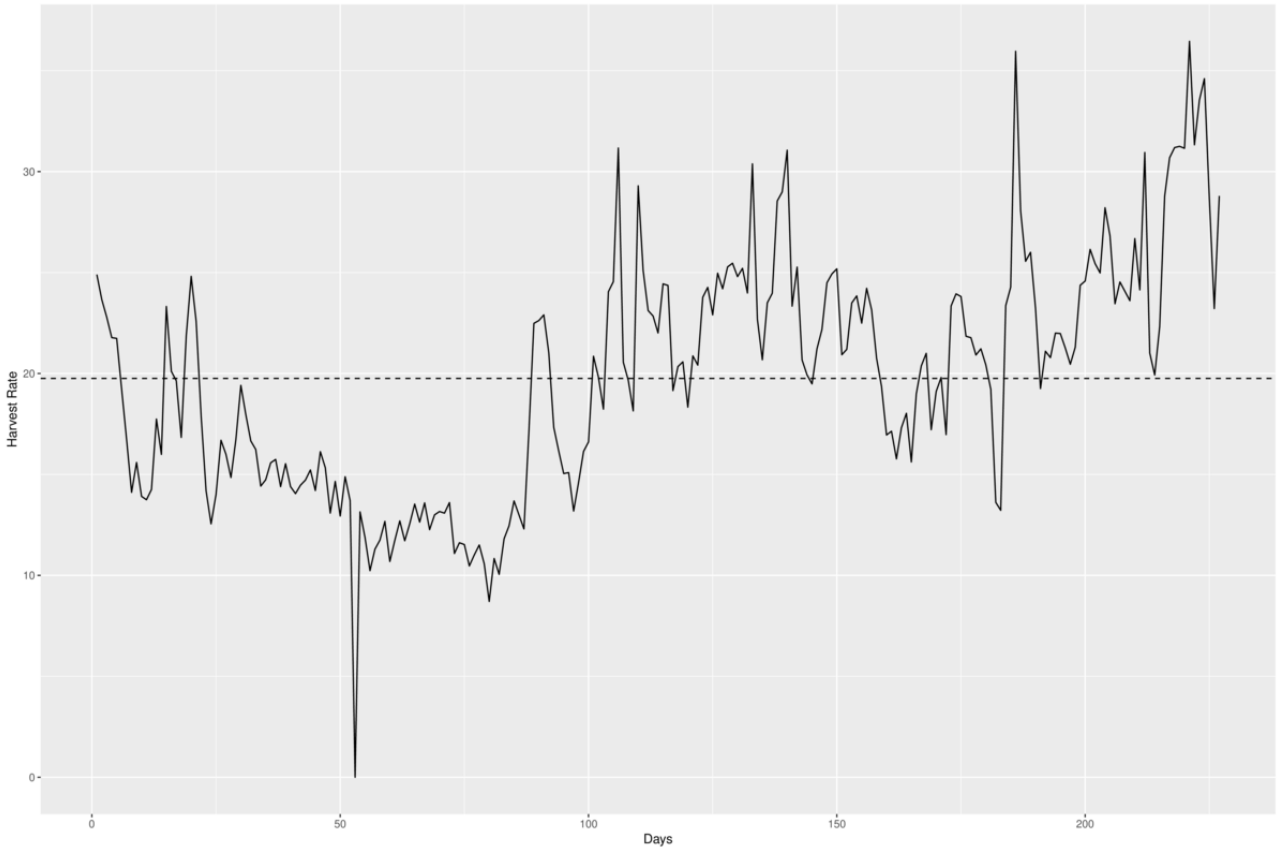
The crawl yielded approximately 13.5 million presumably relevant web pages and approximately 119.5 million nonrelevant web pages. Figure 7 depicts the harvest rate during the crawl.

The overall mean harvest rate was 19.76% ($HR_{t=222}=HR_{max}=36.45\%$; $HR_{t=53}=HR_{min}=0.00\%$). HR_{max} was achieved at day 222 as the crawl was resumed after infrastructure maintenance due to urgent security updates; HR_{min} was recorded on day 53. It was caused by a data center outage in which the infrastructure had to be shut down.

As an additional measure, we estimated the recall of our focused crawling by using the seed-target approach [29]. For this purpose, the initial seed set (n=22,405) was divided into a set of seeds (n=17,405) and targets (n=5000); ccTLD distribution was maintained in each sub set, and 4105 out of 5000 targets (82.10%) were contained in the crawl. This corresponds to an estimated recall of 0.821.

XSL•FO
RenderX

Figure 7. Harvest rate over time measured at the end of each day (dashed line represents the mean harvest rate). Note that the drop at day 53 is related to an outage at our data center. Peak at day 106: storm cluster was extended by three additional virtual machines. Peaks at days 157, 158, 191, 194 and 222: crawl was resumed after infrastructure maintenance due to urgent security updates that required a restart of the host system and/or of the virtual machines.



Graph Structure

The graph database Neo4J in v3.5.4 and its graph algorithm plugins [89] were used to compute the metrics as described in Graph Metrics on an Ubuntu 18.04 LTS 64-bit server. In order to reduce graph complexity, all web pages belonging to the same web site were aggregated and substituted by their parent web site (including outgoing and ingoing links; see Graph Metrics). The resulting graph contains 215,372 nodes (web sites) and 403,175 edges (links between web sites). A total of 82.56% (177,816/215,372) of the web sites belong to the ccTLD .de; 7.95% (17,126/215,372) to .at, and 9.49% (20,430/215,372) to .ch.

The graph has a network diameter of 25. The average path length is 6.466. The average degree is 1.872, the average in-degree is 1.892, and the average out-degree is 1.845. Modularity was computed to be 0.723.

During the analysis, several types of website publishers emerged: public institutions, nonprofit organizations, and private organizations or single individuals. As the ccTLD .de has the highest share within the graph, a global ranking according to PageRank would be dominated by .de web sites. For this reason, the following paragraph will present the top 25 web sites according to PageRank for each ccTLD separately.

Table 4 lists the 25 top-ranked web sites according to PageRank with their respective publisher for .de; 12 out of 25 (48%) are published by public institutions, 32% (8/25) are published by nonprofit organizations, and 20% (5/25) by private organizations. The top-ranked 25 web sites for .at are shown in Table 5; 12 out of 25 (48%) are published by public institutions, 4% (1/25) are published by nonprofit organizations, and 48% (12/25) by private organizations (see Table 5). For the ccTLD .ch, 24% (6/25) are published by public institutions, 40% (10/25) originate from nonprofit organizations, and 9/25 (36%) are published by private organizations (see Table 6).

Overall, 40% (30/75) are web sites published by public institutions, 25% (19/75) are published by nonprofit organizations, and 35% (26/75) by private organizations.

The graph visualization tool Gephi v0.9.2 [90] was used on a bare-metal Windows 10 64-bit computer to explore the host-aggregated graph structure. Unfortunately, we experienced serious performance issues while running Gephi's visualization algorithms. This is a main reason why we illustrate just a small example extract of the host-aggregated graph: Figure 8 consists of 94 nodes and 243 edges and presents basic aspects of the graph's structure. The focus is on www.rki.de as the top-ranked web site for the ccTLD .de (according to our analysis from below). The surrounding nodes represent health-related web sites in close proximity of www.rki.de.

XSL•FO
RenderX

Table 4. Domains of 25 top-ranked web sites for country-code top-level domain .de with their respective publisher according to PageRank.

Rank	Domain	Publisher	Type
1	www.rki.de	Robert Koch Institute	PI[a]
2	www.aerzteblatt.de	Deutscher Ärzte-Verlag GmbH	PI
3	www.charite.de	Charité–Berlin University of Medicine	PI
4	www.deutsche-alzheimer.de	Deutsche Alzheimer Gesellschaft	NPO[b]
5	www.aerztezeitung.de	Springer Medizin Verlag GmbH	PO[c]
6	www.dge.de	Deutsche Gesellschaft für Ernährung	NPO
7	www.g-ba.de	Gemeinsamer Bundesausschuss (Federal Joint Comitee)	PI
8	www.bzga.de	Bundeszentrale für gesundheitliche Aufklärung (Federal Centre for Health Education)	PI
9	www.bundesgesundheitsministerium.de	Bundesministerium für Gesundheit (Federal Ministry of Health)	PI
10	www.apotheken-umschau.de	Wort & Bild Verlag	PO
11	www.dimdi.de	Deutsches Institut für Medizinische Dokumentation und Information (German Institute for Medical Documentation and Information)	PI
12	www.gesundheitsinformation.de	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Healthcare)	PI
13	www.osteopathie.de	Verband der Osteopathen Deutschland eV	NPO
14	www.krebsgesellschaft.de	Deutsche Krebsgesellschaft eV	NPO
15	www.bfarm.de	Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices)	PI
16	www.kbv.de	Kassenärztliche Bundesvereinigung	PI
17	www.krebshilfe.de	Stiftung Deutsche Krebshilfe	NPO
18	www.tk.de	Techniker Krankenkasse (Health Insurance)	PO
19	www.ebm-netzwerk.de	Deutsches Netzwerk Evidenzbasierte Medizin eV	NPO
20	www.bmg.bund.de	Bundesministerium für Gesundheit (Federal Ministry of Health)	PI
21	www.netdoktor.de	NetDoktor.de GmbH	PO
22	www.drk.de	Deutsches Rotes Kreuz eV (German Red Cross)	NPO
23	www.herzstiftung.de	Deutsche Herzstiftung	NPO
24	www.klinikum.uni-heidelberg.de	Universitätsklinikum Heidelberg	PI
25	www.aok.de	AOK Gesundheiskasse (Health Insurance)	PO

[a]PI: public institution.

[b]NPO: nonprofit organization.

[c]PO: private organization.

XSL•FO
RenderX

Table 5. Domains of 25 top-ranked web sites for country-code top-level domain .at with their respective publisher according to PageRank.

Rank	Domain	Publisher	Type
1	www.gesundheit.gv.at	Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (Ministry of Social Affairs)	PI[a]
2	www.meduniwien.ac.at	University of Vienna	PI
3	www.bmgf.gv.at	Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (Ministry of Social Affairs)	PI
4	www.sozialministerium.at	Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (Ministry of Social Affairs)	PI
5	www.apotheker.or.at	Österreichische Apothekenkammer (Austrian Pharmaceutical Association)	PI
6	www.sam-pharma.at	Pharma Handel GmbH	PO[b]
7	www.aerztekammer.at	Österreichische Ärztekammer (Austrian Medical Association)	PI
8	www.univie.ac.at	University of Vienna	PI
9	www.herz-ambulatorium.at	Individual Person	PO
10	www.herz-ordination.at	Individual Person	PO
11	www.tg-steiermark.at	TG Therapeutische Gemeinschaft Betriebs GmbH	NPO[c]
12	www.impuls-fs.at	Institut für medizinisch-physiotherapeutische Untersuchung, Lehre und Schulung	PO
13	www.medunigraz.at	University of Graz	PI
14	www.brustvergroesserung-leicht.at	Individual Person	PO
15	www.bmg.gv.at	Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (Ministry of Social Affairs)	PI
16	www.kages.at	Steiermärkische Krankenanstaltengesellschaft mbH	PO
17	science.orf.at	Österreichischer Rundfunk (Austrian Broadcasting Corporation)	PI
18	www.gynmed.at	Individual Person	PO
19	www.fhstp.ac.at	St. Pölten University of Applied Sciences	PI
20	www.dr-boehm.at	Individual Person	PO
21	bmg.gv.at	Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (Ministry of Social Affairs)	PI
22	www.novartis.at	Novartis AG	PO
23	www.babyforum.at	FOKUS KIND Medien, CRAFT & VALUE	PO
24	femmestyle.at	Schönheitschirurgie femmestyle	PO
25	www.pfizer.at	Pfizer Inc	PO

[a]PI: public institution.

[b]PO: private organization.

[c]NPO: nonprofit organization.

XSL•FO
RenderX

Table 6. Domains of 25 top-ranked web sites for country-code top-level domain .ch with their respective publisher according to PageRank.

Rank	Domain	Publisher	Type
1	www.uzh.ch	University of Zurich	PI[a]
2	www.usz.ch	Universitätsspital Zürich	PI
3	www.srf.ch	Schweizerische Radio- und Fernsehgesellschaft (Swiss Broadcasting Corporation)	PI
4	www.netdoktor.ch	netdoktor GmbH	PO[b]
5	www.pancreas-help.ch	Schweizer Selbsthilfeorganisation Pankreaserkrankungen	NPO[c]
6	www.mutterglueck.ch	Individual Person	PO
7	www.association-osteo-swiss.ch	Schweizerischer Verband der Osteopathen	NPO
8	www.unibas.ch	University of Basel	PI
9	www.ethz.ch	ETH Zurich (Swiss Federal Institute of Technology in Zurich)	PI
10	www.rheumaliga.ch	Rheumaliga Schweiz	NPO
11	www.lungenliga.ch	Lungenliga Schweiz	NPO
12	www.rotpunkt-apotheken.ch	Rotpunkt-Pharma AG	PO
13	www.pharmawiki.ch	PharmaWiki GmbH	PO
14	www.bayer.ch	Bayer AG	PO
15	www.patientensicherheit.ch	Stiftung Patientensicherheit Schweiz	NPO
16	saez.ch	EMH Schweizerischer Ärzteverlag AG	NPO
17	www.swissheart.ch	Schweizerische Herzstiftung	NPO
18	gesundheitsfoerderung.ch	Gesundheitsförderung Schweiz	NPO
19	sensomotorische-lebensweisen.ch	Individual Person	PO
20	www.spitaluster.ch	Spital User	PO
21	symptome.ch	NOXA GmbH	PO
22	www.meineimpfungen.ch	Stiftung meineimpfungen	NPO
23	unicef.ch	United Nations International Children's Emergency Fund	NPO
24	www.bauchtumor.ch	Universitätsspital Bern	PI
25	www.fettabsaugungen.ch	FSnD Ltd	PO

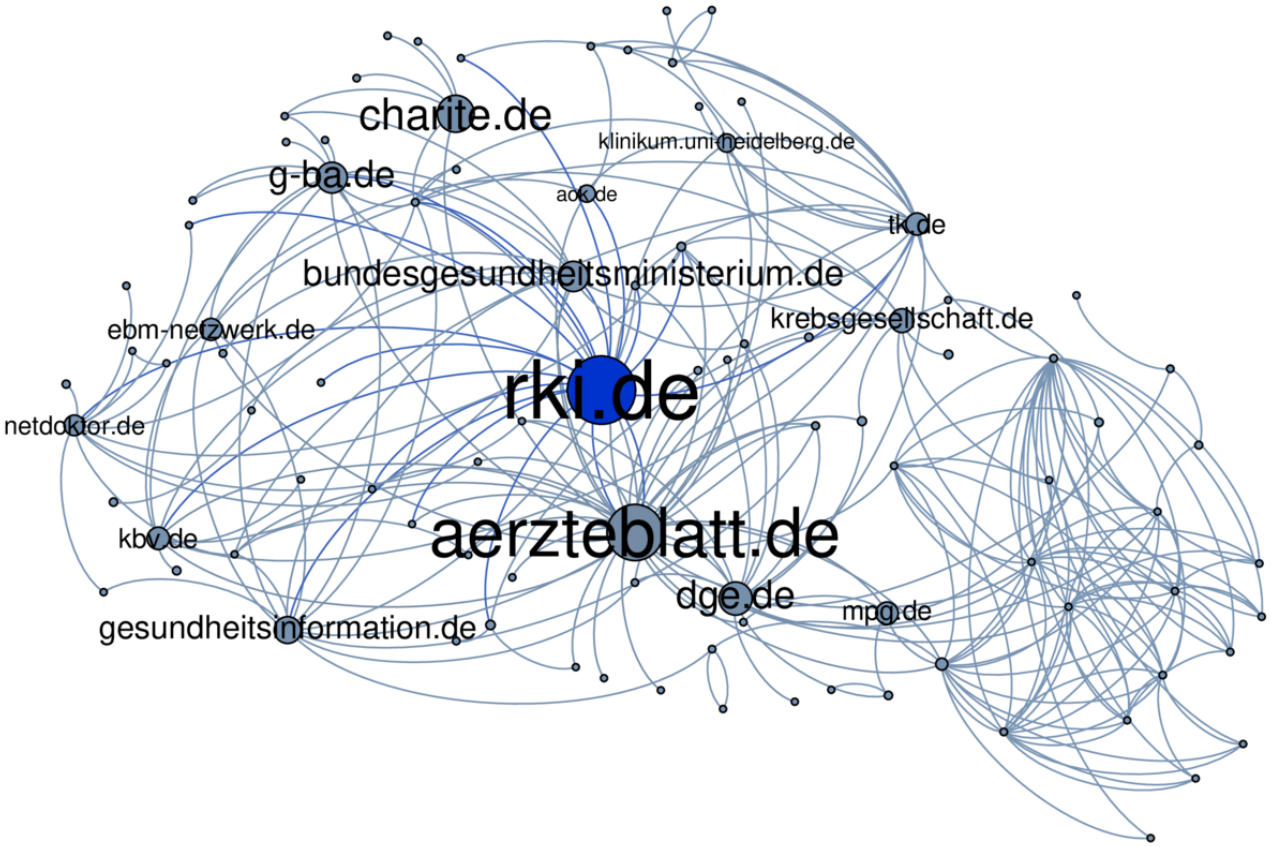
[a]PI: public institution.

[b]PO: private organization.

[c]NPO: nonprofit organization.

XSL•FO
RenderX

Figure 8. A small extract of the host-aggregated web graph with focus on the website www.rki.de. The surrounding nodes represent websites with a maximum link-distance of two starting from www.rki.de. An edge between two nodes implies that there exists at least one hyperlink between some web pages of the hosting websites in either way. Only those websites are included whose content is highly health-related (ie, which were automatically classified as belonging to H with a probability equal to or greater than 0.93). Moreover, they have at least one ingoing and one outgoing link. The bigger a node and its caption, the higher is its page rank. For illustration reasons, directional arrows were not included.



Discussion

Principal Findings

One aim of the study was to demonstrate the suitability of a focused crawler approach for the acquisition of health-related content. Our system achieved an average harvest rate of 19.76% during the entire crawl. In addition, the results show that the majority of the target seeds (4105/5000) could be obtained, which corresponds to a recall of 0.821. Therefore, we are confident that the proposed method is suitable to acquire most health-related content on the web and generate a suitable domain-specific graph representation.

A first manual investigation of several hundred randomly selected pages suggests that our approach produces accurate results. The results indicate that the web sites and web pages of major German, Austrian, and Swiss health-related public institutions have indeed been discovered, even though they were not contained in the initial seeds.

With respect to the study aims 2 and 3, we were able to provide a curated list of 22,405 seed points for the health domain in the D-A-CH region extracted from Curlie (see Multimedia Appendix 2). In addition, a data set with 396 items was created and evaluated by crowd-workers that can be used by other researchers to evaluate similar text classifiers (see Multimedia Appendix 3).

A first analysis of the graph structure (see study aim 4) shows that public institutions and nonprofit organizations have a higher importance according to their PageRank than web sites of private players inside the GHW.

Limitations

Several limitations apply for this study. First, we are not sure whether the seed pages cover a broad spectrum of topics within the health domain as we only acquired seeds from Curlie and a previous health-related crawl [72]. Using specifically crafted queries against established search engines would have increased the amount of available seeds and could have influenced the crawl in a positive way [28,66,69]. However, due to limited amounts of resources and time, we did not follow this approach as the web taxonomy Curlie and a previous crawl gave faster access to seed URLs. As Curlie is a community-driven web taxonomy, the publication process of new URLs is not strictly regulated. This might be a reason for the high share of private players within the top ranks of the web graph as everybody is eligible to publish a web site's URL on Curlie. In addition, the community behind Curlie is rather small compared with its predecessors (ie, URLs pointing to rather new content providers might not be contained in it). Therefore, corresponding web pages and their out-links might have been missed during the crawling process. This implies that reported graph properties might have been influenced by the chosen seed sources.

XSL•FO
RenderX

Second, with a mean accuracy of 0.951, our classifier might have produced false positive results during the crawl process. Third, we only considered the ccTLDs .de, .at, and .ch to avoid the need for a language classification system, as most web sites on these ccTLDs are written in German. Therefore, the data crawled covers only a certain fraction of the GHW, for example, as web sites in German published under .org are not contained.

Comparison With Prior Work

Previous studies investigated the use of focused crawler techniques to harvest biomedical or health-related text material [27,28]. In both analyses, the authors report that the use of focused crawlers requires a lot of computational effort to collect the data and analyze it in an appropriate way, which we can confirm by our observations.

Compared with the study by Rheinländer et al [28] in which they report an harvest rate of 38%, our system achieved an harvest rate of only 19.76%. This might be caused by (1) our system using a soft-focused crawling strategy meant it did not stop at the first encountered irrelevant web page, leading to an increase in irrelevant web pages and crawling time and (2) our crawl was limited to the ccTLDs .de, .at, and .ch as we did not implement a language classifier. This might have influenced the harvest rate of our system as well, yet it achieved a harvest rate in the typical range for such systems [24,28,31,51] (see Related Work).

In contrast to the studies by Rheinländer et al [28] and Amalia et al [27], we focused on the German language and the GHW. This study contributes to the field by demonstrating the suitability of a focused crawler approach for the acquisition of German health-related content in the D-A-CH region. A secondary study outcome is a curated list of seed points for the health domain in the D-A-CH region (see Multimedia Appendix 2). In addition, the crowd-validated evaluation data set (see Multimedia Appendix 3) can be used to evaluate other text classifiers for the given purpose. Moreover, this study gives first insights regarding the graph structure of the health-related web in the D-A-CH region.

Conclusions and Further Research

In this study, a system was presented which uses a focused crawling approach to gather the structure of the GHW. The system used an SVM-based classifier that was trained to assess the relevance of a web page for the health domain. The results indicate that the presented focused crawler is a suitable method for acquiring large health-related textual datasets and can be used to generate domain-specific graph representations. In future work, the authors intend to expand their web crawl by leveraging seed lists generated via search engine providers.

We also plan to analyze the linguistic characteristics of the crawled data as well as identify important topics and trends within this data. This will also include the identification of credible content providers and a comparison of the health-related web between Germany, Austria, and Switzerland. Moreover, future work will include a deeper exploration and analysis as well as a visualization of the resulting graph structure. Using these insights and with the acquired data available, an implementation and evaluation of a health-specific search engine for information seeking citizens will be possible.

Acknowledgments

The authors would like to thank Dr Monika Pobiruchin and Martin Wiesner of Heilbronn University for their feedback and valuable input to the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Instruction for raters written in German.
[DOCX File , 17 KB - jmir_v22i7e17853_app1.docx]

Multimedia Appendix 2
List of seed points.
[XLSX File (Microsoft Excel File), 484 KB - jmir_v22i7e17853_app2.xlsx]

Multimedia Appendix 3
Crowd-annotated corpus of health-related and non–health-related web pages.
[XLSX File (Microsoft Excel File), 26 KB - jmir_v22i7e17853_app3.xlsx]

References

1. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
2. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]

XSL•FO
RenderX

3. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
4. Tu HT, Cohen GR. Striking jump in consumers seeking health care information. Track Rep 2008 Aug(20):1-8. [Medline: 18770913]
5. Fox S, Duggan M. Health Online 2013. Washington: Pew Internet and American Life Project; 2013 Jan 15. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-06-08]
6. Prestin A, Vieux SN, Chou WS. Is online health activity alive and well or flatlining? Findings from 10 years of the Health Information National Trends Survey. J Health Commun 2015 Jul;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
7. Jacobs W, Amuta A, Jeon K. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):442-448. [doi: 10.1080/23311886.2017.1302785]
8. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
9. Wong DK, Cheung M. Online health information seeking and ehealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
10. Wetter T. Consumer Health Informatics New Services, Roles, and Responsibilities. Cham: Springer International Publishing; 2016:978-983.
11. Working Group GPGI. Good practice guidelines for health information. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 2016;110-111:e1-e8. [doi: 10.1016/j.zefq.2016.01.004]
12. Feldwisch-Drentrup H, Kuhrt N. Schlechte und gefährliche Gesundheitsinformationen: Wie sie erkannt und Patienten besser geschützt werden können Bad and harmful health information: How to identify such information to better protect patients. 2019. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_Analyse_Gefaehrliche_Gesundheitsinfos_final.pdf [accessed 2020-06-08]
13. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor—let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
14. van den Bosch A, Bogers T, de Kunder M. Estimating search engine index size variability: a 9-year longitudinal study. Scientometrics 2016;107:839-856 [FREE Full text] [doi: 10.1007/s11192-016-1863-z] [Medline: 27122648]
15. Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, et al. Graph structure in the Web. Computer Networks 2000 Jun;33(1-6):309-320. [doi: 10.1016/s1389-1286(00)00083-9]
16. Meusel R, Mika P, Blanco R. Focused crawling for structured data. Proc 23rd ACM Int Conf Inf Knowl Manag 2014:1039-1048. [doi: 10.1145/2661829.2661902]
17. Meusel R. The Graph Structure in the Web – Analyzed on Different Aggregation Levels. J Web Sci 2015 Jun 24;1(1):33-47. [doi: 10.1561/106.00000003]
18. Liu G, Yu Y, Han J, Xue G. China Web Graph Measurements and Evolution. In: Web Technologies Research and Development. Berlin: Springer; 2005:668-679.
19. Baggio R. The web graph of a tourism system. Physica A Stat Mech Appl 2007 Jun;379(2):727-734. [doi: 10.1016/j.physa.2007.01.008]
20. Kumar M, Bhatia R, Rattan D. A survey of Web crawlers for information retrieval. WIREs Data Mining Knowl Discov 2017 Aug 07;7(6):e1218. [doi: 10.1002/widm.1218]
21. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2016 May 21:146-155. [doi: 10.1111/hir.12143] [Medline: 27207817]
22. Tang T, Hawking D, Craswell N, Griffiths K. Focused crawling for both topical relevance and quality of medical information. Proc 14th ACM Int Conf Inf Knowl Manag 2005:147-154. [doi: 10.1145/1099554.1099583]
23. Pirkola A. Focused crawling: a means to acquire biological data from the web. 2007. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.6495&rep=rep1&type=pdf [accessed 2020-06-08]
24. Abbasi A, Fu T, Zeng D, Adjeroh D. Crawling credible online medical sentiments for social intelligence. 2013 Presented at: International Conference on Social Computing; 2013; Alexandria p. 254-263. [doi: 10.1109/socialcom.2013.43]
25. Abbasi A, Zhao K, Abraham B. A Prototype System for Collecting and Analyzing Credible Online Medical Content. In: Parsons J, editor. Tackling Society's Grand Challenge With Design Science. DESRIST 2016 Lecture Notes in Computer Science. Cham: Springer; 2016:197-201.
26. Xu S, Yoon H, Tourassi G. A user-oriented web crawler for selectively acquiring online content in e-health research. Bioinformatics 2014 Jan 01;30(1):104-114 [FREE Full text] [doi: 10.1093/bioinformatics/btt571] [Medline: 24078710]
27. Amalia A, Gunawan D, Najwan A, Meirina F. Focused crawler for the acquisition of health articles. 2016 Presented at: International Conference on Data and Software Engineering; 2016; Denpasar p. 1-6. [doi: 10.1109/icodse.2016.7936110]
28. Rheinländer A, Lehmann M, Kunkel A, Meier J, Leser U. Potential and Pitfalls of Domain-Specific Information Extraction at Web Scale. Proc 2016 Int Conf Manag Data 2016:759-771. [doi: 10.1145/2882903.2903736]

XSL•FO
RenderX

29. Liu B. Web Data Mining: Exploring Hyperlinks Contents and Usage Data, 2nd Edition. Berlin: Springer; 2011.
30. Brin S, Page L. Reprint of: The anatomy of a large-scale hypertextual web search engine. Comput Netw 2012 Dec;56(18):3825-3833. [doi: 10.1016/j.comnet.2012.10.007]
31. Chakrabarti S, van den Berg M, Dom B. Focused crawling: a new approach to topic-specific Web resource discovery. Comput Netw 1999 May;31(11-16):1623-1640. [doi: 10.1016/s1389-1286(99)00052-3]
32. Davison B. Topical locality in the Web. 2000 Presented at: Proc 23rd Ann Int ACM SIGIR Conf Res Devel Inf Retrieval; 2000; Athens p. 272-279. [doi: 10.1145/345508.345597]
33. Shkapenyuk V, Suel T. Design and implementation of a high-performance distributed Web crawler. 2002 Presented at: 18th Int Conf Data Eng; 2002; San Jose p. 357-368. [doi: 10.1109/icde.2002.994750]
34. Eichmann D. Ethical web agents. Comput Netw and ISDN Syst 1995 Dec;28(1-2):127-136. [doi: 10.1016/0169-7552(95)00107-3]
35. A method for web robots control. Network Working Group. 1996. URL: https://www.robotstxt.org/norobots-rfc.txt [accessed 2019-09-10]
36. Robots.txt-Specifications. Google Developers. URL: https://developers.google.com/search/reference/robots_txt?hl=en [accessed 2019-09-10]
37. Abiteboul S. Querying semi-structured data. In: Database Theory. Berlin: Springer; 1997:1-18.
38. Ofuonye E, Beatty P, Dick S, Miller J. Prevalence and classification of web page defects. Online Inf Rev 2010 Feb 23;34(1):160-174. [doi: 10.1108/14684521011024182]
39. Yi L, Liu B, Li X. Eliminating noisy information in web pages for data mining. Proc Ninth ACM SIGKDD Int Conf Knowl Discovery Data Mining 2003:296-305. [doi: 10.1145/956750.956785]
40. Kohlschütter C, Fankhauser P, Nejdl W. Boilerplate detection using shallow text features. Proc Third ACM Int Conf Web Search Data Mining 2010:441-450. [doi: 10.1145/1718487.1718542]
41. Nioche J. StormCrawler: a collection of resources for building low-latency, scalable web crawlers on Apache Storm. DigitalPebble Ltd. 2019. URL: http://stormcrawler.net/ [accessed 2019-09-10]
42. Allen S, Jankowski M, Pathirana P. Storm Applied: Strategies for Real-Time Event Processing. Shelter Island: Manning Publications Co; 2015.
43. Gormley C, Tong Z. Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine. Sebastopol: O'Reilly Media; 2015.
44. Joachims T. Text categorization with support vector machines: learning with many relevant features. Machine Learning: ECML 1998. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1398 1998:137-142. [doi: 10.1007/bfb0026683]
45. Zowalla R, Wiesner M, Pfeifer D. Automatically Assessing the Expert Degree of Online Health Content Using SVMs. Stud Health Technol Inform 2014;202:48-51. [Medline: 25000012]
46. Cho J, Garcia-Molina H, Page L. Efficient crawling through URL ordering. Comput Netw ISDN Syst 1998 Apr;30(1-7):161-172. [doi: 10.1016/s0169-7552(98)00108-1]
47. Pinkerton B. WebCrawler: Finding What People Want [Dissertation]. University of Washington. 2000. URL: http://www.thinkpink.com/bp/Thesis/Thesis.pdf [accessed 2020-06-03]
48. Rajalakshmi R, Aravindan C. Web page classification using n-gram based URL features. 2013 Fifth International Conference on Advanced Computing (ICoAC) Chennai, India: IEEE; 2013 Presented at: Fifth Int Conf Adv Comput; 2013; Chennia p. 15-21. [doi: 10.1109/icoac.2013.6921920]
49. Samarawickrama S, Jayaratne L. Automatic text classification and focused crawling. 2011 Sixth International Conference on Digital Information Management; Melbourne, Australia: IEEE; 2011 Presented at: Sixth Int Conf Dig Inf Manag; 2011; Melbourne p. 143-148. [doi: 10.1109/icdim.2011.6093329]
50. Bedi P, Thukral A, Banati H, Behl A, Mendiratta V. A multi-threaded semantic focused crawler. J Comput Sci Technol 2012 Nov 15;27(6):1233-1242. [doi: 10.1007/s11390-012-1299-8]
51. Pant G, Srinivasan P. Learning to crawl: comparing classification schemes. ACM Trans Inf Syst 2005 Oct 01;23(4):430-462. [doi: 10.1145/1095872.1095875]
52. Menczer F. ARACHNID: adaptive retrieval agents choosing heuristic neighborhoods for information discovery. Proc 14th Int Conf Mach Learn 1997:227-235.
53. Srinivasan P, Menczer F, Pant G. A general evaluation framework for topical crawlers. Inf Retrieval 2005 Jan;8(3):417-447. [doi: 10.1007/s10791-005-6993-5]
54. Yang Y, Pedersen J. A comparative study on feature selection in text categorization. Int Conf Mach Learn 1997:412-420. [doi: 10.5555/645526.657137]
55. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. URL: https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf [accessed 2020-06-08]
56. Zowalla R, Wiesner M. zlibsvm: an object-oriented Java-binding for support vector machines in the medical domain. Düsseldorf: German Medical Science GMS Publishing House; 2017.
57. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management 1988 Jan;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]

XSL•FO
RenderX

58. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics 2011 Apr 08;4:31 [FREE Full text] [doi: 10.1186/1755-8794-4-31] [Medline: 21477282]
59. Wei Q, Dunbrack RL. The role of balanced training and testing data sets for binary classifiers in bioinformatics. PLoS One 2013;8(7):e67863 [FREE Full text] [doi: 10.1371/journal.pone.0067863] [Medline: 23874456]
60. Meusel R, Vigna S, Lehmberg O, Bizer C. Graph structure in the Web—revisited: a trick of the heavy tail. WWW '14 Companion: Proc 23rd Int Conf World Wide Web 2014:427-432. [doi: 10.1145/2567948.2576928]
61. Lehmberg O, Meusel R, Bizer C. Graph structure in the Web: aggregated by pay-level domain. In: Proc 2014 ACM Conf Web Sci. 2014 Presented at: ACM conference on Web science; 2014; Bloomington p. 119-128. [doi: 10.1145/2615569.2615674]
62. Gross J, Yellen J. Handbook of Graph Theory. Boca Raton: CRC Press; 2004.
63. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
64. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web. 1999. URL: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf [accessed 2020-06-08]
65. Albert R, Barabási A. Statistical mechanics of complex networks. Rev Mod Phys 2002 Jan 30;74(1):47-97. [doi: 10.1103/revmodphys.74.47]
66. Vieira K, Barbosa L, da Silva AS, Freire J, Moura E. Finding seeds to bootstrap focused crawlers. World Wide Web 2015 Feb 26;19(3):449-474. [doi: 10.1007/s11280-015-0331-7]
67. Wu J, Teregowda P, Ramírez J, Mitra P, Zheng S, Giles C. The evolution of a crawling strategy for an academic document search engine: whitelists and blacklists. Proc 4th Ann ACM Web Sci Conf 2012:340-343. [doi: 10.1145/2380718.2380762]
68. Zheng S, Dmitriev P, Giles C. Graph based crawler seed selection. Proc 18th Int Conf World Wide Web 2009:1089-1090. [doi: 10.1145/1526709.1526870]
69. Prasath R, Oztürk P. Finding potential seeds through rank aggregation of web searches. Proc 4th Int Conf Pattern Recog Mach Intell 2011:227-234. [doi: 10.1007/978-3-642-21786-9_38]
70. Curlie. Curlie Project Inc. 2019. URL: https://curlie.org/ [accessed 2019-09-10]
71. Chakrabarti S, Punera K, Subramanyam M. Accelerated focused crawling through online relevance feedback. Proc 11th Int Conf World Wide Web 2002:148-159. [doi: 10.1145/511446.511466]
72. Zowalla R, Wiesner M, Pfeifer D. Analyzing the German Health Web using a focused crawling approach. 2016 Presented at: Health—Exploring Complexity: An Interdisciplinary Systems Approach; 2016; Munich.
73. Ganjisaffar Y. crawler4j: an open source crawler for Java. 2016. URL: https://github.com/yasserg/crawler4j [accessed 2019-09-10]
74. Maier-Hein L, Mersmann S, Kondermann D, Bodenstedt S, Sanchez A, Stock C, et al. Can masses of non-experts train highly accurate image classifiers? Med Image Comput Comput Assist Interv 2014;17(Pt 2):438-445. [doi: 10.1007/978-3-319-10470-6_55] [Medline: 25485409]
75. Maier-Hein L, Kondermann D, Roß T, Mersmann S, Heim E, Bodenstedt S, et al. Crowdtruth validation: a new paradigm for validating algorithms that rely on image correspondences. Int J Comput Assist Radiol Surg 2015 Aug;10(8):1201-1212. [doi: 10.1007/s11548-015-1168-3] [Medline: 25895078]
76. Irshad H, Oh E, Schmolze D, Quintana LM, Collins L, Tamimi RM, et al. Crowdsourcing scoring of immunohistochemistry images: evaluating performance of the crowd and an automated computational method. Sci Rep 2017 Feb 23;7:43286 [FREE Full text] [doi: 10.1038/srep43286] [Medline: 28230179]
77. O'Neil A, Murchison J, van Beek E, Goatman K. Crowdsourcing Labels for Pathological Patterns in CT Lung Scans: Can Non-experts Contribute Expert-Quality Ground Truth? Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Berlin: Springer International Publishing; 2017:96-105.
78. Park J, Mirhosseini S, Nadeem S, Marino J, Kaufman A, Baker K, et al. Crowdsourcing for identification of polyp-free segments in virtual colonoscopy videos. Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications International Society for Optics and Photonics 2017:1. [doi: 10.1117/12.2252281]
79. Jha M, Andreas J, Thadani K, Rosenthal S, McKeown K. Corpus Creation for New Genres: A Crowdsourced Approach to PP Attachment. Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk 2010:13-20 [FREE Full text]
80. Sabou M, Bontcheva K, Derczynski L, Scharl A. Corpus annotation through crowdsourcing: towards best practice guidelines. 2014 Presented at: Proc 9th Int Conf Lang Resources Eval; 2014; Reykjavik URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/497_Paper.pdf
81. Simperl E. How to use crowdsourcing effectively: guidelines and examples. LIBER Q 2015 Aug 18;25(1):18-39. [doi: 10.18352/lq.9948]
82. Gadiraju U, Fetahu B, Kawase R. Training workers for improving performance in crowdsourcing microtasks. Design Teaching Learn Netw World 2015:110-114. [doi: 10.1007/978-3-319-24258-3_8]
83. Hube C, Fetahu B, Gadiraju U. Understanding and mitigating worker biases in the crowdsourced collection of subjective judgments. Conf Hum Factors Comput Syst 2019:1-12. [doi: 10.1145/3290605.3300637]

XSL•FO
RenderX

84. Carvalho A, Dimitrov S, Larson K. How many crowdsourced workers should a requester hire? Ann Math Artif Intell 2016 Jan 6;78(1):45-72. [doi: 10.1007/s10472-015-9492-4]
85. Figure Eight: the Essential High-Quality Data Annotation Platform. URL: https://www.figure-eight.com/ [accessed 2019-10-02]
86. Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: assessment and reporting of intercoder reliability. Human Comm Res 2002 Oct;28(4):587-604. [doi: 10.1111/j.1468-2958.2002.tb00826.x]
87. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
88. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
89. The Neo4j Graph Algorithms User Guide v3.5. URL: https://neo4j.com/docs/graph-algorithms/current/ [accessed 2019-09-13]
90. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009. URL: https://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/154/1009 [accessed 2020-06-08]

Abbreviations

ccTLD: country-code top-level domain
D-A-CH: Germany, Austria, and Switzerland
GHW: German Health Web
SVM: support vector machine
TLD: top-level domain

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by S Nagel, CL Sánchez Bocanegra; comments to author 16.03.20; revised version received 25.03.20; accepted 14.05.20; published 24.07.20.

Please cite as:
Zowalla R, Wetter T, Pfeifer D
Crawling the German Health Web: Exploratory Study and Graph Analysis
J Med Internet Res 2020;22(7):e17853
URL: http://www.jmir.org/2020/7/e17853/
doi:10.2196/17853
PMID:32706701

©Richard Zowalla, Thomas Wetter, Daniel Pfeifer. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Psychometric Properties of the Norwegian Version of the Electronic Health Literacy Scale (eHEALS) Among Patients After Percutaneous Coronary Intervention: Cross-Sectional Validation Study

Gunhild Brørs[1], MSc; Tore Wentzel-Larsen[2,3], MSc; Håvard Dalen[1,4,5], MD, PhD; Tina B Hansen[6,7], MSc, PhD; Cameron D Norman[8], MDes, PhD; Astrid Wahl[9], MSc, PhD; Tone M Norekvål[10,11], MSc, PhD; The CONCARD Investigators[10]

[1]Clinic of Cardiology, St Olavs University Hospital, Trondheim, Norway
[2]Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Oslo, Norway
[3]Norwegian Centre for Violence and Traumatic Stress Studies, Oslo, Norway
[4]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
[5]Department of Internal Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway
[6]Cardiovascular Department, Zealand University Hospital, Roskilde, Denmark
[7]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark
[8]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[9]Department of Health Sciences, University of Oslo, Oslo, Norway
[10]Department of Heart Disease, Haukeland University Hospital, Bergen, Norway
[11]Department of Clinical Science, University of Bergen, Bergen, Norway

Corresponding Author:
Tone M Norekvål, MSc, PhD
Department of Heart Disease
Haukeland University Hospital
Postbox 1400
Bergen, 5020
Norway
Phone: 47 99044635
Email: tone.merete.norekval@helse-bergen.no

Abstract

Background: Web-based technology has recently become an important source for sharing health information with patients after an acute cardiac event. Therefore, consideration of patients' perceived electronic health (eHealth) literacy skills is crucial for improving the delivery of patient-centered health information.

Objective: The aim of this study was to translate and adapt the eHealth Literacy Scale (eHEALS) to conditions in Norway, and to determine its psychometric properties. More specifically, we set out to determine the reliability (internal consistency, test-retest) and construct validity (structural validity, hypotheses testing, and cross-cultural validity) of the eHEALS in self-report format administered to patients after percutaneous coronary intervention.

Methods: The original English version of the eHEALS was translated into Norwegian following a widely used cross-cultural adaptation process. Internal consistency was calculated using Cronbach α. The intraclass correlation coefficient (ICC) was used to assess the test-retest reliability. Confirmatory factor analysis (CFA) was performed for a priori-specified 1-, 2-, and 3-factor models. Demographic, health-related internet use, health literacy, and health status information was collected to examine correlations with eHEALS scores.

Results: A total of 1695 patients after percutaneous coronary intervention were included in the validation analysis. The mean age was 66 years, and the majority of patients were men (1313, 77.46%). Cronbach α for the eHEALS was >.99. The corresponding Cronbach α for the 2-week retest was .94. The test-retest ICC for eHEALS was 0.605 (95% CI 0.419-0.743, *P*<.001). The CFA showed a modest model fit for the 1- and 2-factor models (root mean square error of approximation>0.06). After modifications

XSL•FO
RenderX

in the 3-factor model, all of the goodness-of-fit indices indicated a good fit. There was a weak correlation with age (r=–0.206). Between-groups analysis of variance showed a difference according to educational groups and the eHEALS score, with a mean difference ranging from 2.24 (P=.002) to 4.61 (P<.001), and a higher eHEALS score was found for patients who were employed compared to those who were retired (mean difference 2.31, P<.001). The eHEALS score was also higher among patients who reported using the internet to find health information (95% CI –21.40 to –17.21, P<.001), and there was a moderate correlation with the patients' perceived usefulness (r=0.587) and importance (r=0.574) of using the internet for health information. There were also moderate correlations identified between the eHEALS score and the health literacy domains appraisal of health information (r=0.380) and ability to find good health information (r=0.561). Weak correlations with the mental health composite score (r=0.116) and physical health composite score (r=0.116) were identified.

Conclusions: This study provides new information on the psychometric properties of the eHEALS for patients after percutaneous coronary intervention, suggesting a multidimensional rather than unidimensional construct. However, the study also indicated a redundancy of items, indicating the need for further validation studies.

Trial Registration: ClinicalTrials.gov NCT03810612; https://clinicaltrials.gov/ct2/show/NCT03810612

(J Med Internet Res 2020;22(7):e17312) doi:10.2196/17312

KEYWORDS

eHealth literacy; eHEALS; health literacy; percutaneous coronary intervention; psychometric properties; validation

Introduction

Electronic health (eHealth) delivery provides an opportunity to redesign and improve health care services and health information using web-based technologies that can be accessed over the internet following diagnosis and discharge from hospital [1]. eHealth interventions have shown promising results using a behavioral approach and are recommended for supporting clinical and secondary prevention care for coronary artery disease such as after coronary revascularization (eg, percutaneous coronary intervention) [2]. eHealth has also been shown to be a cost-effective solution essential to increase geographical accessibility to secondary prevention programs, particularly as an addition to existing programs or when other offers are not available. eHealth interventions can be targeted within the natural settings where patients receive access to resources at their discretion [2]. However, patient-related barriers, specifically low health literacy and socioeconomic status, remain obstacles to the large-scale deployment of eHealth in cardiology [1]. Furthermore, patients with low eHealth literacy have lower odds of using eHealth sources to communicate with health care professionals and gain access to health information [3]. Understanding the varying eHealth literacy of patients—defined as the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem [4]—is thereby pivotal when developing and implementing eHealth resources. Assessing eHealth literacy to identify skill gaps makes it possible to better assist those with low comfort levels in taking advantage of the potential benefit that eHealth can offer, and can empower patients to fully participate in health-related decision making [4]. To assess these benefits, the identification and validation of patient reported outcome measures (PROMs) assessing patients' perceived eHealth literacy skills are therefore crucial to developing efficient patient-centered eHealth information strategies in the future [5].

To date, there has been limited evidence on PROMs that are most appropriate for assessing eHealth literacy. Systematic reviews have reported that the eHealth literacy scale (eHEALS) was the only PROM used to measure eHealth literacy in more than one study [6,7]. The eHEALS assesses patients' perceived eHealth literacy skills based on the eHealth literacy Lily model, which combines six literacy types [4,8]. The categories traditional, media, and information literacy are analytic components that involve skills applicable to a broad range of information sources, whereas the scientific, computer, and health literacy categories are context-specific that rely on more situation-specific skills. Combined, these six literacy types form the foundational skills required to fully optimize patients' experiences with eHealth. The underlying theories of the eHEALS are based in part on self-efficacy theory and social cognitive theory. These two theoretical frameworks promote competencies and confidence as precursors to behavior change and skill development [4,8]. More specifically, the eHEALS is based on the premise that the core literacies in the Lily model are not static but rather process-oriented skills that evolve over time as new technologies are introduced and the personal, social, and environmental contexts change [8]. In this way, the Lily model is clearly related to social cognitive theory, as it is based on a model of causation where behavior, environmental influences, and personal factors all interact and influence each other [9]. This means that eHealth literacy is influenced by a patient's presenting health issues, educational background, health status at the time of the eHealth encounter, motivation for seeking the information, and the technologies used [4,8].

The eHEALS has been adapted to different languages in Asia [5,10-13] and Europe [14-18]. Furthermore, the psychometrics properties have been evaluated in different populations such as in students [4,12,15,19,20], adults [9,11,16-18,21], and patients with chronic diseases [5,14,22,23], as well as in different cultures in Australia [9,24] and North America [9,17,20-22]. The internal consistency reliability coefficient was shown to be acceptable (ranging from .80 to .90) in the majority of the linguistic versions of the eHEALS [4,11-13,15-19,23], indicating a reliable scale. According to construct validity, the majority of the studies supported a 1-factor model [5,10,12-14,16,22,23,25] recommended by the original scale

XSL•FO
RenderX

[4], whereas a few other studies have recommended a 2-factor [11,15,18] or 3-factor [9,21,24] model. However, all of these studies varied contextually when evaluating the dimensions of the eHEALS construct. To our knowledge, no validated version of the eHEALS from the Nordic-Baltic countries has been published to date, and there is limited evidence on its use in patients in the acute coronary care setting such as after percutaneous coronary intervention.

Therefore, in this current study, hypotheses were tested and evaluated against existing knowledge. For instance, a lower eHEALS score has been demonstrated among people with chronic illnesses [3], and differences in eHEALS scores according to age and education have been reported [5,14]. Significant relationships between the eHEALS score and physical and mental health composites among patients with heart failure have been described [23]. An association was also reported between eHealth literacy and health literacy on patients' perceptions of the usefulness of eHealth in a population with moderate-to-high cardiovascular risk [26], whereas there was a weak-to-moderate correlation between the eHEALS score and health-related internet use among patients with rheumatic disease [14]. This evidence formed the basis for our hypothesis testing according to the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) criteria for validation [27] on the relationship between eHEALS scores and demographic information, health-related internet use, health literacy, and health status in this study (Table 1).

Table 1. Hypotheses regarding the relationship between eHEALS[a] scores and demographic information, health-related internet use, health literacy, and health status based on previous evidence.

Variables	Evidence (relationship with eHEALS)[a]	CONCARD-PCI hypothesis	Analysis
Demographic information			
Age	Weak [14,16,18] or significant [5]	Weak to moderate relationship	Pearson correlation
Gender	Weak [5,16,18]	Weak relationship	*t* test
Education	Weak [14,16,18] or significant [5]	Weak relationship	ANOVA[b]
Employment	Weak [5]	Weak relationship	ANOVA
Health-related internet use			
Used the internet to find information about health	Weak [14], moderate [18], and significant [5]	Moderate relationship	*t* test
Patient's interest in using the internet for health information in general (frequency of information-seeking)	Significant [5,16]	Moderate relationship	Spearman correlation
Health literacy			
Ability to find good health information	Moderate [11]	Moderate relationship	Pearson correlation
Appraisal of health information	Positive [26]	Moderate relationship	Pearson correlation
Health status based on RAND-12[c] (mental and physical component)	Weak [5] or significant [16,23]	Moderate relationship	Pearson correlation

[a]eHEALS: Electronic health literacy scale.

[b]ANOVA: analysis of variance.

[c]RAND-12: 12-item short-form health survey.

Thus, the aim of this study was to translate and adapt the eHEALS to conditions in Norway, and to determine its psychometric properties. More specifically, we set out to determine the reliability (internal consistency, test-retest) and construct validity (structural validity, hypotheses testing, and cross-cultural validity) of the eHEALS in a self-report format administered to patients after percutaneous coronary intervention.

Methods

Design

This validation study used a cross-sectional design and was part of a larger prospective multicenter cohort study, CONCARD[PCI], which seeks to identify bottlenecks and hurdles in the patient journey, and to suggest the optimal timing of services and alignment with preferences for patients with coronary artery disease undergoing percutaneous coronary intervention [28]. The study adheres to the COSMIN taxonomy of relationships of measurement properties for reliability and construct validity throughout the validation process. The COSMIN taxonomy was developed with the aim to improve the selection of health measurement instruments. It comprises three domains (reliability, validity, and responsiveness), which contain the measurement properties [27]. The COSMIN taxonomy has been widely used for the selection of health measurement instruments for observational studies. To ensure appropriate reporting, the validation study was also performed in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statements, which constitute an established checklist of items that should be addressed in articles reporting within the three main study designs of analytical

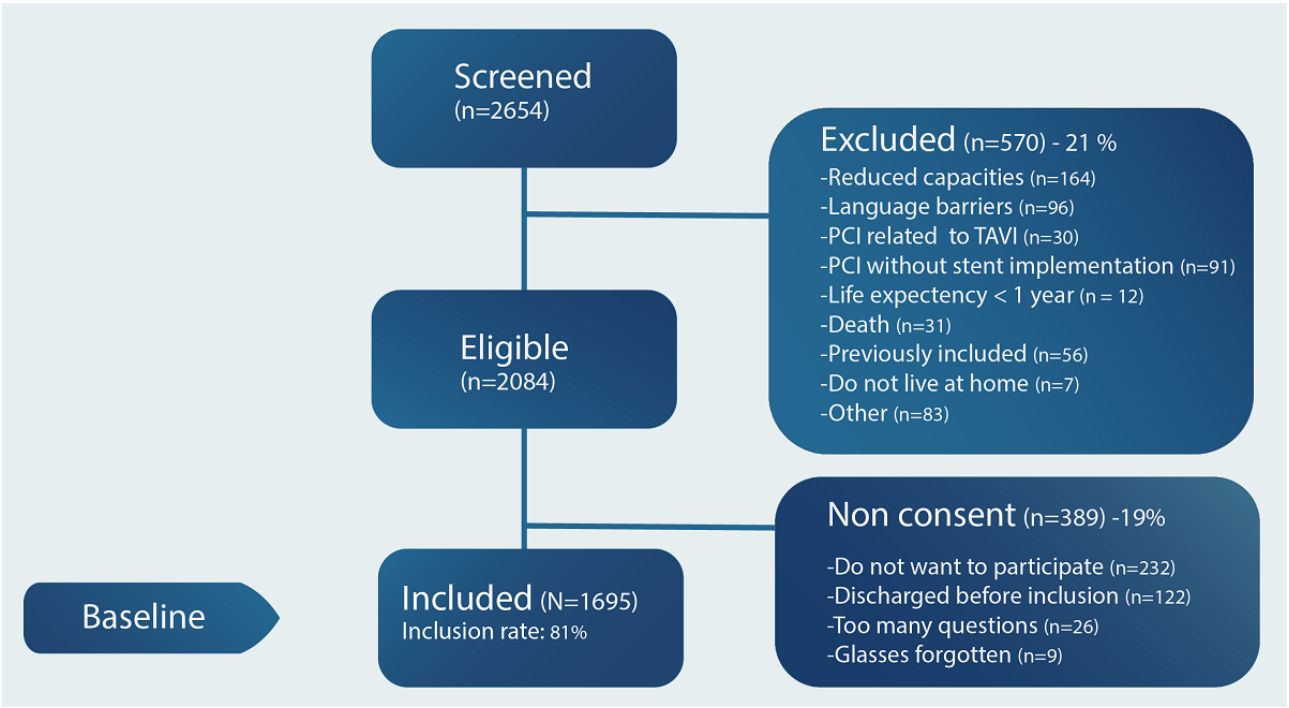
XSL•FO
RenderX

epidemiology: cohort, case-control, and cross-sectional studies [29].

Procedure and Participants

The study included 1695 patients at index admission for percutaneous coronary intervention at three large Norwegian university hospitals from June 12, 2017 through December 2018 (Figure 1). These three Norwegian university hospitals were selected based on the presence of a committed research team, including CONCARD[PCI] study nurses and a local principal investigator, and prior research experience, including research infrastructure, geographic location, and size. The percutaneous coronary intervention centers perform between 900 and 2000 (mean 1531) procedures annually, have 482 to 1400 beds (mean 860), and are referral centers for coronary angiography and percutaneous coronary intervention for a total of 17 local hospitals.

Figure 1. Flow chart of the inclusion process. PCI: percutaneous coronary intervention; TAVI: transcatheter aortic valve implantation.



Inclusion criteria were patients undergoing percutaneous coronary intervention, ≥18 years of age, living at home at the time of inclusion, and having access to electronic equipment with internet access at the time of inclusion. Exclusion criteria were the inability to speak Norwegian or inability to fill out the self-reported questionnaire due to reduced capacity. To prevent a substantial difference in follow-up time or participants responding based on different assumptions, the patients who were likely to die within less than 1 year were excluded from the study. Institutionalized patients, who may be less likely to have follow up by a primary health care provider or to use the internet to find health information, were also excluded. Similarly, patients undergoing percutaneous coronary intervention without stent implantation and patients undergoing percutaneous coronary intervention related to transcatheter aortic valve implantation or MitralClip often have other indications for the examination or treatment including other follow-up routines and were therefore also excluded from this study.

Self-reports relating to eHealth literacy, health-related internet use, health literacy, health status, as well as demographic information and clinical data identified through the Norwegian Registry for Invasive Cardiology and patient medical records were obtained before discharge from hospital after percutaneous coronary intervention. The self-report was administered using a pencil and paper survey delivered with other PROMs used as part of the CONCARD[PCI] study. A random subgroup of 100 patients was approached for an eHEALS retest after 2 weeks, 74 (74.0%) of whom completed the retest.

Translation and Cross-Cultural Adaptation of the eHEALS

A cross-cultural adaptation process was conducted to reach equivalence between the original source and the Norwegian target version of the eHEALS [30]. The translation process was conducted systematically in six steps [30]. Pilot testing was performed with a prefinal version of the eHEALS including 150 patients before being employed in the main cohort study. A summary of the overall translation procedure is described in Textbox 1. The research team encountered some difficulties in translating all of the words and phrases in the original English version of the eHEALS into Norwegian. The original eHEALS questions related to “health resources” were translated as “sources of health information” after approval from the developer of the eHEALS. The patient representatives who participated in the cognitive interviews reported that they clearly understood the items and response options and did not provide any suggestion for additional changes to the prefinal version of the instrument.

XSL•FO
RenderX

Textbox 1. Steps for translation and cross-cultural adaptation of the eHealth literacy scale (eHEALS) into Norwegian.

- Step 1: Forward translation
  - i. Two forward translations of the English eHEALS were made by two bilingual translators for whom the target language (Norwegian) was their mother tongue.
  - ii. The translators worked independently, and wrote a report (TL1 and TL2) that identified challenging phrases and described their rationale for final translation choices. An example of a difficult phrase to translate into Norwegian was "health resources."
  - iii. The two translations were compared and discrepancies were identified.
- Step 2: Synthesis
  - i. The research team synthesized the reports (TL1 and TL2) into one consensus version (TL3) and described how they resolved discrepancies.
- Step 3: Back translation
  - i. Two individuals who had a good understanding of English and also spoke Norwegian fluently independently translated TL3 back into English (TL4 and TL5). Neither of the translators who spoke English as their native language was aware of the original version of the eHEALS.
- Step 4: Synthesis and back translation
  - i. The research team agreed on the modified Norwegian version of the eHEALS (TL6).
  - ii. The research team discussed the timing of administration and meaning of certain words and sentences, and the Likert-type scale.
- Step 5: Instrument pilot testing
  - i. The prefinal version (TL6) was discussed with patient representatives and piloted before being employed in the large-scale cohort study.
  - ii. A cognitive interview was conducted to test the feasibility and understanding of the items. The patients were asked to read the questionnaire items as well as the instructions.
- Step 6: Revised instrument
  - i. The researchers evaluated the adapted eHEALS questionnaire (TL6) and all necessary changes were made.
  - ii. Patients who answered that they did not have access to electronic equipment with internet access in the pilot found it challenging to answer the eHEALS items. Therefore, the research team decided that the eHEALS items had low relevance and released these patients from answering the questionnaire.

Study Instruments and Measures

Characteristics of the Study Population

Demographic information collected included age, gender, civil status, smoking status, education level (secondary school, trade school, high school, college/university less than 4 years, college/university 4 years or more), and employment status (working, retired, or other, including sick leave, disability pension, seeking employment). Clinical data included medical history (peripheral vascular disease, stroke, myocardial infarction, diabetes, previous percutaneous coronary intervention, previous coronary artery bypass grafting, previous other heart surgery) and indication for percutaneous coronary intervention (stable angina pectoris, unstable angina pectoris, nonST-segment elevation myocardial infarction [NSTEMI] or ST-segment elevation myocardial infarction [STEMI]).

eHealth Literacy Scale

The original English eHEALS comprises eight items and assesses patients' own perception of their knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information [4]. The questionnaire contains two supplementary items to use alongside the eHEALS to better understand patients' interest in using the internet for health information in general. These items are not a formal part of the eHEALS and are not included in the total score. The original English questionnaire showed high levels of internal consistency (Cronbach α=.88) and modest test-retest reliability [4]. The eHEALS items were used to calculate a mean score using the half rule and were linearly transformed to a 0-100 scale, with higher scores indicating better eHealth literacy. To be able to compare the mean eHEALS score with those reported in other studies, the scale was linearly converted to an 8-40 scale, computed as 8 + (scale from 0 to 100) × (40 – 8)/100.

Health-Related Internet Use

To assess patients' health-related internet use, the following two supplementary items in the eHEALS were used: 1. How useful do you feel the internet is in helping you in making decisions about your health? 2. How important is it for you to be able to access health resources on the internet? [4]. Another question was also developed specifically for this study with "yes" or "no" as the response options: Have you used the internet to find information about health?

XSL•FO
RenderX

Health Literacy Questionnaire

The health literacy questions were selected from the Health Literacy Questionnaire (HLQ), which assesses nine separate domains of health literacy. In this study, two domains reflecting skills to use resources and critical evaluation were used: HLQ domain 5 (appraisal of health information, 5 items) and HLQ domain 8 (ability to find good information, 5 items). The first domain had a 4-point response option scale (strongly disagree to strongly agree) and the second domain had a 5-point response option scale (ranging from cannot achieve or always difficult to always easy). The items in the appraisal of the health information domain were used to calculate a total mean score ranging between 1 and 4, and the items in the ability to find good information domain were used to calculate a total mean score between 1 and 5. A low HLQ score indicates that the respondent has difficulties within the domain, and a high score indicates greater health literacy ability. In the event of more than two missing items, the domain score was regarded as missing [31]. The HLQ shows sufficient psychometric properties [31,32], and the Norwegian version of the HLQ has been translated and validated [33].

Health Status Questionnaire

The 12-item short-form survey RAND-12 [34] was used to assess overall generic health status through 12 items covering eight domains: physical functioning (2 items), bodily pain (1 item), physical role functioning (2 items), general health (1 item), vitality (1 item), social functioning (1 item), emotional role functioning (2 items), and mental health (2 items), summarized into physical and mental health composite scores. The RAND-12 questionnaire has been validated in European populations and shows sufficient properties [35].

Expected Relationships and Subgroup Means

The hypotheses testing (convergent validity, known-groups validity, and divergent validity) regarding the relationship between eHEALS scores and demographic information, health-related internet use, health literacy, and health status was formulated a priori. The hypotheses are based on evidence from previous studies on eHEALS as summarized in Table 1.

Statistical Analyses

We investigated the psychometric properties of the Norwegian version of the eHEALS by assessing the construct validity of three aspects: structural validity, hypotheses testing, and cross-cultural validity [27]. Descriptive statistics were used to summarize patients' sociodemographic characteristics, clinical data, health-related internet use, health status, and eHEALS scores. Floor and ceiling effects were estimated. Nonparametric tests were used for ordinal variables and parametric tests were used for comparisons of continuous variables. Continuous variables were characterized by the mean (SD). The missing rates for each item were calculated.

The reliability of the eHEALS was assessed by determining its internal consistency and test-retest reliability. Test-retest reliability was calculated by the intraclass correlation coefficient (ICC). Internal consistency reliability (ie, how well the items on a tool fit together) was calculated using Cronbach α, in which α>.70 was considered to be acceptable [27].

Confirmatory factor analyses (CFAs) were used to validate the extent to which the a priori hypotheses concerning dimensions of the eHEALS construct, based on theory and previous analyses, fit the actual data. CFA was used to explore the model fit of eHEALS as a 1-factor model as recommended by the original scale developer [4], in addition to a 2-factor model (information seeking: items 1-5 and 8; information appraisal: items 6 and 7) proposed by Soellner et al [15] and a 3-factor model (awareness: items 1 and 2; skills: items 3-5; and evaluate: items 6-8) proposed by Sudbury-Riley et al [9]. For the CFAs, the robust weighted least square mean and variance adjusted procedure (WLSMV) was used since the items are ordinal. The model fit was evaluated by various goodness-of-fit measures, including the model Chi square statistic with its degrees of freedom and *P* value, in addition to the root mean square error of approximation (RMSEA) (good fit<0.06) and its associated 95% CI, standardized root mean square residual (SRMR; good fit<0.08), comparative fit index (CFI; good fit>0.95), and Tucker-Lewis index (TLI; good fit>0.95) [27].

The convergent validity and divergent validity between the eHEALS and other constructs were assessed by computing Pearson correlation coefficients (*r*) between continuous variables and Spearman correlation coefficients (ρ) between ordinal variables. Correlation coefficients were interpreted such that 0.3 is considered a weak correlation, 0.3 to 0.6 is considered a moderate correlation, and above 0.6 is considered a strong correlation [27,36]. For known-groups validity (eg, gender, education level, employment status, and use of internet), a *t* test or one-way analysis of variance was used. If there was an indication of significant differences (*P*<.05) between groups, a posthoc analysis test for multiple comparisons that does not assume equal variances (Tamhane T2 statistic) was used to investigate where the differences occurred.

SPSS (IBM Corp. Released 2016, IBM SPSS Statistics for Windows, Version 24.0; Armonk, NY, USA) was used for summary statistics and correlations, and for conducting statistical comparisons. Mplus (Computer software, 1998-2018, version 7) developed by BO Muthén and LK Muthen, was used to perform CFAs.

Ethical Considerations

The study was approved by the Norwegian Regional Committee for Ethics in Medical Research (REK 2015/57). All patients provided written informed consent, and confidentiality and the right to withdraw from the study were assured. The study conformed with the ethical principles outlined in the Declaration of Helsinki.

Results

Characteristics of the Study Population

A total of 1695 patients consented to participate in the study (Figure 1). The mean age of the patients was 66 years, ranging from 30 to 96 years. The majority of patients were men, married/living with partner, and hospitalized for an acute coronary event (unstable angina pectoris, NSTEMI, or STEMI). Of the patients included at index admission for percutaneous coronary intervention, a strong majority reported that they have

XSL•FO
RenderX

access to electronic equipment with internet access. Most patients reported that they used the internet to find information about health (Table 2). Overall, 37.27% (499/1339) of the patients stated that the internet was useful for making decisions concerning their health and 41.51% (555/1337) stated that it was important to them that they could access health resources on the internet (Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Demographic and clinical characteristics of patients after percutaneous coronary intervention (N=1695).[a]

Characteristic	Value	N[a]
Age (years), mean (SD)	66 (10)	1695
Gender (male), n (%)	1313 (77.46)	1695
Civil status, n (%)		1529
Married/Living with partner	1173 (76.72)	
Living alone	356 (23.28)	
Smoking status, n (%)		1561
Current smoker	372 (23.83)	
Previous smoker (>1 month)	713 (45.68)	
Never smoked	476 (30.49)	
Education level attained, n (%)		1522
Secondary school	331 (21.75)	
Trade school	543 (35.68)	
High school	156 (10.25)	
College/university (<4 years)	269 (17.67)	
College/university (≥4 years)	223 (14.65)	
Employed, n (%)		1544
Working	559 (36.20)	
Retired	771 (49.94)	
Other (sick leave, disability pension, seeking employment)	214 (13.86)	
Medical history, n (%)		1685
Peripheral vascular disease	129 (7.66)	
Stroke	72 (4.27)	
Myocardial infarction	346 (20.53)	
Diabetes	314 (18.63)	
Previous PCI[b]	426 (25.28)	
Previous CABG[c]	180 (10.68)	
Previous other heart surgery	19 (1.13)	
Indication for PCI, n (%)		1695
SAP[d]	473 (27.91)	
UAP[e]	266 (15.69)	
NSTEMI[f]	522 (30.80)	
STEMI[g]	346 (20.41)	
Other	88 (5.19)	
Access to electronic equipment with internet access, n (%)	1402 (93.66)	1497
Used the internet to find information about health, n (%)	980 (66.08)	1483

[a]Number of observations for each characteristic may not total 1695 because of missing data.

[b]PCI: percutaneous coronary intervention.

[c]CABG: coronary artery bypass grafting.

[d]SAP: stable angina pectoris.

[e]UAP: unstable angina pectoris.

[f]NSTEMI: nonST-segment elevation myocardial infarction.

XSL•FO
RenderX

[g]STEMI: ST-segment elevation myocardial infarction.

Psychometric Analyses

General Properties

The mean eHEALS score was 25.66 (SD 6.23). The highest mean of the eHEALS items was 3.40 and the lowest mean was 2.92. Among all respondents, 80% were most likely to select one (41%) or two (39%) response options across all items, with 34%-51% responding "undecided" and 22%-47% responding "agree" (Multimedia Appendix 1). In total, 45 (3%) maximum possible scores and 27 (2%) minimum possible scores were identified, indicating limited ceiling and floor effects. The total mean eHEALS score for the retest was 53.52 (SD 19.79), with a floor of 5.6% (n=4) and ceiling of 1.4% (n=1).

Reliability

Cronbach α for the eHEALS was >.99 (Table 3). The corresponding Cronbach α for the 2-week retest was .94. The ICC for the eHEALS was 0.605 (95% CI 0.419-0.743, *P*<.001), indicating moderate stability over time.

Table 3. Mean (SD) scores and Cronbach α values of the eHEALS[a], HLQ[b], and RAND-12[c] of patients after percutaneous coronary intervention (N=1659).

Item	Mean (SD)	Cronbach α
eHEALS[a]	25.66 (6.23)	>.999
HLQ[b]		
HLQ 5[e]	2.43 (0.66)	.844
HLQ 8[d]	3.22 (0.73)	.875
RAND-12[e]		
PCS[f]12	43.93 (10.88)	N/A[g]
MCS[h]12	46.48 (11.14)	N/A

[a]eHEALS: eHealth literacy scale.

[b]HLQ: health literacy questionnaire.

[c]HLQ 5:Appraisal of health information.

[d]HLQ 8: Ability to find good health information.

[e]RAND-12: 12-item short-form health survey.

[f]PCS: physical composite score.

[g]N/A: not applicable; since PCS12 and MCS12 of RAND-12 are not computed as means or sum scores, there is no Cronbach α.

[h]MCS: mental health composite score.

Structural Validity

The strong standardized factor loadings for the 1-factor model, ranging from 0.79 to 0.93, indicated promising item properties. The Chi square test of model fit (*P*<.001), SRMR, CFI, and TLI indices suggested a good fit. However, the high RMSEA value suggested a poor structural fit of the eHEALS in the 1-factor model.

For the 2-factor model, standard factor loadings ranged from 0.80 to 0.93. Similar to the 1-factor model, this model suggested a good fit based on the SRMR, CFI, and TLI, and a poor fit for the RMSEA (Table 4). Furthermore, examination of the modification indices through pairing items 5 and 8 to the second (appraisal) factor did not suggest an appreciable improvement in fit for the RMSEA, although it was slightly reduced (0.176, 90% CI 0.165-0.187).

Standard factor loadings in the 3-factor model ranged from 0.84 to 0.97. Similar to the 1- and 2-factor models, the CFA supported a good fit for the three indices SRMR, CFI, and TLI, whereas the RMSEAs remained high (Table 4). Examination of the modification shown in the output file conducted by pairing item 3 in the first (awareness) and second (skills) factor, and item 5 in the second (skills) and third factor (evaluate) suggested an improvement in the model fit (SRMR=0.008, CFI=0.999, TLI=0.997, RMSEA=0.057; 90% CI 0.045-0.070). The standard factor loadings and correlations for the modifications are presented in Figure 2.

XSL•FO
RenderX

Table 4. Goodness-of-fit indices of the eHEALS[a] 1-, 2-, and 3-factor structure model.

Model	Chi square (df)	RMSEA[b] (90% CI)	SRMR[c]	CFI[d]	TLI[e]
Model 1[f]	1649.256 (20)	0.247 (0.237-0.257)	0.045	0.966	0.952
Model 2[g]	1482.130 (19)	0.240 (0.230-0.251)	0.040	0.969	0.955
Model 3[h]	510.925 (17)	0.148 (0.137-0.159)	0.019	0.990	0.983

[a]eHEALS: eHealth literacy scale.

[b]RMSEA: root mean square error of approximation.

[c]SRMR: standardized root mean square residual.

[d]CFI: comparative fit index.

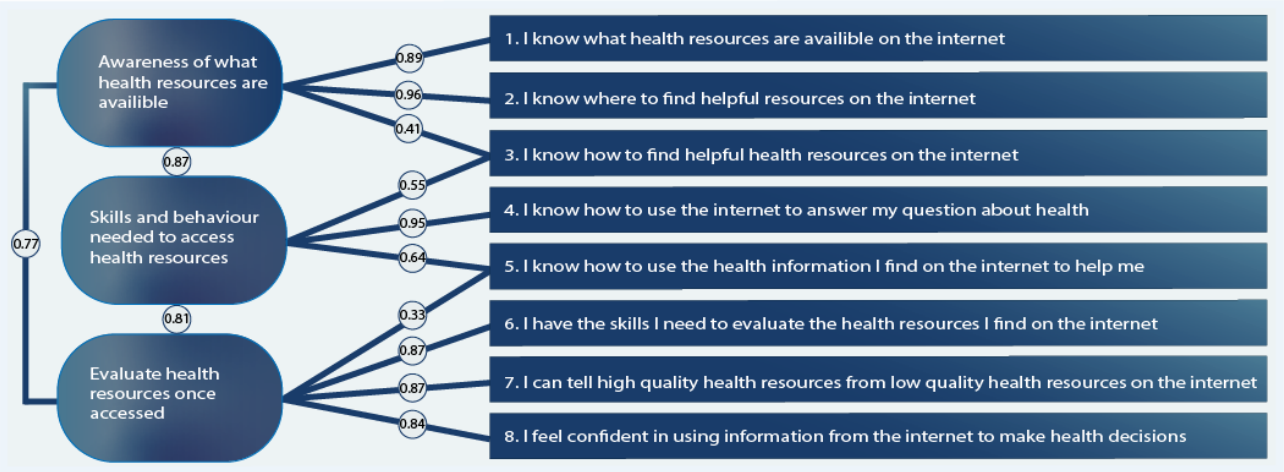
[e]TLI: Tucker-Lewis index.

[f]1-factor model: Factor 1:1-8 [4].

[g]2-factor model: Factor 1: 1-5, 8; Factor 2: 6, 7 [15].

[h]3-factor model: Factor 1: 1, 2; Factor 2: 3-5; Factor 3: 6-8 [9].

Figure 2. Electronic health literacy scale (eHEALS) 3-factor model proposed by Sudbury-Riley et al [9] with modification for items 1 and 5.



Convergent and Discriminant Validity

Demographic Information

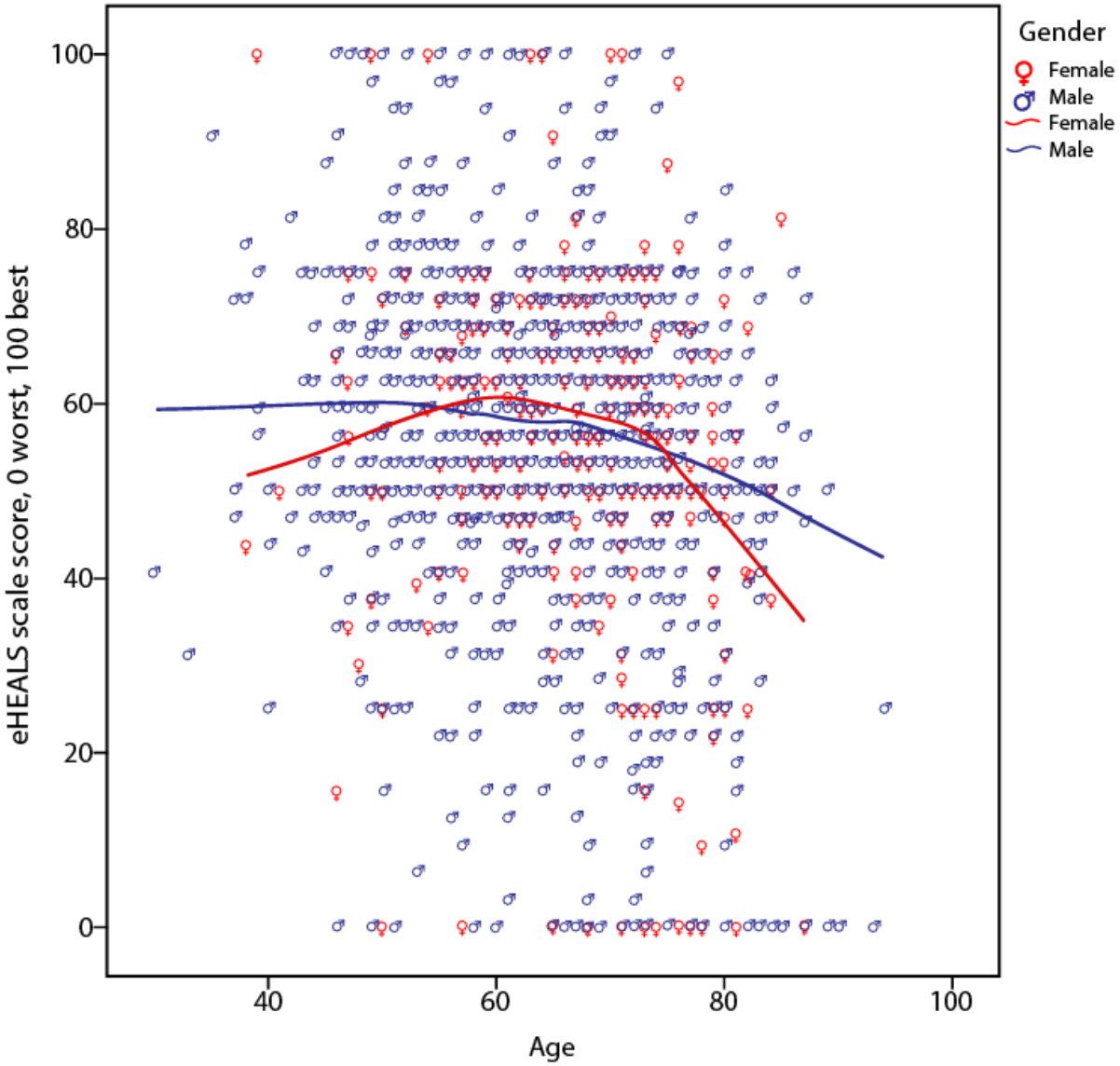
Pearson correlation analysis showed a weak negative correlation between the eHEALS score and age. An independent-sample *t* test did not indicate substantial differences in the eHEALS score between men (mean 25.64, SD 6.15) and women (mean 25.47, SD 6.60) (t_{1313}=–0.526, *P*=.60) (Table 5 and Figure 3).

As shown in Table 5, the between-group analysis of variance showed a difference according to educational groups in the eHEALS score. A posthoc test indicated that patients with 4 or more years of college/university education scored higher on the eHEALS compared to those with secondary school (mean difference=4.61, *P*<.001), trade school (mean difference=3.23, *P*<.001), and high school (mean difference=2.24, *P*=.002) education. Patients with less than 4 years of college/university education scored higher on the eHEALS than those with secondary (mean difference=3.39, *P*<.001) and trade school (mean difference=2.00, *P*<.001) education. Patients with high school education had higher eHEALS scores than those with secondary school education (mean difference=2.37, *P*<.001).

The between-groups analysis of variance also indicated a difference according to employment groups and eHEALS scores (Table 5). The posthoc test indicated a higher eHEALS score for patients who were employed compared to those who were retired (mean difference=2.31, *P*<.001). Similarly, the “other” patients group showed a higher eHEALS score compared to that of the retired patients (mean difference=1.28, *P*=.05).

XSL•FO
RenderX

Figure 3. Association between electronic health literacy scale (eHEALS) scores, gender, and age. The scale was linearly transformed to a 0-100 scale. The scale was linearly converted to an 8-40 scale (scale from 8 to 40 computed as 8 + [scale from 0 to 100] × [40 – 8]/100). The eHEALS scale in the figure is: 0=8, 20=14.4, 40=20.8, 60=27.2, 80=33.6, 100=40.



Health-Related Internet Use

An independent sample *t* test showed a higher eHEALS score for patients who reported that they used the internet to find information about health (mean 27.45, SD 5.10) than that of patients who did not (mean 21.27, SD 6.31) (95% CI –21.40 to –17.21, *P*<.001). Spearman correlation analysis revealed a moderate positive correlation between the eHEALS score and patients' perceived usefulness and importance of using the internet to find health information (Table 5).

Health Literacy

Pearson correlation analysis revealed a moderate positive correlation between the eHEALS score and the HLQ scale for appraisal of health information and the HLQ scale for ability to find good information (Table 5).

Health Status

Pearson correlation analysis revealed a weak positive correlation between the eHEALS score and self-reported health assessed with RAND-12 (Table 5).

XSL•FO
RenderX

Table 5. Group statistics and correlations between eHEALS[a] score, patients' demographic, and other instruments.

Variable	Statistic	*P* value
Demographic information		
Age, Pearson correlation coefficient	–0.206	<.001
Gender, 95% CI	–3.38-1.95	.60
Education, $F_{4,1280}$ (ANOVA[b])	21.085	<.001
Employment, $F_{2,1296}$ (ANOVA)	19.615	<.001
Health-related internet use		
Use of internet, 95% CI	–21.40 to –17.21	<.001
eHEALS supp. 1[c], Spearman correlation coefficient	0.587	<.001
eHEALS supp. 2[d], Spearman correlation coefficient	0.574	<.001
Health literacy		
HLQ[e] 5[f], Pearson correlation coefficient	0.380	<.001
HLQ 8[g], Pearson correlation coefficient	0.561 (<.001)	<.001
Health status (RAND-12[h])		
Mental component, Pearson correlation coefficient	0.116	<.001
Physical component, Pearson correlation coefficient	0.112	<.001

[a]eHEALS; eHealth literacy scale.

[b]ANOVA: analysis of variance.

[c]eHEALS supp.1: How useful do you feel the internet is in helping you in making decisions about your health?

[d]eHEALS supp.2: How important is it for you to be able to access health resources on the internet?

[e]HLQ: health literacy questionnaire.

[f]HLQ domain 5: appraisal of health information.

[g]HLQ domain 8: ability to find good information.

[h]RAND-12: 12-item short-form health survey.

Discussion

Principal Findings

To our knowledge, this is the first study to determine the psychometric properties of the eHEALS in patients after percutaneous coronary intervention. The Norwegian translation of the eHEALS appears to have acceptable construct validity. However, the high internal item consistency and the high RMSEA suggest that the fit of the data to the hypothesized models based on existing knowledge is not entirely adequate to fully capture the construct validity in this setting.

The structural validity was confirmed by three (SRMR, CFI, and TLI) out of four goodness-of-fit indices, indicating an adequate fit of the three hypothesized models. The RMSEA was lower in the 3-factor model [9] than in the 1-factor [4] and 2-factor [15] models. After two modifications of the 3-factor model, all four goodness-of-fit indices indicated a good fit. These results suggest that the eHEALS is a multidimensional construct that, as proposed by other studies, is a better fit for the 2-factor model [11,15,18] and 3-factor model [9,21,22,24]. However, in line with the current study, several previous studies conducting CFA showed an RMSEA value above the cut-off criteria for the 1-factor model [9,11,22], 2-factor model [15,18,22], and 3-factor model [9,24]. The high RMSEA indicated a poor fit, suggesting that complexity exists in all three models, but the 3-factor model was found to have acceptable fit after a low number of reasonable modifications. The differences in model fit among the three CFA models suggest that it is possible to distinguish between different conceptualizations even with high redundancy.

A high proportion of patients were most likely to select the response "undecided" or "agree" across all items, suggesting that most of the patients either considered themselves as neutral (neither disagree or agree) or relatively confident about their knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information. Although there is inconclusive evidence about the number of categories in a response scale and whether the neutral category has an impact on measurement quality [37], it has been recommended that future research should explore which response options are most appropriate to include in the eHEALS to obtain a more precise measure of eHealth literacy [22]. One explanation for the high proportion of respondents most likely to select these two response options could be that the patients experienced difficulties filling out the questionnaire in the context of an acute coronary event. Furthermore, patients with acute coronary syndrome may be more prone to survey response bias in such a manner that they select a neutral option irrespective of their

XSL•FO
RenderX

actual attitude or behavior. This suggests that the scale may work differently for patients with chronic diseases than for patients in acute settings. This finding also underpins that eHealth literacy is a process-oriented skill that evolves over time as new technologies are introduced and the personal, social, and environmental contexts change [8]. Similarly, it is also possible that the very high estimate of internal consistency is attributed to the patients' difficulties in differentiating between the meanings of the items in an acute coronary setting. A Cronbach α value above .90 has also been reported in other studies [5,10,14]. However, the very high Cronbach α in this study may indicate that there is a potential redundancy of items, wherein patients immediately after percutaneous coronary intervention may perceive that the same items are essentially rephrased in several different ways. The need for further research investigating item interpretation across populations has been suggested [14,24].

The current study indicates adequate discriminant validity of the eHEALS, and most of the demographic information and other instruments confirmed the hypotheses defined a priori. As confirmed in previous psychometric studies within general adult European populations [14,16,18], the measurement properties of the eHEALS were affected by age and gender to a lesser extent. However, the Chinese version showed a difference in eHEALS scores according to age among chronic disease patients [5]. In addition, the latter study showed that the eHEALS score was affected by education levels [5], whereas other studies reported a weak correlation with education in general adult European populations [14,18]. The results of previous studies and those of the current study diverge in this respect, indicating insufficient evidence to link perceived eHealth literacy with education. However, these findings may also be related to the fact that different methods were used to categorize and analyze education levels, and therefore should be interpreted with caution.

As the promotion of eHealth literacy takes place within a larger context, the original scale developer recommended involving other groups engaged in the literacy sectors in the work on validating the eHEALS [4]. Additionally, to address the patients' eHealth literacy level, patient integration in the evaluation is of great importance, specifically in accordance with PROMs such as health-related internet use, health literacy, and health status. A moderate correlation between the eHEALS score and patients' interest in using the internet for health information in general has been suggested, which strengthens the discriminant validity [16,18]. This provides a direction for skills, motivation, and interest that is applicable to a broad range of information sources, irrespective of the topic or context, in line with the analytic components of the Lily model [8]. Furthermore, according to the context-specific nature of eHealth literacy skills [8], a moderate correlation was found between the eHEALS score and health literacy. Consistent with previous studies that showed a relationship between eHealth literacy and health literacy [11,26], the current study indicates that patients with higher eHEALS scores tend to be "information explorers," able to identify good information and reliable sources of information, and to resolve conflicting information by themselves or with the help of others [31]. This relationship suggests that the construct validation in the scale is adequate. However, this study also showed a weak correlation between the eHEALS score and health status. This differs from other studies [16,23], suggesting modest divergent validity of the eHEALS in terms of its relationship with health status.

Strengths and Limitations

The current study has several methodological strengths and limitations that should be addressed. The stringent linguistic, cultural, and measurement adaptation procedures are likely to have contributed to strengthening the conduct of the study. However, the Norwegian eHEALS showed mixed psychometric performance, which is likely due to the context of an acute coronary event. This indicates that hospitalization can affect the response to this type of PROM. Another key strength of the study is the large sample size, which allowed us to investigate the correlations between eHEALS scores, other PROMs, and subgroups. However, the analysis of the translated eHEALS was determined to be specific to patients who underwent percutaneous coronary intervention and cannot be generalized to other scenarios. There is therefore a need to determine the psychometric properties of the eHEALS in a more diverse population and in other settings to provide empirical evidence of the generalizability of the Norwegian eHEALS. Finally, the study only determined the administration of the eHEALS in self-reported written format (paper and pencil) in a hospital acute care setting. Further work should explore other modes of administration, including online administration developed for eHealth sources such as tablets, smartphones, and email.

Conclusion

This study provides new information on the psychometric properties of the eHEALS for patients after percutaneous coronary intervention, suggesting that the eHEALS is a multidimensional construct. Nevertheless, the RMSEA is not entirely adequate to fully capture the construct validity based on existing knowledge, and further factorial validation studies are needed. The internal item consistency was very high, indicating a redundancy of items. There is nonetheless a need for more research on the psychometric properties of the eHEALS. Moreover, use of the eHEALS in this study identified areas of eHealth literacy that are important for the further development of eHealth as a source of health information.

Acknowledgments

The authors are grateful to Trond R Pettersen for forward translation, and Marie Norekvål Hayes and Alexander Sæløen for backward translation. Furthermore, Jon Martin Brørs and Rune Stiansen participated in the expert group. The authors are grateful to the CONCARD^{PCI} team for including the patients, and to the patients who agreed to participate in the study and share their experiences. The authors thank Marie Norekvål Hayes for development of Figures 1-3, and Linn Benjaminsen Hølvold for

XSL•FO
RenderX

technical support with Endnote. A major grant from the Western Norway Health Authority supports the CONCARD[PCI] project (grant no. 912184). GB was supported by an educational grant from the CONCARD[PCI] project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Response frequencies (%) and mean (SD) for all the eHEALS items, including supplementary items (N=1695).
[DOCX File , 17 KB - jmir_v22i7e17312_app1.docx]

References

1. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-Cardiology Working Group Position Paper: Overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
2. Brørs G, Pettersen TR, Hansen TB, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. BMC Health Serv Res 2019 Jun 10;19(1):364 [FREE Full text] [doi: 10.1186/s12913-019-4106-1] [Medline: 31182100]
3. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
4. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
5. Chang A, Schulz PJ. The Measurements and an Elaborated Understanding of Chinese eHealth Literacy (C-eHEALS) in Chronic Patients in China. Int J Environ Res Public Health 2018 Jul 23;15(7):1553 [FREE Full text] [doi: 10.3390/ijerph15071553] [Medline: 30041420]
6. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowl Manag E-Learn Int J 2015;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
7. Neter E, Brainin E. Association Between Health Literacy, eHealth Literacy, and Health Outcomes Among Patients With Long-Term Conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
8. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
9. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
10. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
11. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Heal Psychologist 2015;17(6):275-280 [FREE Full text]
12. Bazm S, Mirzaei M, Fallahzadeh H, Bazm R. Validity and reliability of Iranian version of eHealth literacy scale. J Community Health Res 2016;5(2):121-130.
13. Chung S, Park BK, Nahm ES. The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
14. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
15. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
16. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jun;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
17. Aponte J, Nokes KM. Validating an electronic health literacy scale in an older hispanic population. J Clin Nurs 2017 Sep;26(17-18):2703-2711. [doi: 10.1111/jocn.13763] [Medline: 28207962]
18. Diviani N, Dima AL, Schulz PJ. A Psychometric Analysis of the Italian Version of the eHealth Literacy Scale Using Item Response and Classical Test Theory Methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]

XSL•FO
RenderX

19. Gazibara T, Cakic J, Cakic M, Pekmezovic T, Grgurevic A. eHealth and adolescents in Serbia: psychometric properties of eHeals questionnaire and contributing factors to better online health literacy. Health Promot Int 2019 Aug 01;34(4):770-778. [doi: 10.1093/heapro/day028] [Medline: 29800141]
20. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct Validity of the eHealth Literacy Scale (eHEALS) Among Two Adult Populations: A Rasch Analysis. JMIR Public Health Surveill 2016 May 20;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
21. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and Validity of the Telephone-Based eHealth Literacy Scale Among Older Adults: Cross-Sectional Survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
22. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
23. Lin C, Broström A, Griffiths MD, Pakpour AH. Psychometric Evaluation of the Persian eHealth Literacy Scale (eHEALS) Among Elder Iranians With Heart Failure. Eval Health Prof 2019 Feb 11:163278719827997. [doi: 10.1177/0163278719827997] [Medline: 30744419]
24. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-Factor Structure of the eHealth Literacy Scale Among Magnetic Resonance Imaging and Computed Tomography Outpatients: A Confirmatory Factor Analysis. JMIR Hum Factors 2018 Feb 19;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
25. Paramio Pérez G, Almagro BJ, Hernando Gómez Á, Aguaded Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
26. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a Population With Moderate-to-High Cardiovascular Risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
27. De Vet HCW, Terwee C, Mokkink L, Knol D. Measurement in medicine: A practical guide. Cambridge: Cambridge University Press; 2011:1-328.
28. Norekvål TM, Allore HG, Bendz B, Bjorvatn C, Borregaard B, Brørs G, CONCARD Investigators. Rethinking rehabilitation after percutaneous coronary intervention: a protocol of a multicentre cohort study on continuity of care, health literacy, adherence and costs at all care levels (the CONCARD). BMJ Open 2020 Feb 12;10(2):e031995 [FREE Full text] [doi: 10.1136/bmjopen-2019-031995] [Medline: 32054625]
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
30. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
31. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
32. Elsworth GR, Beauchamp A, Osborne RH. Measuring health literacy in community agencies: a Bayesian study of the factor structure and measurement invariance of the health literacy questionnaire (HLQ). BMC Health Serv Res 2016 Sep 22;16(1):508 [FREE Full text] [doi: 10.1186/s12913-016-1754-2] [Medline: 27659559]
33. Urstad K, Andenæs R, Wahl AK, Kvarme LG, Helseth S, Moum TA. he Health Literacy Questionnaire (HLQ): Initial validity testing in a Norwegian sample. Health Lit Res Pract 2020;[Accepted August 26, 2019]:(forthcoming) [Accepted August 26, 2019].
34. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
35. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7] [Medline: 9817135]
36. Brown TA. Confirmatory factor analysis for applied research 2nd edition. New York: Guildford Press; 2006.
37. DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. Qual Quant 2018;52(4):1523-1559. [doi: 10.1007/s11135-017-0533-4] [Medline: 29937582]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
COSMIN: Consensus-based standards for the selection of health status measurement instruments
eHEALS: electronic health literacy scale

XSL•FO
RenderX

eHealth: electronic health
HLQ: health literacy questionnaire
ICC: intraclass correlation coefficient
NSTEMI: nonST-segment elevation myocardial infarction
PROM: patient reported outcome measure
RAND-12: 12-item short-form health survey
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
STEMI: ST-segment elevation myocardial infarction
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by L Sudbury-Riley, E Neter; comments to author 23.01.20; revised version received 10.02.20; accepted 23.04.20; published 28.07.20.

Please cite as:
Brørs G, Wentzel-Larsen T, Dalen H, Hansen TB, Norman CD, Wahl A, Norekvål TM, The CONCARD Investigators
Psychometric Properties of the Norwegian Version of the Electronic Health Literacy Scale (eHEALS) Among Patients After Percutaneous Coronary Intervention: Cross-Sectional Validation Study
J Med Internet Res 2020;22(7):e17312
URL: https://www.jmir.org/2020/7/e17312
doi:10.2196/17312
PMID:32720900

©Gunhild Brørs, Tore Wentzel-Larsen, Håvard Dalen, Tina B Hansen, Cameron D Norman, Astrid Wahl, Tone M Norekvål, The CONCARD Investigators. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Requirements of Health Data Management Systems for Biomedical Care and Research: Scoping Review

Leila Ismail[1], PhD; Huned Materwala[1], MS; Achim P Karduck[2], PhD; Abdu Adem[3], PhD

[1]Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

[2]Faculty of Informatics, Furtwangen University, Furtwangen, Germany

[3]College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

Corresponding Author:
Leila Ismail, PhD
Department of Computer Science and Software Engineering
College of Information Technology
United Arab Emirates University
Al Maqam Campus
Al Ain, Abu Dhabi, 15551
United Arab Emirates
Phone: 971 37673333 ext 5530
Email: leila@uaeu.ac.ae

Abstract

Background: Over the last century, disruptive incidents in the fields of clinical and biomedical research have yielded a tremendous change in health data management systems. This is due to a number of breakthroughs in the medical field and the need for big data analytics and the Internet of Things (IoT) to be incorporated in a real-time smart health information management system. In addition, the requirements of patient care have evolved over time, allowing for more accurate prognoses and diagnoses. In this paper, we discuss the temporal evolution of health data management systems and capture the requirements that led to the development of a given system over a certain period of time. Consequently, we provide insights into those systems and give suggestions and research directions on how they can be improved for a better health care system.

Objective: This study aimed to show that there is a need for a secure and efficient health data management system that will allow physicians and patients to update decentralized medical records and to analyze the medical data for supporting more precise diagnoses, prognoses, and public insights. Limitations of existing health data management systems were analyzed.

Methods: To study the evolution and requirements of health data management systems over the years, a search was conducted to obtain research articles and information on medical lawsuits, health regulations, and acts. These materials were obtained from the Institute of Electrical and Electronics Engineers, the Association for Computing Machinery, Elsevier, MEDLINE, PubMed, Scopus, and Web of Science databases.

Results: Health data management systems have undergone a disruptive transformation over the years from paper to computer, web, cloud, IoT, big data analytics, and finally to blockchain. The requirements of a health data management system revealed from the evolving definitions of medical records and their management are (1) medical record data, (2) real-time data access, (3) patient participation, (4) data sharing, (5) data security, (6) patient identity privacy, and (7) public insights. This paper reviewed health data management systems based on these 7 requirements across studies conducted over the years. To our knowledge, this is the first analysis of the temporal evolution of health data management systems giving insights into the system requirements for better health care.

Conclusions: There is a need for a comprehensive real-time health data management system that allows physicians, patients, and external users to input their medical and lifestyle data into the system. The incorporation of big data analytics will aid in better prognosis or diagnosis of the diseases and the prediction of diseases. The prediction results will help in the development of an effective prevention plan.

(J Med Internet Res 2020;22(7):e17508) doi:10.2196/17508

XSL•FO
RenderX

KEYWORDS

big data; blockchain; data analytics; eHealth; electronic medical records; health care; health information management; Internet of Things; medical research; mHealth

Introduction

The notion of health data management systems has evolved during the last century. With the evolution of medical records from paper charts to electronic health records (EHRs) [1], health data management has undergone disruptive transitions to provide more accurate and better patient care and make qualitative use of these records. This shift is underpinned by the advancement in information technologies that led to the development of several notions of health data management systems. Those health data management systems were often misaligned with the goals of biomedical care and research. This misalignment is caused particularly by the discrepancies between advanced technologies and their adoption for biomedical care and research. Consequently, it becomes vital to address this gap by developing a new framework for the health data management system. In this paper, we provide a broader history and evolution of health data management systems underpinned by the changing definition of medical records, discuss the issues prevailing within, introduce the modern aspects of health data management systems supporting the growing size of medical data, and discuss insights and provide solutions aiming for a better health care ecosystem.

The introduction of EHRs has transformed the health care industry by providing more services, improving the quality of patient care, and enhancing the data access ability in real time, thereby creating a diverse set of health data management systems [2]. Our understanding of EHRs is that it provides a longitudinal view of a patient's medical history over his or her lifetime generated by one or more health care providers or medical organizations delivering treatments to that patient. These cohesive and summarized records include the patient's demographic and personal information, past and current diagnoses and treatments, progress notes, laboratory and radiology results, allergies, and immunizations [1]. However, an earlier form of EHRs referred to as paper charts involves written records of a patient's diagnosis and treatments for the purpose of medical teaching. Next, the term has been revised to computer-based patient records, electronic medical records, and currently EHRs. With the advancement in technological developments and the goal to provide better and efficient health care, health data management systems have evolved from a computer-based approach to client-server–based, cloud, the Internet of Things (IoT), and finally to blockchain-based system.

With the rise of big health care data and the realization of using medical data for governance and research, it becomes necessary to integrate big data analytics within health data management systems [3]. However, this brings new challenges of data aggregation and preprocessing from multiple sources to develop insights, data security, and privacy to cope with an increasing number of data breaches and hacking incidents [4]. Further challenges have been imposed on biomedical care and research by the nature and types of medical data being generated at a rapid pace. These challenges have developed the need for a new health data management framework.

This paper analyzes the requirements for better patient care and predictive analysis that must be considered when implementing a health data management system. Considering these requirements will make the health care data management system more accurate, efficient, and cost-effective. To our knowledge, this is the first analysis of the temporal evolution of health data management systems to give insights into the system requirements for better health care.

The contributions of this paper are three-fold. First, the paper provides a taxonomy of health data management systems based on their technological advancement, and the inherent challenges and issues are discussed therein. Second, we present the reforming definitions of medical records and extract the requirements of a health data management system. Third, the paper provides insights into the health data management system research and guidelines for the future research area.

Related Works

Health data management systems are evolving for better health care. Literature reviews on these systems are classified into 2 categories: (1) electronic health (eHealth) [5-8] and (2) mobile health (mHealth) [9].

Regarding eHealth, the study by Jamal et al [5] reviews the impact of a computerized system on the quality of health care. The results showed that a health information system, if properly designed, can prevent medical errors and can support doctors and medical providers in diagnosis. The study by Van De Belt et al [6] reviews the definitions of health and medicine over 2 years (from 2007 to 2009), coming up with a common definition involving the web, patients, professionals, social networking, health information content, and collaboration. In this study, we reveal additional requirements needed for better health care: privacy, security, public insights, and patient participation in accessing and monitoring medical data. The studies by Hans et al [7] and Cunningham et al [8] focus on the definitions of eHealth from 1999 to 2004. The authors found that the themes health and technologies are most recurrent in all definitions.

Concerning mHealth, Silva et al [9] provide a review of mHealth apps and services. It highlights that the coordination, integration, and interoperability between different mHealth apps is important for better health care as well as improved performance of mobile devices in terms of device battery, storage, computation, and network.

In this study, we reviewed health data management systems based on the following 7 requirements across studies conducted over the years: (1) medical record data, (2) real-time data access, (3) patient participation, (4) data sharing, (5) data security, (6) patient identity privacy, and (7) public insights.

XSL•FO
RenderX

Methods

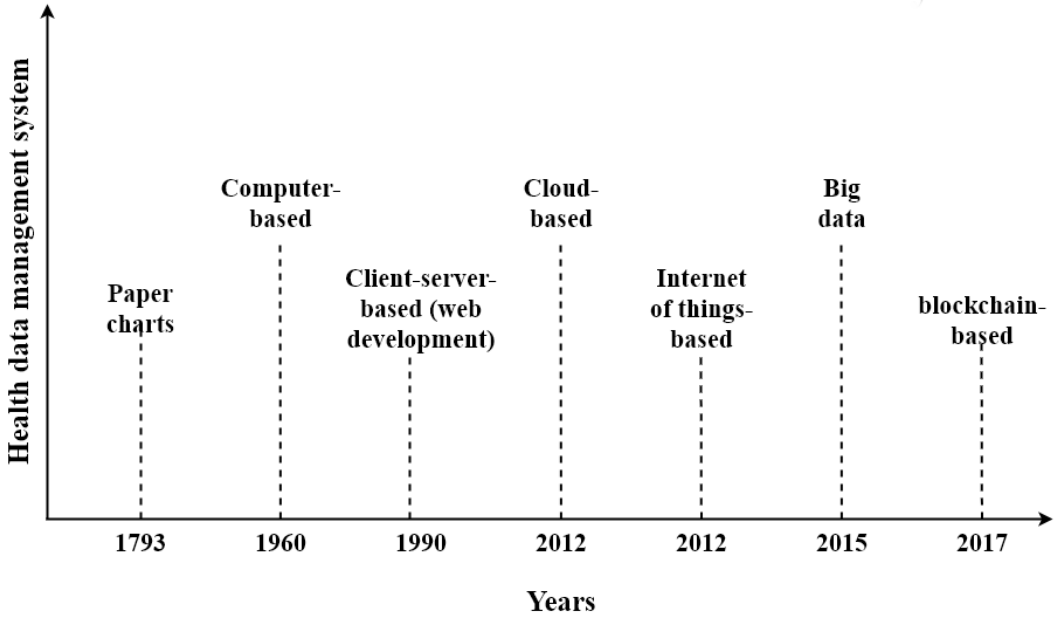
For the analysis and study of the evolution of health data management systems, we reviewed published research articles, reports, medical lawsuits, and health care regulations; acts about the methods of organizing medical record data; and the needs of a health data management system. The literature was searched in the Institute of Electrical and Electronics Engineers, Association for Computing Machinery, Elsevier, MEDLINE, PubMed, Scopus, and Web of Science databases from 1793 to 2020. We selected the papers that included incidents that involved the definitions of a health data management system, triggered the introduction of a new system, and/or implemented technologies for better health care. The analysis of these papers shows that advances in technologies are being adopted for accurate and efficient patient care.

Results

Taxonomy of Health Data Management Systems

Before satisfying the requirements of biomedical care and research, the evolution of the underlying health data management systems and their limitations must be understood. The capabilities of the health data management should ensure that the requirements of patient care are met. Health data management systems have undergone multiple transitions over the years alongside the advancement in information technologies as shown in Figure 1. During this evolution, several programs were established and regulation acts were passed to improve the quality of patient care. Table 1 presents the events that triggered the evolution of health data management systems. Table 2 presents the limitations of health data management systems.

Figure 1. Evolution of the health data management system.



XSL•FO
RenderX

Table 1. Events that have triggered the evolution of health data management systems.

Year	Responsible authority	Evolutionary change
1793	Board of Governors of the Society of the New York Hospital	Rule to record patients' data for hospital expenditure justification was passed [10].
1805	Board of Governors of the Society of the New York Hospital	Rule to record major medical cases for education was passed provoked by a fatal dispute between an American statesman and an American politician. According to the rule, the recorded cases should be bounded in a book for inspection by the governors, medical professionals and students, and the friends of the patients [11].
1830	Board of Governors of the Society of the New York Hospital	Rule to maintain a record of all the medical cases [11].
1918	American College of Surgeons	A hospital standardization program was established to standardize the format of medical records for improved patient care [12].
1928	American College of Surgeons	American Association of Record Librarians of North America was established to enhance the standards of medical records [13].
The 1960s	Lawrence Weed	The idea to use computers for medical records was proposed to allow doctors to track a patient's medical history and provide evidence for the treatment [14].
		A problem-oriented medical records model was developed to standardize the method of EMRs[a] that provided a structure to help doctors record their notes [14].
The 1960s	N/A[b]	Paper charts were termed as EMRs.
1965	Centers for Medicare and Medicaid Services	Rule to record patients' data by medical nurses for medical insurance reimbursement with the introduction of Medicare and Medicaid laws [15].
1965	Lockheed Corporation	First commercial computerized health data management system known as Clinical Information System was developed for El Camino Hospital. The system included features for laboratory tests, appointment scheduling, and pharmacy management [16].
1967	University of Utah, 3M and Latter-Day Saints Hospital	First clinical decision support system known as Health Evaluation of Logical Processing was developed to support clinical operations. The system helped doctors to identify cardiac contraction based on a patient's test results' analysis and to select an appropriate medication for infectious disease cases [17].
1968	Massachusetts General Hospital and Harvard University	The first modular computer-based health data management system known as Computer Stored Ambulatory Record was implemented. The system accommodated clinical vocabularies through clinical mapping to recognize different terms used for the same disease [18].
The 1980s	Indian Health Service	MPI[c] was introduced to keep track of patients' medical data to reduce unnecessary testing and adverse drug effects [19].
1987	Health Level Seven	Electronic standards were developed to address the standardization issues of health data management system development and adaption. The standards allowed the use of components from different vendors in a health data management system [20].
1991	Institute of Medicine	The term computer-based patient records was introduced in a report studying the benefits of electronic management of health records [21].
1996	US Congress	The Health Insurance Portability and Accountability Act was passed to safeguard patients' medical records by involving role-based access control, automatic data backup, audit trails, and data encryption [22].
1999	John Mitchell	The term eHealth[d] was coined that refers to the integration of electronic communication and information technologies for electronic transmission, storage, and retrieval of medical records both locally and remotely [23].
2000	S Laxminarayan and Robert SH Istepanian	The term mHealth[e] was coined that refers to wireless telemedicine using mobile telecommunications and multimedia technologies for the new mobile health care system [24].
2001	Gunther Eysenbach	The definition of eHealth was expanded by incorporating business and public health to health services and defining the outcomes and stakeholders of eHealth [25].
2004	Stephen S Intille	The term uHealth[f] was coined that refers to the use of biometric sensors and medical devices to monitor and improve a patient's medical health [26].
The 2000s	Health care organizations	Proposed a formal definition of the term personal health records that allows patients to access their medical history and to manage it by making part of it available to selected participants by defining access control rights [27].
2003	Institute of Medicine	The term electronic health records [28] was coined.

XSL•FO
RenderX

Year	Responsible authority	Evolutionary change
2006	Commonwealth of Massachusetts	Massachusetts health care reform law was passed that mandated for residents to have minimum medical insurance coverage and for employers with more than 10 full-time employees to provide medical insurance coverage [29].
2006	Elliott Fisher	The term Accountable Care Organizations was coined that refers to a group of doctors, hospitals, and other health care providers who volunteer to give high-quality care to their patients to avoid unnecessary duplication of services and reduce medical errors [30].
2009	US Department of Justice, Office of Inspector General, and Human and Health Services	The Health care Fraud Prevention and Enforcement Action was established to strengthen the existing programs to prevent and reduce Medicare and Medicaid frauds [31].
2010	US President Barack Obama	The Patient Protection and Affordable Care Act was signed into law with an objective to provide an expansion of medical insurance coverage [32].

[a]EMR: electronic medical record.

[b]N/A: not applicable.

[c]MPI: master patient index.

[d]eHealth: electronic health.

[e]mHealth: mobile health.

[f]uHealth: ubiquitous health

Table 2. Limitations of health data management systems.

Health data management system	Limitation
Paper charts	Illegible handwriting resulting in incorrect treatments [33] and deaths [34,35]. Requires physical storage and are susceptible to unplanned destruction such as flood, fire, rodents, and degradation. Physically cumbersome to read, understand, and search for specific information. The cost and time required for paper charts to be requested for duplication and then delivered are unacceptably high.
Computer-based	Medical records are managed by the physicians and cannot be accessed by the patients. Physicians visiting a patient have to note down or memorize the patient's medical data to return to the hospital and record it digitally, which may lead to error.
Client-server-based	A patient has no traceability on how his or her data are used. The issues of security, privacy, and single point of failure. In addition, a cohesive view of a patient's medical data from multiple hospitals is difficult. Requires repeating medical tests at times, which results in more time, cost, and effect on health conditions.
Cloud-based	Single point of failure, loss of data control and stewardship, a requirement of steady internet connection, and data reliability [36,37].
IoT[a]-based	Data security and patient privacy are a major concern.
Big-data–based	The process of data aggregation from different storage sites is time consuming, complex, and expensive. The data are stored using different formats and requires preprocessing. In addition, preserving the security of the data and privacy of the patient identity while maintaining the usefulness of data for analysis and studies is quite challenging.
Blockchain-based	The process of ledger update on multiple nodes is energy consuming [38] and suffers from the issue of low throughput [39].

[a]IoT: Internet of Things.

Requirements of a Health Data Management System

Over the last century, the definition of health data management has undergone numerous reformations to address the need for better and advanced patient care alongside technological advances. We evaluated these differing examinations and rationalized the definition used in the remainder of the paper. It is important to note that, as the term health data management is rather recent, the listed definitions were taken from different legislations and health data recording systems, even if the exact phrase *health data management* was not used. Table 3 shows the evolving definitions of health data management systems from being purely medical practice and learning-based definitions to being more patient-centric and research-based definitions. We classified health data management systems based on 7 requirements that underpin the evolution in the field as shown in Figure 2. Each number in the figure represents a definition stated in Table 3.

XSL•FO
RenderX

Table 3. The definitions of health data management systems.

Number	Year	Source	Definition
1	1793	Siegler [10]	"[...] Names and Diseases of the Persons, received, deceased or discharged in the same, with the date of each event, and the Place from whence the Patients last came [...]"
2	1805	Siegler [10]	"The house physician, with the aid of his assistant, under the direction of the attending physician, shall keep a register of all medical cases which occur in the hospital, and which the latter shall think worthy of preservation, which book shall be neatly bound, and kept in the library for the inspection of the friends of the patients, the governors, physicians and surgeons, and the students attending the hospital."
3	1941	Sayles and Gordon [12]	"Accurate and complete medical records [...] which includes identification data; complaint; personal and family history; history of the present illness; physical examination; special examinations such as consultations, clinical laboratory, x-ray and other examinations; provisional or working diagnosis; medical or surgical treatment; gross or microscopical pathological findings; progress notes; final diagnosis; condition on discharge; follow-up; and, in case of death, autopsy findings."
4	1968	Weed [14]	"The computer is making a major contribution [...] the patient will gain from his physician an immediate sympathetic understanding [...] inadequate analysis by the medical profession can be avoided."
5	1968	Weed [14]	"[...] orient data around each problem [...] complete list of all the patient's problems [...] diagnosis and all other unexpected findings or symptoms [...] The list is separated into active and inactive problems, and in this way, those of immediate importance are easily discernible [...] orders, plans, progress notes and numerical data can be recorded under the numbered and titled problem [...]"
6	1993	Cynthia [40]	"Digital versions of paper charts that contain the medical and treatment history of the patients from one practice for providers to use for diagnosis and treatment"
7	1997	Dick et al [21]	"Electronic patient record [...] support users through availability of complete and accurate data, practitioner reminders and alerts, clinical decision support systems, links to bodies of medical knowledge, and other aids."
8	2001	Eysenbach [25]	"[…] medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies […] an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology."
9	2002	Cameron and Turtle-Song [41]	"The subjective component contains information about the problem [...] objective information consists of those observations made by the counselor [...] assessment section demonstrates how [...] data are formulated, interpreted, and reflected upon, and the plan section summarized the treatment direction."
10	2003	Markle Foundation [42]	"[…] electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment."
11	2003	HIMSS[a] [1]	"[...] longitudinal electronic record of patient health information generated by one or more encounters [...] patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports [...] automates and streamlines the clinician's workflow. The EHRs has the ability to generate a complete record of a clinical patient encounter [...] evidence-based decision support, quality management, and outcomes reporting."
12	2003	HIMSS [43]	"The Electronic Health Record (EHR) is a secure, real-time, point-of-care, patient-centric information resource […] decision making by providing access to patient health record information where and when they need it and by incorporating evidence-based decision support […] billing, quality management, outcomes reporting, resource planning, and public health disease surveillance and reporting."
13	2005	AHIMA[b] [44]	"[...] lifelong resource of health information needed by individuals to make health decisions. Individuals own and manage the information [...] is maintained in a secure and private environment, with the individual determining rights of access [...]"
14	2008	Böcking and Trojanus [45]	"Health data management […] acquiring, entering, processing, coding, outputting, retrieving, and storing of data gathered in the different areas of health care […] also embraces the validation and control of data according to legal or professional requirements."
15	2013	HIPAA[c] [22]	"A major goal […] to protect the privacy of individuals' health information […] adopt new technologies to improve the quality and efficiency of patient care."

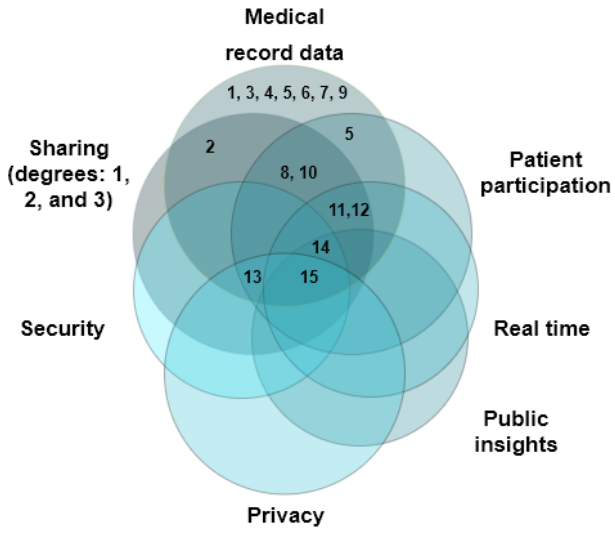
[a]HIMSS: Healthcare Information and Management Systems Society.

[b]AHIMA: American Health Information Management Association.

[c]HIPAA: Health Insurance Portability and Accountability Act.

XSL•FO
RenderX

Figure 2. Requirements of a health data management system.



Medical Record Data

Medical record data that describes the identity and health of a patient based on the personal and demographic identity, history of the medical condition, ongoing treatment, laboratory tests, and radiology results are the common requirement of a health data management system. The medical records have been a primary component throughout the evolution of health data management systems, whether in the form of printed documents or digital records.

Real-Time Data

To improve the quality of patient care, the requirement of real-time data access was highlighted in the definitions of health data management systems. This requirement reduces the medical incidents owing to the delay in data updates by the physicians. However, this requirement cannot be fulfilled by the paper-based and computer-based health data management systems. This requirement was introduced with the client-server–based management system [46-52] that enables the physicians to access and update the patient medical records in real time.

Patient Participation

With the medical records maintained by the hospitals or third-party cloud service providers, the patients cannot track how their medical data are used. Consequently, patient participation in accessing and monitoring medical data is a key requirement to develop trust in health data management systems. In addition to data access, the participation of patients in providing health conditions and lifestyle data to the physicians will aid in better prognosis and diagnosis. The introduction of IoT-based health data management system involving sensors and medical devices that monitor a patient's health and lifestyle conditions enables the patient to input their medical conditions to the system [53-59]. An analysis of personal health records management platforms based on users' perception shows that a simple easy-to-use system is required for patient engagement and satisfaction [60].

Sharing

Sharing of medical records is a vital requirement with the patient's treatment being spread across various health care providers. This is to aid other physicians to study the patient's medical history for better treatment and to avoid repetition of laboratory and radiology tests. On the basis of the list of definitions in Table 3, we classified sharing based on the users allowed to access the data into 3 different categories: (1) degree 1, where the information is shared within the same medical organization where the patient is currently receiving treatment, (2) degree 2, where the information is shared with the patient, patient's friends, and family, and (3) degree 3, where the information is shared with other medical organizations and government. The requirement of sharing is complemented by the introduction of the cloud-based health data management system [61-63]. However, to share medical record data between different health care organizations and to efficiently use the shared information, the systems should support interoperability. Interoperability can be achieved by using a standard format to store, manage, and share the medical data. There are several standard formats to store medical data and images [64]. Some of the major file formats used for medical images are Analyze [65], Neuroimaging Informatics Technology Initiative [66], Minc [67], and Digital Imaging and Communications in Medicine [68]. Health Level 7 International, standardized by the American National Standards Institute, is a health care protocol for sharing medical data [20]. It includes the rules for the integration, exchange, and management of EHRs. Wen et al [69] assessed the interoperability of eHealth systems in Taiwan for exchanging data. This is to reduce repeated medical examinations and medications for better health care. They concluded that the government should define policies to enforce interoperability.

Security

With increasing incidents of data breaching and phishing attacks, and the adoption of a third-party service provider, the security of the patients' sensitive and important data is essential. Compared with 477 health data breaches reported in 2017,

XSL•FO
RenderX

affecting 5,579,438 patient records in 2017, 503 breaches affecting 15,085,302 records were reported in 2018 [70]. The requirement of security is even high when patients' medical records are handled by a cloud service provider or when medical sensors and devices are used to gather patients' medical and lifestyle data. According to a report by Intel Security, the use of cloud services by the health care provider has reduced owing to the lack of cyber security methods implemented by the cloud service provider [71]. A report states that, on average, hospitals lose track of around 30% of their networked medical devices, making it much harder to protect against vulnerabilities [72]. More than 61% of all IoT devices and sensors on a hospital network are at high risk of cyber-attack. In recent years, blockchain technology [73,74] has gained wide popularity and has penetrated into the domain of health care to address the need for a more patient-centric supportive system for the professionals, to connect disparate systems for improved patient care, and to increase the accuracy of EHRs [75-81].

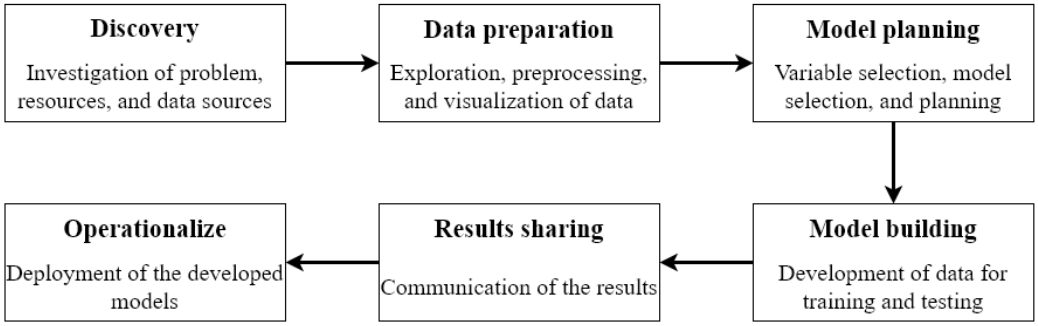
Privacy

The privacy requirement of a patient's identity in a health data management system is crucial with the increasing number of medical frauds and fake medications. The privacy of the patient cannot be compromised, especially with the rise of data analytics, where the medical record data of the patients are used for analysis. The blockchain-based health data management system aims to address this issue.

Public Insights

Prediction of health conditions is important to avoid life-threatening situations. The increasing amount of health care data [82], if properly analyzed, can facilitate the prediction of health conditions. The process of gathering, organizing, storing, and analyzing big data to discover correlations, hidden patterns, and other valuable insights is known as big data analytics. Figure 3 shows the life cycle of big data analytics.

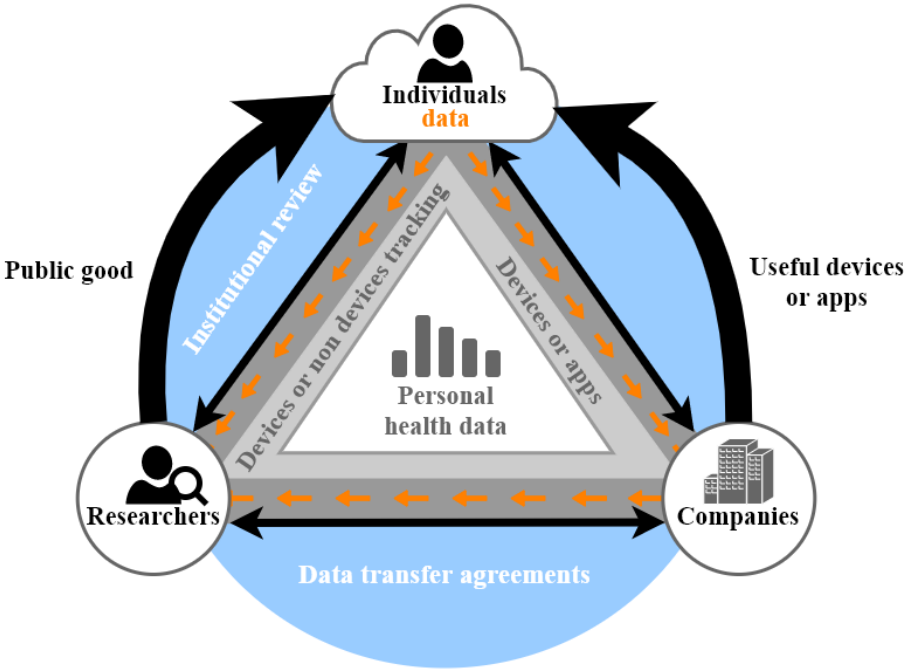
Figure 3. Lifecycle of big data analytics.



The prediction from health care data for public insights allows to actively improve public health and to react faster to a situation [83-91]. Using personal health care data requires, of course, a well-defined balance between the assurance of the privacy of personal health care data with respect to transparency, for example, toward insurance companies. Insights into genetical personal risk factors for chronic diseases should not lead to a situation where a person has disadvantages concerning the insurance status. Moreover, the monitoring of the public health situation has to be based on the health care data of individuals. Consequently, research projects have recently addressed the balance of personal health care data as a public good [92]. Figure 4 [92] shows the relationship between the 3 key stakeholders for defining the balance between personal health care data and the potential of these data as a public good. Companies could be health insurance providers, hospitals, pharmaceutical companies, and government organizations.

XSL•FO
RenderX

Figure 4. Personal health care data ecosystem.



The diabetes mellitus crisis or the growth of cardiovascular problems caused by nutrition patterns and lifestyle behavior in many countries and regions of the world, changing patterns of Alzheimer and dementia, or microbiome research, and the abuse of antibiotics would benefit tremendously from personal health care data as a public good [93,94]. Bringing together the insights of large initiatives such as the Health Data Exploration Project and Computational Health Sciences [92,94] promises the key for future advancement in the area of private and personal health care data for the public good. Health care data analytics can help researchers and government officials for better prediction of chronic diseases, the development of effective therapeutic drugs, more accurate patient care, and the development of a nation-wide effective prevention plan.

Table 4 shows health data management systems presented in the taxonomy and evaluates them in terms of their adherence to the defined requirements.

XSL•FO
RenderX

Table 4. Health data management systems in the literature vs the requirements.

System	Medical record data	Real-time data	Patient participation		Sharing			Security	Privacy	Public insights
			Data access	Data input	Degree 1	Degree 2	Degree 3			
Paper-based	Allows recording of medical data for eventual use	Encounters high delays	Does not allow patients to track the use of their medical data	Does not allow patients to provide their health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Does not allow data sharing with the patient, patient's friends, and family	Does not provide methods against cybersecurity attacks	Does not provide methods for preserving a patient's privacy	Does not support prediction
Computer-based	Allows recording of medical data for eventual use	Encounters high delays	Does not allow patients to track the use of their medical data	Does not allow patients to provide their health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Does not provide methods against cybersecurity attacks	Does not provide methods for preserving a patient's privacy	Does not support prediction
Client-server–based	Allows recording of medical data for eventual use	Allows data retrieval in real time	Allows patients to access and monitor their medical data	Does not allow patients to provide their health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Does not provide methods against cybersecurity attacks	Does not provide methods for preserving a patient's privacy	Does not support prediction
Cloud-based	Allows recording of medical data for eventual use	Allows data retrieval in real time	Allows patients to access and monitor their medical data	Does not allow patients to provide their health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Does not provide methods against cybersecurity attacks	Does not reveal a patient's identity	Does not support prediction
IoT[a]-based	Allows recording of medical data for eventual use	Allows data retrieval in real time	Allows patients to access and monitor their medical data	Allows patients to provide health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Does not provide methods against cybersecurity attacks	Does not provide methods for preserving a patient's privacy	Provides methods for the prediction of health conditions
Big data analytics	Allows recording of medical data for eventual use	Allows data retrieval in real time	Allows patients to access and monitor their medical data	Allows patients to provide health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Does not provide methods against cybersecurity attacks	Does not reveal a patient's identity	Provides methods for the prediction of health conditions

XSL•FO
RenderX

System	Medical record data	Real-time data	Patient participation		Sharing			Security	Privacy	Public insights
			Data access	Data input	Degree 1	Degree 2	Degree 3			
Blockchain-based	Allows recording of medical data for eventual use	Allows data retrieval in real time	Allows patients to access and monitor their medical data	Allows patients to provide health conditions	Supports data sharing only within the same hospital	Allows data sharing with the patient, patient's friends, and family	Allows data sharing with other medical organizations and government	Ensures the protection of medical data against cybersecurity attacks	Does not reveal a patient's identity	Provides methods for the prediction of health conditions

[a]IoT: Internet of Things.

Discussion

Principal Findings

This study revealed that there is a need for a secure and efficient health data management system that will allow physicians and patients to update decentralized medical records and to analyze the medical data for supporting more precise diagnoses, prognoses, biomedical research, and public insights. The early form of health data management using the manual recording of a patient's diagnosis and treatment on sheets of paper was introduced almost a century ago. Later, with the advancement in technology, health data management systems evolved to web, cloud, IoT, big data analytics, and blockchain-based systems. The definition of medical records has reformed alongside this temporal evolution of the system. The requirements for a health data management system extracted from these definitions are medical record data, real-time data, patient participation, sharing, security, privacy, and public insights. The paper-based health data management system fulfills the requirements of medical record data and sharing. However, paper charts are prone to misplacement, occupy large physical space, and involve a time-consuming and expensive data sharing process. Over time, the paper charts were replaced by electronic records in the computer-based system with the same requirements.

To achieve the requirement of real-time data access in addition to medical record data and sharing, a client-server–based health data management system was introduced. This system allows patients and health care providers to access medical data over the internet using a mobile device or a desktop computer. However, it suffers from the issues of single point of failure, data fragmentation, system vulnerability, low scalability, and high data security and patient privacy risks. To minimize the infrastructural cost and to address the issue of data fragmentation, the medical organizations and health care providers transitioned to a cloud-based system. The cloud service provider ensures the requirement of privacy of patient identity, but the security of the data is not ensured in addition to the issue of a single point of failure.

The requirement of patient participation to feed their medical data and lifestyle conditions for better prognosis and diagnosis was achieved with the introduction of the IoT-based management system. However, with the increasing number of data breaches and hacking of the medical sensors and devices, there prevails a constant threat to the security of data and privacy of a patient's identity. With the advancement in big data analytics, increasing amount of health care data are being studied to gain insights for better prognosis and diagnosis of diseases. However, the privacy of a patient's identity still remains a concern.

The blockchain technology, which recently attracted the attention of industries, shows potential in the field of health care. A blockchain-based health data management system satisfies all the requirements needed for better patient care. However, it consumes a high amount of energy [95,96] and has low throughput [39]. There are increasing research efforts to solve these issues. For instance, to address the problem of energy consumption, Milutinovic et al [97] proposed the proof of luck consensus mechanism that ensures energy-efficient and low-latency transaction validation. Ismail et al [98] and Dorri et al [99] proposed scalable blockchain architectures for health care that use a clustering approach to increase transactions throughput.

The main requirements of a health care data management system are security and privacy, especially with the increasing number of data breaching and hacking attacks. Furthermore, the adoption of patient participation to feed health data to a health system is increasing with the introduction of disruptive technologies, such as the IoT and big data analytics. Big data analytics requires the sharing of medical information among hospitals to get insights and predictive analysis from the data. This paves the way toward a health data management system as a support to physicians and medical professionals for better diagnosis and prognosis of chronic diseases. In addition, such a system allows to derive public insights from data to develop a nation-wide prevention plan for certain diseases. The traceability feature of the blockchain ensures that the data used for developing the predictive models is accurate, leading to a precise prognosis, diagnosis, and decision support system. Consequently, we suggest an integrated blockchain-, IoT- and big data–based health data management system to ensure the requirements of smart health care: real-time access to data by physicians and patients, health data input from patients through medical sensors and lifestyle, security, privacy, and public insights. This integrated health management system should be scalable and energy-efficient, presenting new research challenges in the research era of a smart health data management system.

XSL•FO
RenderX

Conclusions

The objective of this paper was to highlight the requirements of a health data management system for biomedical care and research. In summary, it discussed the temporal evolution of health data management systems from paper charts to blockchain-based systems, along with the reformation of the definition of what we call EHRs today. The system should satisfy the requirements of medical record data, real-time access, patient participation, data sharing, data security, patient identity privacy, and public insights. The incorporation of big data analytics aids in better prognosis and diagnosis of the diseases and the prediction of risk for the development of chronic diseases.

Acknowledgments

This work was supported by the Emirates Center for Energy and Environment Research of the United Arab Emirates University under grant 31R101. The authors would like to thank the anonymous reviewers for their valuable comments, which helped them improve the content, quality, and presentation of this paper.

Conflicts of Interest

None declared.

References

1. Healthcare Information and Management Systems Society. What Are Electronic Health Records (EHRs)? URL: https://www.himss.org/electronic-health-records [accessed 2020-02-13]
2. King J, Patel V, Jamoom EW, Furukawa MF. Clinical benefits of electronic health record use: national findings. Health Serv Res 2014 Feb;49(1 Pt 2):392-404 [FREE Full text] [doi: 10.1111/1475-6773.12135] [Medline: 24359580]
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2(1):3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
4. Kuo MH, Sahama T, Kushniruk AW, Borycki EM, Grunwell DK. Health big data analytics: current perspectives, challenges and potential solutions. Int J Big Data Intell 2014;1(1/2):114. [doi: 10.1504/ijbdi.2014.063835]
5. Jamal A, McKenzie K, Clark M. The impact of health information technology on the quality of medical and health care: a systematic review. Health Inf Manag 2009;38(3):26-37. [doi: 10.1177/183335830903800305] [Medline: 19875852]
6. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of health 2.0 and medicine 2.0: a systematic review. J Med Internet Res 2010 Jun 11;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
7. Oh H, Rizo C, Enkin M, Jadad A. What is ehealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
8. Cunningham SG, Wake DJ, Waller A, Morris AD. Definitions of eHealth. In: Gaddi A, Manca M, editors. eHealth, Care and Quality of Life. New York, USA: Springer; 2013:15-30.
9. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
10. Siegler EL. The evolving medical record. Ann Intern Med 2010 Nov 16;153(10):671-677. [doi: 10.7326/0003-4819-153-10-201011160-00012] [Medline: 21079225]
11. Thomas E, John M, Charles R, Society of the New York Hospital. An Account of the New York Hospital. Medical Center Archives 1811.
12. Sayles NB, Gordon LL. Health Information Management Technology: An Applied Approach. Chicago, USA: American Health Information Management Association; 2013.
13. American Health Information Management Association. AHIMA History URL: http://bok.ahima.org/doc?oid=58133#.XnNMSIgzbIU [accessed 2020-02-13]
14. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 14;278(11):593-600. [doi: 10.1056/NEJM196803142781105] [Medline: 5637758]
15. Centers for Medicare & Medicaid Services. 1965. CMS' Program History URL: https://www.cms.gov/About-CMS/Agency-information/History/ [accessed 2020-02-13]
16. Tripathi M. EHR evolution: policy and legislation forces changing the EHR. J AHIMA 2012 Oct;83(10):24-9; quiz 30. [Medline: 23061349]
17. Gardner RM, Pryor T, Warner HR. The HELP hospital information system: update 1998. Int J Med Inform 1999 Jun;54(3):169-182. [doi: 10.1016/s1386-5056(99)00013-1] [Medline: 10405877]
18. Amatayakul MK. Electronic Health Records: A Practical Guide for Professionals and Organizations. Chicago, USA: American Health Information Management; 2004.
19. IHS Markit. 1980. Master Patient Index (MPI) URL: https://www.ihs.gov/hie/masterpatientindex/ [accessed 2020-02-13]
20. Hammond WE. Health level 7: an application standard for electronic medical data exchange. Top Health Rec Manage 1991 Jun;11(4):59-66. [Medline: 10112038]

XSL•FO
RenderX

21. Dick RS, Steen EB, Detmer DE. The Computer-based Patient Record: An Essential Technology for Health Care. Washington, DC: National Academies Press; 1997.
22. United States Department of Health and Human Services. Summary of the HIPAA Security Rule URL: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html [accessed 2020-02-13]
23. John M. From Telehealth to E-health: The Unstoppable Rise of E-health. Australia: Department of Communications, Information Technology and the Arts; 1999.
24. Laxminarayan S, Istepanian R. Unwired e-med: the next generation of wireless and internet telemedicine systems. IEEE Trans Inf Technol Biomed 2000 Sep;4(3):189-193. [doi: 10.1109/titb.2000.5956074] [Medline: 11026588]
25. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
26. Intille SS. Ubiquitous computing technology for just-in-time motivation of behavior change. Stud Health Technol Inform 2004;107(Pt 2):1434-1437. [doi: 10.3233/978-1-60750-949-3-1434] [Medline: 15361052]
27. Kim MI, Johnson KB. Personal health records: evaluation of functionality and utility. J Am Med Inform Assoc 2002;9(2):171-180 [FREE Full text] [doi: 10.1197/jamia.m0978] [Medline: 11861632]
28. Gillies J, Holt A. Anxious about electronic health records? No need to be. N Z Med J 2003 Sep 26;116(1182):U604. [Medline: 14581956]
29. Holahan J, Blumberg L. Massachusetts health care reform: a look at the issues. Health Aff (Millwood) 2006;25(6):w432-w443. [doi: 10.1377/hlthaff.25.w432] [Medline: 16973652]
30. Centers for Medicare & Medicaid Services. 2006. Accountable Care Organizations (ACOs) URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO [accessed 2020-02-13]
31. US Department of Justice. 2016. Fact Sheet: The Health Care Fraud and Abuse Control Program Protects Conusmers and Taxpayers by Combating Health Care Fraud URL: https://www.justice.gov/opa/pr/fact-sheet-health-care-fraud-and-abuse-control-program-protects-conusmers-and-taxpayers [accessed 2020-02-13]
32. HealthCare. 2010. Patient Protection and Affordable Care Act URL: https://www.healthcare.gov/glossary/patient-protection-and-affordable-care-act/ [accessed 2020-02-13]
33. Sokol DK, Hettige S. Poor handwriting remains a significant problem in medicine. J R Soc Med 2006 Dec;99(12):645-646 [FREE Full text] [doi: 10.1258/jrsm.99.12.645] [Medline: 17139073]
34. Leonidas LL. Opinion - Inquirer.net. 2014. Death by Bad Handwriting URL: https://opinion.inquirer.net/79623/death-by-bad-handwriting [accessed 2020-02-13]
35. Charatan F. Family compensated for death after illegible prescription. Br Med J 1999 Dec 4;319(7223):1456 [FREE Full text] [doi: 10.1136/bmj.319.7223.1456] [Medline: 10582922]
36. Davis J. Healthcare IT News. 2017. eClinicalWorks Sued for Nearly $1 Billion for Inaccurate Medical Records URL: https://www.healthcareitnews.com/news/eclinicalworks-sued-nearly-1-billion-inaccurate-medical-records [accessed 2020-02-13]
37. Amazon Web Services. 2018. United States District Court Northern District of Illinois URL: https://s3.amazonaws.com/assets.fiercemarkets.net/public/004-Healthcare/external_Q12018/SurfsidevAllscripts.pdf [accessed 2020-02-13]
38. The Guardian. 2017. Bitcoin Mining Consumes More Electricity a Year Than Ireland URL: https://www.theguardian.com/technology/2017/nov/27/bitcoin-mining-consumes-electricity-ireland [accessed 2020-02-13]
39. Scherer M. UMEA University. 2017. Performance and Scalability of Blockchain Networks and Smart Contracts URL: https://umu.diva-portal.org/smash/get/diva2:1111497/FULLTEXT01.pdf [accessed 2020-03-24]
40. Miller C. The electronic medical record: a definition and discussion. Top Health Inf Manage 1993 Feb;13(3):20-29. [Medline: 10124869]
41. Cameron S, Turtle-Song I. Learning to write case notes using the SOAP format. J Couns Dev 2002;80(3):286-292. [doi: 10.1002/j.1556-6678.2002.tb00193.x]
42. Markle: Advancing America's Future. 2003. Personal Health Working Group Final Report URL: https://www.markle.org/publications/1429-personal-health-working-group-final-report [accessed 2020-02-13]
43. Handler R, Holtmeier R, Metzger J, Overhage M, Taylor S, Underwood C. Healthcare Information and Management Systems Society. 2003. HIMSS Electronic Health Record Definitional Model URL: http://www.providersedge.com/ehdocs/ehr_articles/HIMSS_EMR_Definition_Model_v1-0.pdf [accessed 2020-02-13]
44. AHIMA e-HIM Personal Health Record Work Group. Defining the Personal Health Record. J AHIMA 2005 Jun 11;156(24):786-786. [doi: 10.1136/vr.156.24.786-a]
45. Böcking W, Trojanus D. Health Data Management. In: Kirch W, editor. Encyclopedia of Public Health. New York, USA: Springer; 2008.
46. Kohane IS, Greenspun P, Fackler J, Cimino C, Szolovits P. Building national electronic medical record systems via the World Wide Web. J Am Med Inform Assoc 1996;3(3):191-207 [FREE Full text] [doi: 10.1136/jamia.1996.96310633] [Medline: 8723610]
47. Rind DM, Kohane IS, Szolovits P, Safran C, Chueh HC, Barnett GO. Maintaining the confidentiality of medical records shared over the internet and the world wide web. Ann Intern Med 1997 Jul 15;127(2):138-141. [doi: 10.7326/0003-4819-127-2-199707150-00008] [Medline: 9230004]

XSL•FO
RenderX

48. Schoenberg R, Safran C. Internet based repository of medical records that retains patient confidentiality. Br Med J 2000 Nov 11;321(7270):1199-1203 [FREE Full text] [doi: 10.1136/bmj.321.7270.1199] [Medline: 11073513]
49. Uckert F, Görz M, Ataian M, Prokosch HU. Akteonline-an electronic healthcare record as a medium for information and communication. Stud Health Technol Inform 2002;90:293-297. [Medline: 15460705]
50. Grant RW, Wald JS, Poon EG, Schnipper JL, Gandhi TK, Volk LA, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for diabetes collaborative care. Diabetes Technol Ther 2006 Oct;8(5):576-586 [FREE Full text] [doi: 10.1089/dia.2006.8.576] [Medline: 17037972]
51. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
52. Marceglia S, Bonacina S, Braidotti A, Nardelli M, Pinciroli F. Towards a web-based system for family health record. AMIA Annu Symp Proc 2006:1023 [FREE Full text] [Medline: 17238642]
53. Laplante PA, Kassab M, Laplante NL, Voas JM. Building caring healthcare systems in the internet of things. IEEE Syst J 2018;12(3):- [FREE Full text] [doi: 10.1109/JSYST.2017.2662602] [Medline: 31080541]
54. Wu F, Wu T, Yuce M. An internet-of-things (IoT) network system for connected safety and health monitoring applications. Sensors (Basel) 2018 Dec 21;19(1):E21 [FREE Full text] [doi: 10.3390/s19010021] [Medline: 30577646]
55. Meinert E, van Velthoven M, Brindley D, Alturkistani A, Foley K, Rees S, et al. The internet of things in health care in Oxford: protocol for proof-of-concept projects. JMIR Res Protoc 2018 Dec 4;7(12):e12077 [FREE Full text] [doi: 10.2196/12077] [Medline: 30514695]
56. Mavrogiorgou A, Kiourtis A, Perakis K, Pitsios S, Kyriazis D. IoT in healthcare: achieving interoperability of high-quality data acquired by IoT medical devices. Sensors (Basel) 2019 Apr 27;19(9):1-24 [FREE Full text] [doi: 10.3390/s19091978] [Medline: 31035612]
57. Valluru D, Jeya IJ. IoT with cloud based lung cancer diagnosis model using optimal support vector machine. Health Care Manag Sci 2019 Jul 20:- epub ahead of print. [doi: 10.1007/s10729-019-09489-x] [Medline: 31327114]
58. Ramirez Lopez LJ, Puerta Aponte G, Rodriguez Garcia A. Internet of things applied in healthcare based on open hardware with low-energy consumption. Healthc Inform Res 2019 Jul;25(3):230-235 [FREE Full text] [doi: 10.4258/hir.2019.25.3.230] [Medline: 31406615]
59. Qu Y, Ming X, Qiu S, Zheng M, Hou Z. An integrative framework for online prognostic and health management using internet of things and convolutional neural network. Sensors (Basel) 2019 May 21;19(10):1 [FREE Full text] [doi: 10.3390/s19102338] [Medline: 31117213]
60. Rau H, Wu Y, Chu C, Wang F, Hsu M, Chang C, et al. Importance-performance analysis of personal health records in Taiwan: a web-based survey. J Med Internet Res 2017 Apr 27;19(4):e131 [FREE Full text] [doi: 10.2196/jmir.7065] [Medline: 28450273]
61. Bahga A, Madisetti VK. A cloud-based approach for interoperable electronic health records (EHRs). IEEE J Biomed Health Inform 2013 Sep;17(5):894-906. [doi: 10.1109/JBHI.2013.2257818] [Medline: 25055368]
62. Fernández-Cardeñosa G, de la Torre-Díez I, López-Coronado M, Rodrigues JJ. Analysis of cloud-based solutions on EHRs systems in different scenarios. J Med Syst 2012 Dec;36(6):3777-3782. [doi: 10.1007/s10916-012-9850-2] [Medline: 22492177]
63. Zangara G, Corso PP, Cangemi F, Millonzi F, Collova F, Scarlatella A. A cloud based architecture to support electronic health record. Stud Health Technol Inform 2014;207:380-389. [doi: 10.3233/978-1-61499-474-9-380] [Medline: 25488244]
64. Schulz S, Stegwee R, Chronaki C. Standards in healthcare data. In: Kubben P, Dumontier M, Dekker A, editors. Standards in Healthcare Data. New York, USA: Springer; 2019:19-36.
65. MRC Cognition and Brain Sciences Unit. The Analyze Data Format URL: http://imaging.mrc-cbu.cam.ac.uk/imaging/FormatAnalyze [accessed 2020-02-13]
66. NIfTI: Neuroimaging Informatics Technology Initiative. URL: https://nifti.nimh.nih.gov/ [accessed 2020-02-13]
67. Vincent RD, Neelin P, Khalili-Mahani N, Janke AL, Fonov VS, Robbins SM, et al. MINC 2.0: a flexible format for multi-modal images. Front Neuroinform 2016;10:35 [FREE Full text] [doi: 10.3389/fninf.2016.00035] [Medline: 27563289]
68. Digital Imaging and Communications in Medicine. URL: https://www.dicomstandard.org/ [accessed 2020-02-13]
69. Wen H, Chang W, Hsu M, Ho C, Chu C. An assessment of the interoperability of electronic health record exchanges among hospitals and clinics in Taiwan. JMIR Med Inform 2019 Mar 28;7(1):e12630 [FREE Full text] [doi: 10.2196/12630] [Medline: 30920376]
70. Davis J. HealthITSecurity. 2019. 15 Million Patient Records Breached in 2018; Hacking, Phishing Surges URL: https://healthitsecurity.com/news/15-million-patient-records-breached-in-2018-hacking-phishing-surges [accessed 2020-02-13]
71. Business Wire. New Intel Security Cloud Report Reveals IT Departments Find It Hard to Keep the Cloud Safe URL: https://www.businesswire.com/news/home/20170212005011/en/ [accessed 2020-02-13]
72. Zorz Z. Help Net Security - Information Security News. 2019. Healthcare's Blind spot: Unmanaged IoT and Medical Devices URL: https://www.helpnetsecurity.com/2019/07/22/healthcare-iot/ [accessed 2020-02-13]
73. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. New York, USA: BN Publishing; 2008.

XSL•FO
RenderX

74. Ismail L, Heba H, AlShamsi M, AlHammadi M, AlDhanhani N. Towards a Blockchain Deployment at UAE University: Performance Evaluation and Blockchain Taxonomy. In: Proceedings of the 2019 International Conference on Blockchain Technology. 2019 Presented at: ICBCT'19; March 15-18, 2019; Hawaii, USA p. 30-38. [doi: 10.1145/3320154.3320156]
75. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using Blockchain for Medical Data Access and Permission Management. In: Proceedings of the 2nd International Conference on Open and Big Data. 2016 Presented at: OBD'16; August 22-24, 2016; Vienna, Austria. [doi: 10.1109/obd.2016.11]
76. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-based data preservation system for medical data. J Med Syst 2018 Jun 28;42(8):141. [doi: 10.1007/s10916-018-0997-3] [Medline: 29956058]
77. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustain Cities Soc 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
78. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: efficient and secure medical data sharing via blockchain. J Med Syst 2018 Jun 21;42(8):136. [doi: 10.1007/s10916-018-0993-7] [Medline: 29931655]
79. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
80. Dey T, Jaiswal S, SunderKrishnan S, Katre N. HealthSense: A Medical Use Case of Internet of Things and Blockchain. In: Proceedings of the International Conference on Intelligent Sustainable Systems. 2017 Presented at: ICISS'17; December 7-8, 2017; Palladam, India. [doi: 10.1109/iss1.2017.8389459]
81. Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Continuous patient monitoring with a patient centric agent: a block architecture. IEEE Access 2018;6:32700-32726. [doi: 10.1109/access.2018.2846779]
82. Evariant: Healthcare's Only Patient for Life Platform. What is Healthcare Data Management and Why is it Important? URL: https://www.evariant.com/faq/why-is-healthcare-data-management-important [accessed 2020-02-13]
83. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015 Aug 26;1(1):15030-15037 [FREE Full text] [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]
84. Yu K, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016 Aug 16;7:12474 [FREE Full text] [doi: 10.1038/ncomms12474] [Medline: 27527408]
85. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NN, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Sci Rep 2017 Apr 18;7:46450 [FREE Full text] [doi: 10.1038/srep46450] [Medline: 28418027]
86. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, et al. arXiv. 2017. Detecting Cancer Metastases on Gigapixel Pathology Images URL: https://arxiv.org/abs/1703.02442 [accessed 2020-03-24]
87. Richter AN, Khoshgoftaar TM. Efficient learning from big data for cancer risk modeling: a case study with melanoma. Comput Biol Med 2019 Jul;110:29-39. [doi: 10.1016/j.compbiomed.2019.04.039] [Medline: 31112896]
88. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-Learning Algorithms to Automate Morphological and FunctionaMachine-learning algorithms to automate morphological and functional assessments in 2D echocardiographyl Assessments in 2D Echocardiography. J Am Coll Cardiol 2016 Nov 29;68(21):2287-2295 [FREE Full text] [doi: 10.1016/j.jacc.2016.08.062] [Medline: 27884247]
89. Kiral-Kornek I, Roy S, Nurse E, Mashford B, Karoly P, Carroll T, et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. EBioMedicine 2018 Jan;27:103-111 [FREE Full text] [doi: 10.1016/j.ebiom.2017.11.032] [Medline: 29262989]
90. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. Digit Med 2018;1:18.
91. Hsiao P, Chin-Ming CH, Li Y. Applied wearable devices for digital health based on novel cardiac force index of running performance:cross-sectional study. JMIR mHealth and uHealth (forthcoming). [doi: 10.2196/15331]
92. Health Data Exploration-Personal Data for the Public. 2014. Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health URL: http://hdexplore.calit2.net/wp-content/uploads/2015/08/hdx_final_report_small.pdf [accessed 2020-02-13]
93. Knight Lab. Personal Health Data as Public Good URL: https://knightlab.ucsd.edu/wordpress/?page_id=19 [accessed 2020-02-13]
94. Bakar Institute. Bringing the Power of Computation to Today’s Spectrum of Data Will Yield Untold Health Insights and Patterns URL: https://bakarinstitute.ucsf.edu/research/ [accessed 2020-02-13]
95. Digiconomist. Bitcoin Energy Consumption Index URL: https://digiconomist.net/bitcoin-energy-consumption [accessed 2020-02-13]
96. Crush Crypto. What is Practical Byzantine Fault Tolerance (PBFT)? URL: https://crushcrypto.com/what-is-practical-byzantine-fault-tolerance/ [accessed 2020-02-13]

XSL•FO
RenderX

97. Milutinovic M, He W, Wu H, Kanwal M. Proof of Luck: An Efficient Blockchain Consensus Protocol. In: Proceedings of the 1st Workshop on System Software for Trusted Execution. 2016 Presented at: SysTEX'16; December 12-16, 2016; Trento, Italy. [doi: 10.1145/3007788.3007790]
98. Ismail L, Materwala H, Zeadally S. Lightweight blockchain for healthcare. IEEE Access 2019;7:149935-149951. [doi: 10.1109/access.2019.2947613]
99. Dorri A, Kanhere SS, Jurdak R. Towards an Optimized BlockChain for IoT. In: Proceedings of the Second International Conference on Internet-of-Things Design and Implementation. 2017 Presented at: IoTDI'17; April 18-21, 2017; Pittsburgh, Pennsylvania. [doi: 10.1145/3054977.3055003]

Abbreviations

EHR: electronic health record
eHealth: electronic health
IoT: Internet of Things
mHealth: mobile health

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by Z Sherali, A Behmanesh, CM Chu; comments to author 03.02.20; revised version received 13.02.20; accepted 01.03.20; published 07.07.20.

Please cite as:
Ismail L, Materwala H, Karduck AP, Adem A
Requirements of Health Data Management Systems for Biomedical Care and Research: Scoping Review
J Med Internet Res 2020;22(7):e17508
URL: https://www.jmir.org/2020/7/e17508
doi:10.2196/17508
PMID:32348265

©Leila Ismail, Huned Materwala, Achim P Karduck, Abdu Adem. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Influence of Electronic Health Record Use on Physician Burnout: Cross-Sectional Survey

Tania Tajirian[1,2], MD; Vicky Stergiopoulos[1,2], MD, MHSc; Gillian Strudwick[1,2], RN, PhD; Lydia Sequeira[1,2], MHI; Marcos Sanches[3], MSc; Jessica Kemp[4], BSc; Karishini Ramamoorthi[5], BSc; Timothy Zhang[4]; Damian Jankowicz[1], PhD

[1]Centre for Addiction and Mental Health, Toronto, ON, Canada

[2]University of Toronto, Toronto, ON, Canada

[3]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

[4]University of Waterloo, Waterloo, ON, Canada

[5]McMaster University, Hamilton, ON, Canada

Corresponding Author:
Tania Tajirian, MD
Centre for Addiction and Mental Health
6168F, 100 Stokes St.
Toronto, ON, M6J1H4
Canada
Phone: 1 4165358501 ext 30515
Email: Tania.Tajirian@camh.ca

Abstract

Background: Physician burnout has a direct impact on the delivery of high-quality health care, with health information technology tools such as electronic health records (EHRs) adding to the burden of practice inefficiencies.

Objective: The aim of this study was to determine the extent of burnout among physicians and learners (residents and fellows); identify significant EHR-related contributors of physician burnout; and explore the differences between physicians and learners with regard to EHR-related factors such as time spent in EHR, documentation styles, proficiency, training, and perceived usefulness. In addition, the study aimed to address gaps in the EHR-related burnout research methodologies by determining physicians' patterns of EHR use through usage logs.

Methods: This study used a cross-sectional survey methodology and a review of administrative data for back-end log measures of survey respondents' EHR use, which was conducted at a large Canadian academic mental health hospital. Chi-square and Fisher exact tests were used to examine the association of EHR-related factors with general physician burnout. The survey was sent out to 474 individuals between May and June 2019, including physicians (n=407), residents (n=53), and fellows (n=14), and we measured physician burnout and perceptions of EHR stressors (along with demographic and practice characteristics).

Results: Our survey included 208 respondents, including physicians (n=176) and learners (n=32). The response rate was 43.2% for physicians (full-time: 156/208, 75.0%; part-time: 20/199, 10.1%), and 48% (32/67) for learners. A total of 25.6% (45/176) of practicing physicians and 19% (6/32) of learners reported having one or more symptoms of burnout, and 74.5% (155/208) of all respondents who reported burnout symptoms identified the EHR as a contributor. Lower satisfaction and higher frustration with the EHRs were significantly associated with perceptions of EHR contributing toward burnout. Physicians' and learners' experiences with the EHR, gathered through open-ended survey responses, identified challenges around the intuitiveness and usability of the technology as well as workflow issues. Metrics gathered from back-end usage logs demonstrated a 13.6-min overestimation in time spent on EHRs per patient and a 5.63-hour overestimation of after-hours EHR time, when compared with self-reported survey data.

Conclusions: This study suggests that the use of EHRs is a perceived contributor to physician burnout. There should be a focus on combating physician burnout by reducing the unnecessary administrative burdens of EHRs through efficient implementation of systems and effective postimplementation strategies.

(J Med Internet Res 2020;22(7):e19274) doi:10.2196/19274

XSL•FO
RenderX

KEYWORDS

electronic health record; physician; burnout; psychiatry; medical informatics

Introduction

Overview of Physician Burnout

Physician wellness is vital to the delivery of high-quality health care and greatly affects the performance of health care systems [1]. Organizations have started including wellness programs among their top priorities in an attempt to reduce physician burnout [2]—a work-related syndrome involving emotional exhaustion, depersonalization, and a sense of reduced personal accomplishment [3]. Physician burnout has been associated with career dissatisfaction [4], absenteeism and job turnover [5], reduced quality of care [6], and medical errors [7]. It is increasingly being measured, with high prevalence rates (78%) among American physicians [8], and almost 30% of Canadian physicians [9] have reported symptoms of burnout. The major contributors to physician burnout include *individual factors* (eg, age and education debt) [10] and *work factors* (eg, inefficient work processes, negative leadership, and limited interprofessional collaboration or advancement) [11]. Within psychiatry, workplace variables have been found to be a major stressor and may be more likely to perpetuate burnout [12]. Demands including navigating the working relationship with clients experiencing trauma—while often becoming the target of anger, hatred, and even violence—can be emotionally draining [13]. Moreover, burnout has also been known to span all phases of a physician's career, including during medical school and residency, and a recent systematic review calculated a 33.7% burnout rate for psychiatry residents [14].

Among workplace inefficiencies, the use of health information technology such as electronic health records (EHRs) has been suggested to contribute to physician burnout in psychiatry [15,16]. With proper implementation, EHRs can improve the quality of health care by increasing time efficiency and guideline adherence and reducing medication errors [17]. However, the promise of improved quality of patient care through fast access to patient information and improved clinical decision-making support has not been attained in many health care organizations as the unintended consequences of EHRs proliferate [18].

These unintended consequences can be *technical* factors, such as poor software design, or *sociotechnical* factors, such as poor usability or workflow integration [19-22]. EHRs have often added to physicians' cognitive load through excessive data entry requirements [23,24] and reduction of time spent with patients [25,26]. The provision of mental health care, in particular, poses an added level of complexity, including navigating multidisciplinary treatment plans, varying levels of care (ie, residential and partial hospitals) that do not fit neatly into the clinical or scheduling workflows [27], and the inability to capture and find important documentation [28,29]. Despite the challenges, the adoption of EHRs remains an important policy priority in most countries, showing a steady increase over the past 10 years and reaching an adoption rate of 75% in American [30] hospitals and 81% in Canadian [31] hospitals. Moving forward, the National Academy of Medicine calls for a *human-centered approach* to combat physician burnout by reducing unnecessary administrative burdens through improved design and implementation of technology and supportive regulatory policies [32].

Although there are numerous editorial and opinion pieces identifying EHRs as possible contributors to physician burnout [33,34], research in determining physicians' perceptions of the impact of EHRs on burnout is scant [35]. In addition, there is a need to apply a variety of research methods to fully understand the complexity of the phenomenon to optimize technology and clinical workflows [35], given that previous studies used subjective, perception-based data for measuring both EHR-related stressors and burnout variables.

Study Aims

This study aimed to (1) identify the extent of burnout and the perceived contribution of the EHR toward burnout among our population of physicians and learners (residents and fellows); (2) identify significant contributors of burnout and EHR-related burnout; (3) explore differences between physicians and learners among factors previously identified as contributing to EHR-related burnout (time spent, use, and documentation styles within the EHR; EHR proficiency and training; and perceived usefulness of EHR); and (4) compare self-reported perceptions on EHR usage metrics using log data.

Methods

Design, Setting, and Participants

This study used a cross-sectional survey design to gather the perceptions of physicians and learners and back-end EHR usage logs to gather use patterns. It was conducted at Canada's largest academic mental health hospital located in Toronto, Ontario. At the time of the survey distribution, the Centre for Addiction and Mental Health had 407 physicians, 208 of which were considered full-time (ie, ones who had a *primary* appointment with the hospital). Among these physicians were hospitalists and psychiatrists spread across 7 different clinical divisions. In addition, the study included 53 residents and 14 clinical fellows. Of the 37,065 unique patients admitted in the hospital, 63.10% (23,388/37,065) were treated for an admission diagnosis falling within two groups: schizophrenia or/psychotic disorders and substance use disorders [36]. Physicians used a comprehensive EHR, which was implemented 5 years before this study.

Depending on the email address used for the electronic survey links, usage log data for 201 participants in May and 198 participants in June were identified, which were then used to compare metrics that were also asked within the survey.

Ethical approval was obtained from the organization's quality improvement projects ethics review board.

XSL•FO
RenderX

Data Collection

Survey

The survey was administered to the study's target population of physicians and learners (residents and clinical fellows) between May 2019 and June 2019 by sending a link to the anonymous electronic survey via email. To maximize survey participation, weekly reminders were sent out for the duration of survey recruitment (6 weeks), and engagement methods, including advertising in physician newsletters, discussions at hospital-wide meetings, and resident lunches, were used.

Research Electronic Data Capture (REDCap)—a secure web app—was used to manage the survey [37]. The app collected information about demographics, practice setting and EHR-related usage, and burnout (Table 1) and open-ended responses to capture the respondents' experience with the EHR and explore unique ways in which they use the EHR (Multimedia Appendix 1). The survey was developed using previous literature, including the study by Gardner et al [15] and KLAS Arch Collaborative Impact Report on Clinician Burnout [38], and was tested with physicians, researchers, and divisional chiefs, and their feedback was incorporated.

Table 1. Independent and dependent variables.

Variable type and category	Variables
Independent	
Demographics	• Age • Gender (male, female, or nonbinary) • Role (physician versus learners)
Practice	• Clinical academic division • Length of practice • Patient load (number of patients per week)
EHR[a]-related factors	• Time spent in EHR (time per patient and time after hours per week) • Frustration and satisfaction with EHR • Documentation styles (typing, back-end transcription, and voice recognition software) • EHR proficiency • EHR training needs • Perceived usefulness of EHR (on improving communication, enabling high-quality care, and patient safety)
Dependent	
Burnout	• General burnout: Measured using a single question from the Mini-Z survey [39] (The Mini-Z is a 10-item instrument developed from the Physician Worklife Study [40]). This single question has been previously validated for physicians [41] against the detailed Maslach Burnout Inventory [3], and it achieved a sensitivity of 83.2% and specificity of 87.4% [41]. Respondents were asked to identify their symptoms of burnout based on a 5-point scale: (1) "I enjoy my work. I have no symptoms of burnout," (2) "I am under stress, and don't always have as much energy as I did, but I don't feel burned out," (3) "I am definitely burning out and have one or more symptoms of burnout, eg. emotional exhaustion," (4) "The symptoms of burnout I am experiencing won't go away. I think about work frustrations a lot," and (5) "I feel completely burned out. I am at the point where I may need to seek help." Participants considered as "burned out" include those having one or more symptoms of burnout (ie, a score of ≥3 on the above scale) • Contribution of EHRs toward burnout: Measured by a single question where physicians and learners were asked "Do you think [EHR name] contributes to your symptoms of burnout?", and responses were captured on a 4-point scale: "Always," "Almost always", "Some of the time," and "Almost never."

[a]EHR: electronic health record.

Usage Logs

As unique electronic survey links were used, a list of those who responded to the survey was assembled. Usage logs were extracted for all these individuals for the 2-month period (May-June 2019) of the survey administration. Although we were able to determine whether a participant had responded to the survey or not (a feature of REDCap), we were not able to identify their individual survey responses because of the anonymity of the survey. Therefore, variables extracted from usage log data were compared with responses from the survey in aggregate and not at an individual level. The data extracted were *EHR-related factors* including (1) number of patients seen per month, (2) time spent in EHR per patient, and (3) time spent in EHR after hours per month (details of the source of the back-end EHR analytics can be found in the study by Overhage et al [42]).

The inactive time when the physician was logged into the EHR but not actively engaged in using it (eg, typing) was excluded from the analyses, and *after-hours* were defined as 6 PM to 6 AM and weekends—similar to how it was defined in the survey.

Data Analysis

Survey

Descriptive statistics were calculated for all numeric and categorical variables. The association of independent variables (including demographics, practice styles, and EHR factors) with variables measuring burnout was analyzed using Fisher exact tests and chi-square tests (when the content table was too large

XSL•FO
RenderX

for the exact Fisher test to be calculated). Fisher exact tests were used to identify differences between physicians and learners for the following variables (counts of 5 or lower in chi-square tables): age, patient load, time spent in EHRs per patient, frustration with EHR, satisfaction with EHR, perceived usefulness (on improving communication within the circle of care, enabling delivery of high-quality care, and keeping patients safe), and documentation type. All descriptive and chi-square analyses were conducted using the Statistical Package for the Social Sciences software [43], and Fisher exact tests were conducted in R [44]. When significance at. 05 level was not achieved, no *P* values were reported within the study's results.

For open-ended survey responses, inductive content analysis was used [45]. After dividing the responses into two groups (low and high satisfaction with EHR based on a quantitative rating scale), the data were read and coded to capture key thoughts or concepts. Investigator triangulation was used to refine the coding scheme, and one investigator proceeded to code all data. Emergent codes were clustered into broader subcategories or themes.

Comparison: Self-Reports and Usage Log Data

As back-end EHR usage log data were extracted on a monthly basis, the 2 survey variables (patients seen per week and time spent in EHR after hours per week) were changed from *weekly* to *monthly* to ensure practical comparison. Descriptive statistics gathered from the survey responses and usage logs of all respondents were compared.

Results

Participant Profile

Demographic and practice characteristics of the study population are shown in Table 2. The survey was answered by 176 physicians and 32 learners. Response rates were 43.2% for physicians (full-time: 75% and part-time: 10%) and 47.7% for learners (fellows: 86% and residents: 40%). A total of 44.3% (78/176) of physicians and 50% (16/32) of learners were female; 46.0% (81/176) of physicians were in the 0 to 10 years practice timeframe and 26.1% (46/176) practiced for 21 years or more. Physicians saw a mean of 27 patients (median 25) per week and learners saw a mean of 14 patients (median 15).

XSL•FO
RenderX

Table 2. Demographic and practice characteristics, by experience level.

Demographics	Total sample (N=208), n (%)	Physicians (n=176), n (%)	Learners (residents and fellows; n=32), n (%)
Age (years)			
<30	17 (8.2)	4 (2.3)	13 (41)
31-40	81 (38.9)	63 (35.8)	18 (56)
41-50	59 (28.4)	58 (33.0)	1 (3)
51-60	23 (11.1)	23 (13.1)	0 (0)
≥61	28 (13.5)	28 (15.9)	0 (0)
Gender			
Female	94 (45.2)	78 (44.3)	16 (50)
Male	105 (50.5)	89 (50.6)	16 (50)
Gender fluid or nonbinary or two-spirit	1 (0.5)	1 (0.6)	0 (0)
I prefer not to answer	8 (3.8)	8 (4.5)	0 (0)
Practice setting[a]			
Emergency mental health	40 (19.2)	22 (12.5)	18 (56)
Inpatient mental health	82 (39.4)	69 (39.2)	13 (41)
Outpatient mental health	159 (76.4)	131 (74.4)	28 (88)
Telehealth	25 (12.0)	21 (11.9)	4 (13)
Outreach	8 (3.8)	8 (4.5)	0 (0)
Unknown	2 (1.0)	2 (1.1)	0 (0)
Patient load/week			
≤10	53 (25.5)	39 (22.2)	14 (43)
11-20	55 (26.4)	39 (22.2)	16 (50)
21-30	46 (22.1)	45 (25.6)	1 (3)
≥31	53 (25.5)	52 (29.5)	1 (3)
Unknown	1 (0.5)	1 (0.6)	0 (0)

[a]Practice setting was a multi-select question.

Burnout and the Perceived Contribution of the Electronic Health Record Toward Burnout

A total of 25.6% (45/176) of all physicians and 19% (6/32) of learners identified as having one or more symptoms of burnout, as measured using the single-item measure from the Mini-Z. When asked about EHR contributing to burnout, 69.3% (122/176) of physicians and 67% (22/32) of learners reported feeling that the EHR always or almost always contributes to their symptoms of burnout. Within the subset of those individuals who experienced one or more symptoms of burnout (n=51), this perception of the EHR contributing to burnout was slightly more prevalent (155/208, 74.5%; Table 3).

XSL•FO
RenderX

Table 3. Burnout prevalence, by experience level.

Burnout measure	Total sample (N=208), n (%)	Physicians (n=176), n (%)	Learner (residents and fellows; n=32), n (%)
General physician burnout			
1: "I enjoy my work. I have no symptoms of burnout"	45 (21.6)	38 (21.6)	7 (22)
2: "I am under stress, and don't always have as much energy as I did, but I don't feel burned out"	111 (53.4)	92 (52.3)	19 (59)
3: "I am definitely burning out and have one or more symptoms of burnout, e.g. emotional exhaustion"	35 (16.8)	31 (17.6)	4 (13)
4: "The symptoms of burnout I am experiencing won't go away. I think about work frustrations a lot"	15 (7.2)	13 (7.4)	2 (6)
5: "I feel completely burned out. I am at the point where I may need to seek help."	1 (0.5)	1 (0.6)	0 (0)
Unknown	1 (0.5)	1 (0.6)	0 (0)
One or more symptoms of burnout[a]	51 (24.5)	45 (25.6)	6 (19)
EHR[b] contributing to physician burnout			
Always/almost always	144 (69.2)	122 (69.3)	22 (67)
Some of the time/almost never	63 (30.3)	53 (30.1)	10 (31)
Unknown	1 (0.48)	1 (0.57)	0 (0)

[a]General physician burnout score ≥3.

[b]EHR: electronic health record.

Significant Contributions to Burnout

Physicians and learners with higher levels of frustration and lower satisfaction with the EHR were significantly more burned out (Figure 1). Other variables that were significantly associated with burnout included participants' perceptions of the EHR on keeping their patients safe (*P*=.002) and whether physicians perceived communication regarding EHR upgrades as efficient (*P*=.047).

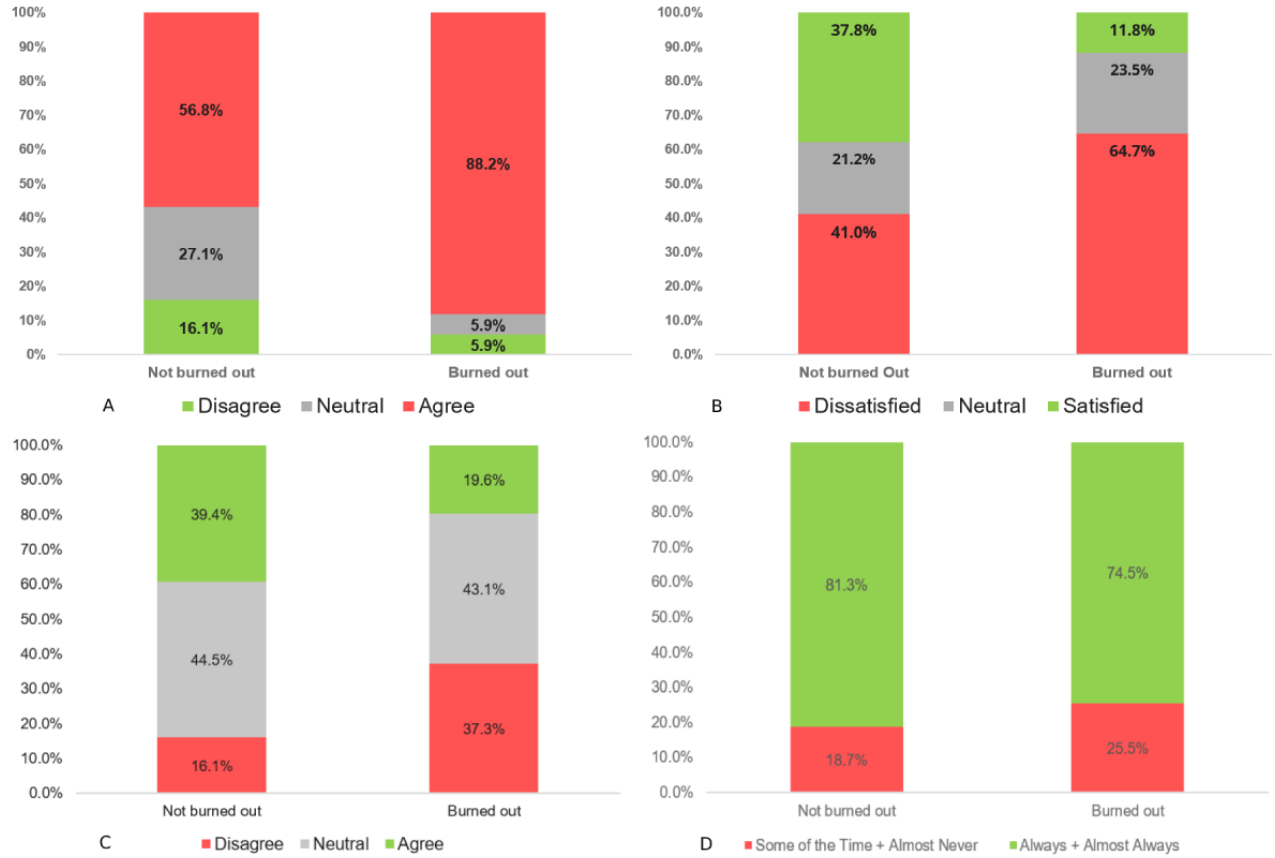
Variables that had a significant association with perceptions of EHR contributing to burnout were low satisfaction with the EHR (*P*<.001) and frustration with the EHR (*P*<.001; Figure 1). Similarly, as with burnout, those who were content with communication surrounding EHR updates were also significantly less likely to perceive the EHR as a contributory factor to burnout (*P*=.003). Finally, those who were more proficient with the EHR were also significantly less likely to perceive it as contributing to their burnout (*P*=.01).

The experiences of physicians and learners who had *low satisfaction* with the EHR focused on usability issues and unintended consequences of the EHR on patient care. A total of 39 individuals reported usability issues such as it being nonintuitive, having too many clicks, or not being user-friendly, with one participant referring to it as "death by a thousand clicks" (Participant #94). Moreover, 48 respondents discussed difficulties with finding or retrieving information, including inaccessible documentation. Respondents also discussed time sinks because of the system being "slow" and "clunky" and noted the impact that technology has on direct patient care, such as "...has a negative impact on...the amount of quality face-to-face time I can spend with patients" (Participant #118).

Respondents with *high EHR satisfaction* used workarounds to complete tasks in the EHR, such as "type long consult notes in [Microsoft] word then copy" (Participant #33) or "enter my appointments in my Outlook calendar..." (Participant #66). Others discussed their knowledge of customization, such as "know how to insert personal short cuts" (Participant #103), or the use of back-end dictation or "use [voice recognition software] exclusively instead of typing progress notes" (Participant #125). Satisfied participants thought that the EHR allows for "communication with other care providers" and has "the ability to forward things easily to care providers in the circle of care" (Participant #168).

XSL•FO
RenderX

Figure 1. Significant contributors to physician burnout. A: [EHR name] adds to my daily frustration (*P*<.001); B: How would you rate your satisfaction with [EHR name]? (*P*<.001); C. [EHR name] helps keep my patients safe (*P*<.002); D: Do you feel communication regarding [EHR name] changes are efficient? (*P*<.047).



Physician and Learner Differences for Electronic Health Record–Related Factors

Time Spent, Use, and Documentation Styles Within the Electronic Health Record

Time spent, use, and documentation styles were not significantly different between those participants who were burned out and those who were not burned out. There were significant differences between physicians and learners with respect to the time spent on the EHR per patient, where 47% (15/32) of learners spent over 50 min per patient compared with only 16.5% (29/176) of physicians spending the same amount of time (*P*=.03). In total, 84.1% (148/176) of physicians and 78% (25/32) of learners reported spending time on the EHR after hours, with 22.2% (2/208) of respondents spending 4 or more hours per week.

Compared with physicians (96/176, 54.5%), a greater proportion of learners (24/32, 75%) document in the EHR through direct typing only. In contrast, 2.8% (5/176) of physicians rely solely on dictation through back-end transcription or voice recognition. None of the learners used dictation only to document in the EHR. Leaners that did not use direct typing as their documentation style (8/32, 25%) used a combination of direct typing and dictation, whereas 41.5% (73/176) of physicians used a combination of documentation styles.

Electronic Health Record Proficiency and Training

A total of 31.3% (55/176) of physicians and 41% (13/32) of learners felt that their initial EHR training prepared them well, and a large portion of physicians and learners (89/176, 50.6% and 13/32, 41%, respectively) felt they had ongoing training available to meet their needs. In addition, more than half of the physicians (94/176, 53.4%) and learners (22/32, 69%) felt proficient in their use of the EHR. The majority of physicians (14/176, 79.5%) and learners (24/32, 75) reported that communication regarding changes to the EHR was effective.

Perceived Usefulness of Electronic Health Records

A total of 62.5% (110/176) of physicians and 72% (23/32) of learners indicated that the EHR adds to their daily frustration (Table 4). Although 51.1% (90/176) of physicians felt that the EHR improved communication in their circle of care, only 34% (11/32) of learners agreed with this statement. With regard to the impact of the EHR on improving patient safety, 38.1% (67/176) and 40.3% (71/176) of physicians agreed with the statement or were neutral, respectively. In contrast, 63% (20/32) of learners felt neutral about this statement and 13% (4/32) agreed. Physicians had a significantly more positive perspective on the EHR in terms of quality of care (*P*=.007), with 38.6% (68/176) agreeing that it enabled them to deliver high-quality care, compared with only 9% (3/32) of learners.

XSL•FO
RenderX

Table 4. Electronic health record factors, by experience level.

Demographics	Total sample (N=208), n%	Physicians (n=176), n (%)	Learner (residents and fellows; n=32), n (%)
Satisfaction with EHR[a]			
Very satisfied/somewhat satisfied	97 (46.6)	1 (43.2)	21 (66)
Neither satisfied nor dissatisfied	45 (21.6)	37 (21.0)	8 (25)
Somewhat dissatisfied/very dissatisfied	65 (31.3)	62 (35.2)	3 (9)
Unknown	1 (0.5)	1 (0.6)	0 (0)
Frustration with EHR			
Disagree	26 (12.5)	25 (14.2)	3 (9)
Neutral	45 (21.6)	39 (22.2)	6 (19)
Agree	133 (63.9)	110 (62.5)	23 (72)
Unknown	2 (1.0)	2 (1.1)	0 (0)
Time spent in EHR per patient (min)			
≤10	32 (15.4)	31 (17.6)	1 (3)
11-20	56 (26.9)	51 (29.0)	5 (16)
21-50	74 (35.6)	63 (35.8)	11 (34)
≥50	44 (21.2)	29 (16.5)	15 (47)
Unknown	2 (1.0)	2 (1.1)	0 (0)
EHR improves quality of care			
Disagree	46 (22.1)	34 (19.3)	12 (38)
Neutral	89 (42.8)	72 (40.9)	17 (53)
Agree	71 (34.1)	68 (38.6)	3 (9)
Unknown	2 (0.9)	2 (1.1)	0 (0)

[a]EHR: electronic health record.

Self-Reported Perceptions and Electronic Health Record Usage Log Data Comparison

As gathered from usage logs, the median *number of patients seen per month* for all the survey respondents over the months of May and June 2019 was 60 patients (May, N=201, June, N=198), compared with the self-reported median of 80 patients (N=207). The median *time spent on EHRs per patient* for all survey respondents for the months of May and June 2019 was 16.4 min (May, N=201, June, N=198), compared with the self-reported median of 30 min (N=206). The median *time spent on the EHR after hours* (defined as 6 PM to 6 AM and weekends) for all survey respondents for the months of May and June 2019 was 2.37 hours (May, N=201, June, N=198), compared with the self-reported median of 8 hours (N=201).

Discussion

Principal Findings

Although overall satisfaction with EHRs remains low, reverting to paper documentation is not a viable alternative. This study adds to a growing body of evidence calling for a focus on EHR improvement [46]. The study data demonstrate that 67% of learners and 43% of physicians were satisfied with the system, which is comparable with other studies [47]. Among those who perceived the EHR in a negative light, a majority (>65%) of the respondents expressed dissatisfaction with the EHR (n=97).

Burnout and the Perceived Contribution of the Electronic Health Record Toward Burnout

This study helps in understanding physician burnout attributed to technology within the Canadian mental health context. Although the general burnout rate of physicians and learners (24.6%) was comparable with the Canadian national average (30%) [9], our survey found that the majority (69.6%) of physicians and learners attributed EHR to their symptoms of burnout, even when they did not identify as being burned out. Although other institutions have completed surveys to examine the role of technology in physician burnout [15,16,48], this study adds to the existing literature, providing data from a different geography and a robust baseline at our facility. Measuring burnout rates, as well as the significant EHR-related contributors to burnout within the hospital, have direct implications on practice. Our organization has created a multipronged approach toward improving physicians' experience with the EHR, which includes direct feedback channels, improved education and communication around EHR updates, implementing speech recognition technology, and developing physician efficiency dashboards. Having a strong baseline measure of burnout allows us to measure the short- and

XSL•FO
RenderX

long-term impact of initiatives at our hospital which aim to reduce physician burnout.

Significant Contributors to Burnout

Gardner et al [15] found that those who spent excessive time on the EHR at home had a 1.9 times higher rate of burnout, and Privitera et al [48] found that EHR use at home increased burnout by 46% within their population. However, this study found no significant differences in the time spent after hours between those respondents who were and were not burned out.

Another previously affiliated factor with EHR satisfaction was gender, with men reporting significantly higher EHR workload stress than women [49]. This study did not find any significant differences between men, women, or nonbinary individuals when it came to satisfaction and frustration with EHR, as well as other EHR-related factors.

The results from this study did identify low proficiency with the EHR as a significant factor that leads physicians and learners to perceive the EHR as contributing toward their burnout, which supports the hypothesis that improved education and training can help in reducing this negative perception. Research by Dastagir et al [50] demonstrated the impact of proficiency training on significant improvements in self-reported efficiency and satisfaction, which could eventually have an effect on burnout.

Physician and Learner Differences for Electronic Health Record–Related Factors

In addition, this study found significant differences between physicians and learners with respect to the time spent in the EHR per patient, with a higher number of learners spending >50 min. Such a difference could be attributed to learners getting used to a new EHR system (as they often work with several EHRs across the various training sites) and mastering clinical practice and documentation standards. They could also be tasked with doing more designated EHR work, including documentation and orders, allocated by their supervising physician. Similar results demonstrating the extent of indirect patient care that residents take on were found in a time-motion study conducted by Penn Medicine and John Hopkins University. The study found that the residents spend almost 66% of their time interacting with patients' medical records or documentation [51].

Self-Reported Perceptions and Electronic Health Record Usage Log Data Comparison

This study found that perceptions of time spent in the EHR after hours were much higher than the actual time spent, as gathered by back-end usage logs, with an >5-hour difference between these 2 averages. It is possible that the time spent after hours for logging in and out of the system, on email, and for other digital administrative activities could be included within participants' perceived estimates. This difference resembles previous research that has found overestimations of 1.83 hours in learners and up to 4.04 hours in attending physicians [52]. This study's respondents also overestimated the time spent in the EHR per patient compared with how long they spend according to back-end usage logs (14-min difference), which could be because of interruptions in the workflow. It is important to note that, in general, employees have been shown to overestimate the hours that they work [53].

This discrepancy between self-reported and back-end usage log data has implications for future research, where a combination of methods should be used for studying the link between EHR-related stressors and physician burnout. Although burnout is primarily measured through perceptions, the stressors related to the EHR, such as time spent in the EHR after hours, primary documentation method, and amount and frequency of training, can all be measured through more objective means.

Limitations

Due to the nature of this study, we were only able to report on associations between variables rather than causal relationships.

To improve the understanding of the complexity of EHR use, we used back-end EHR usage logs. However, because of the anonymity of the survey, we were unable to compare self-reported data with usage logs on an individual basis, and we could only provide an aggregate comparison of 3 measures. Furthermore, usage logs can lack the vital context around clinical workflows, and there has been varied validity and sensitivity of using such logs for mapping out clinical activity [54]. Validation of back-end EHR usage logs through direct observation was not carried out within this study environment; however, this analytics system has been used in previously published literature describing physicians' EHR usage [42].

Despite numerous discussions and publications, there are striking differences in the understanding of what constitutes burnout and substantial variability in prevalence estimates of burnout among physicians [55]. This study used a single question from the Mini-Z, which was previously validated by physicians, and yielded results similar to those of the more commonly used Maslach Burnout Inventory [3].

Finally, because of the heterogeneity of EHRs, implementation practices, training, and organizational contexts, there exists a potential limitation in generalizing such results to other contexts.

Conclusions

This work is the first step in better understanding EHR-related physician burnout in a Canadian academic mental health environment, where we measured general burnout rates and its perceived link to EHR use through a survey that gathered self-perceptions. In addition, we compared self-perceptions with a back-end usage log for 3 important metrics and found that participants tended to overestimate their time spent on the EHR. This finding provides a valuable contribution toward the methodology for studying physician burnout and demonstrates the need to combine self-reported perceptions with objective data sources.

The contribution of this study to the literature on physician burnout demonstrates the importance of increasing end-user satisfaction and minimizing end-user frustration with the EHR, both significant factors that were associated with burnout within the study population. The results of this study emphasize the value of developing *human-centered* effective strategies to improve physicians' experiences with EHRs, including efficient

XSL•FO
RenderX

communication about EHR upgrades. Measuring burnout and understanding the impact of EHR-related stressors within the study population serves as a strong baseline, allowing us to measure the short- and long-term impact of multiple initiatives underway at our hospital aimed at reducing physician burnout.

Acknowledgments

The authors would like to acknowledge Thanara Rajakulendran for her work with content analysis for the open-ended survey data. The authors would also like to acknowledge the Centre for Addiction and Mental Health in Toronto, Canada, for their in-kind contributions to this work.

Authors' Contributions

DJ, TT, GS, and VS conceptualized the study. TT led the design of the survey and recruitment efforts for survey completion and analysis. JK developed data collection tools and managed the data. TT, GS, LS, and TZ drafted the manuscript and performed data analysis, along with the help of KR and MS. All authors edited and provided feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed survey questions.
[PDF File (Adobe PDF File), 202 KB - jmir_v22i7e19274_app1.pdf]

References

1. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. Lancet 2009 Nov 14;374(9702):1714-1721. [doi: 10.1016/S0140-6736(09)61424-0] [Medline: 19914516]
2. Nahum M, van Vleet TM, Sohal VS, Mirzabekov JJ, Rao VR, Wallace DL, et al. Immediate mood scaler: tracking symptoms of depression and anxiety using a novel mobile mood scale. JMIR Mhealth Uhealth 2017 Apr 12;5(4):e44 [FREE Full text] [doi: 10.2196/mhealth.6544] [Medline: 28404542]
3. Maslach C, Jackson S, Leiter M, Schaufeli W, Schwab R. The maslach burnout inventory manual. In: Zalaquett CP, Wood RJ, editors. Evaluating Stress: A Book of Resources. Palo Alto, CA: The Scarecrow Press; 1986.
4. Leigh JP, Kravitz RL, Schembri M, Samuels SJ, Mobley S. Physician career satisfaction across specialties. Arch Intern Med 2002 Jul 22;162(14):1577-1584. [doi: 10.1001/archinte.162.14.1577] [Medline: 12123400]
5. Williams ES, Rathert C, Buttigieg SC. The personal and professional consequences of physician burnout: a systematic review of the literature. Med Care Res Rev 2019 Jun 19:- epub ahead of print. [doi: 10.1177/1077558719856787] [Medline: 31216940]
6. Wetterneck TB, Linzer M, McMurray JE, Douglas J, Schwartz MD, Bigby JA, Society of General Internal Medicine Career Satisfaction Study Group. Worklife and satisfaction of general internists. Arch Intern Med 2002 Mar 25;162(6):649-656. [doi: 10.1001/archinte.162.6.649] [Medline: 11911718]
7. Shanafelt TD, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, et al. Burnout and medical errors among American surgeons. Ann Surg 2010 Jun;251(6):995-1000. [doi: 10.1097/SLA.0b013e3181bfdab3] [Medline: 19934755]
8. Physicians Foundation. 2018. 2018 Survey of America's Physicians: Practice Patterns and Perspectives URL: https://physiciansfoundation.org/wp-content/uploads/2018/09/physicians-survey-results-final-2018.pdf [accessed 2020-01-01]
9. Canadian Medical Association. 2018. CMA National Physician Health Survey: A National Snapshot URL: https://www.cma.ca/sites/default/files/2018-11/nph-survey-e.pdf [accessed 2020-01-01]
10. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529. [doi: 10.1111/joim.12752] [Medline: 29505159]
11. Dyrbye LN, West CP, Burriss TC, Shanafelt TD. Providing primary care in the United States: the work no one sees. Arch Intern Med 2012 Oct 8;172(18):1420-1421. [doi: 10.1001/archinternmed.2012.3166] [Medline: 22911276]
12. Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry 2016 Jun;15(2):103-111 [FREE Full text] [doi: 10.1002/wps.20311] [Medline: 27265691]
13. Pai DD, Lautert L, Souza SB, Marziale MH, Tavares JP. [Violence, burnout and minor psychiatric disorders in hospital work]. Rev Esc Enferm USP 2015 Jun;49(3):460-468 [FREE Full text] [doi: 10.1590/S0080-623420150000300014] [Medline: 26107707]
14. Chan MK, Chew QH, Sim K. Burnout and associated factors in psychiatry residents: a systematic review. Int J Med Educ 2019 Jul 30;10:149-160 [FREE Full text] [doi: 10.5116/ijme.5d21.b621] [Medline: 31381505]

XSL•FO
RenderX

15. Gardner R, Cooper E, Haskell J, Harris D, Poplau S, Kroth P, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 1;26(2):106-114. [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
16. Domaney NM, Torous J, Greenberg WE. Exploring the association between electronic health record use and burnout among psychiatry residents and faculty: a pilot survey study. Acad Psychiatry 2018 Oct;42(5):648-652. [doi: 10.1007/s40596-018-0939-x] [Medline: 29785625]
17. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
18. Gephart S, Carrington JM, Finley B. A systematic review of nurses' experiences with unintended consequences when using the electronic health record. Nurs Adm Q 2015;39(4):345-356. [doi: 10.1097/NAQ.0000000000000119] [Medline: 26340247]
19. Vanderhook S, Abraham J. Unintended consequences of EHR systems: a narrative review. Proc Int Sym Hum Fact Ergo 2017 May 15;6(1):218-225. [doi: 10.1177/2327857917061048]
20. Pirnejad H, Bal R, Shahsavar N. The nature of unintended effects of health information systems concerning patient safety: a systematic review with thematic synthesis. Stud Health Technol Inform 2010;160(Pt 1):719-723. [Medline: 20841780]
21. Eikey E, Chen Y, Zheng K. Unintended adverse consequences of health it implementation: workflow issues and their cascading effects. In: Cognitive Informatics: Reengineering Clinical Workflow for Safer and More Efficient Care. New York, USA: Springer; 2019:31-43.
22. Rahadhan P, Poon S, Land L. Understanding unintended consequences for EMR: a literature review. Stud Health Technol Inform 2012;178:192-198. [Medline: 22797041]
23. Kroth P, Morioka-Douglas N, Veres S, Pollock K, Babbott S, Poplau S, et al. The electronic elephant in the room: physicians and the electronic health record. JAMIA Open 2018 Jul;1(1):49-56 [FREE Full text] [doi: 10.1093/jamiaopen/ooy016] [Medline: 31093606]
24. Kroth PJ, Morioka-Douglas N, Veres S, Babbott S, Poplau S, Qeadan F, et al. Association of electronic health record design and use factors with clinician stress and burnout. JAMA Netw Open 2019 Aug 2;2(8):e199609 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9609] [Medline: 31418810]
25. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
26. Varpio L, Rashotte J, Day K, King J, Kuziemsky C, Parush A. The EHR and building the patient's story: a qualitative investigation of how EHR use obstructs a vital clinical activity. Int J Med Inform 2015 Dec;84(12):1019-1028. [doi: 10.1016/j.ijmedinf.2015.09.004] [Medline: 26432683]
27. Busch A, Bates D, Rauch S. Improving adoption of EHRs in psychiatric care. N Engl J Med 2018 May 03;378(18):1665-1667 [FREE Full text] [doi: 10.1056/NEJMp1800465] [Medline: 29719176]
28. Kobus AM, Harman JS, Do HD, Garvin RD. Challenges to depression care documentation in an EHR. Fam Med 2013 Apr;45(4):268-271. [Medline: 23553091]
29. Triplett P. Psychiatry and the meaningful use of electronic health records. Perspect Biol Med 2013;56(3):407-421. [doi: 10.1353/pbm.2013.0028] [Medline: 24375121]
30. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
31. Gheorghiu B, Hagens S. Measuring interoperable EHR adoption and maturity: a Canadian example. BMC Med Inform Decis Mak 2016 Jan 25;16:8 [FREE Full text] [doi: 10.1186/s12911-016-0247-x] [Medline: 26810606]
32. The Lancet. Physician burnout: the need to rehumanise health systems. Lancet 2019 Nov 2;394(10209):1591. [doi: 10.1016/S0140-6736(19)32669-8] [Medline: 31690430]
33. Gawande A. The New Yorker. 2018 Nov 05. Why Doctors Hate Their Computers URL: https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers [accessed 2020-01-01]
34. Collier R. Electronic health records contributing to physician burnout. Can Med Assoc J 2017 Nov 13;189(45):E1405-E1406 [FREE Full text] [doi: 10.1503/cmaj.109-5522] [Medline: 29133547]
35. Bakken S. Building the evidence base on health information technology-related clinician burnout: a response to impact of health information technology on burnout remains unknown-for now. J Am Med Inform Assoc 2019 Oct 1;26(10):1158. [doi: 10.1093/jamia/ocz078] [Medline: 31135874]
36. The Centre for Addiction and Mental Health. 2019. CAMH Annual Report 2018-2019 URL: https://www.camh.ca/-/media/images/all-other-images/annual-report-2018-2019/camh_annual-report_2018-19.pdf?la=en&hash=5B3C801DF55B334133ED542C5ACE50C34730415F [accessed 2020-05-08]
37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

XSL•FO
RenderX

38. KLAS Research. 2019. Clinician Burnout: An Arch Collaborative Impact Report 2019 URL: https://klasresearch.com/home [accessed 2020-05-26]
39. Linzer M, Poplau S, Babbott S, Collins T, Guzman-Corrales L, Menk J, et al. Worklife and wellness in academic general internal medicine: results from a national survey. J Gen Intern Med 2016 Sep;31(9):1004-1010 [FREE Full text] [doi: 10.1007/s11606-016-3720-4] [Medline: 27138425]
40. Williams ES, Konrad TR, Linzer M, McMurray J, Pathman DE, Gerrity M, et al. Refining the measurement of physician job satisfaction: results from the physician worklife survey. SGIM career satisfaction study group. Society of general internal medicine. Med Care 1999 Nov;37(11):1140-1154. [doi: 10.1097/00005650-199911000-00006] [Medline: 10549616]
41. Rohland BM, Kruse GR, Rohrer JE. Validation of a single-item measure of burnout against the Maslach burnout inventory among physicians. Stress and Health 2004 Apr 7;20(2):75-79. [doi: 10.1002/smi.1002]
42. Overhage JM, McCallie D. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. Ann Intern Med 2020 Feb 4;172(3):169-174. [doi: 10.7326/M18-3684] [Medline: 31931523]
43. IBM - United States. 2016. Downloading IBM SPSS Statistics 25 URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2019-09-01]
44. The Global Biodiversity Information Facility. 2013. R: A Language and Environment for Statistical Computing URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2020-05-26]
45. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
46. National Academies of Sciences Engineering and Medicine, National Academy of Medicine, ommittee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. Washington, DC: The National Academies Press; 2019.
47. Stanford Medicine. 2019. How Doctors Feel About Electronic Health Records URL: http://med.stanford.edu/content/dam/sm/ehr/documents/EHR-Poll-Presentation.pdf [accessed 2020-01-01]
48. Privitera MR, Atallah F, Dowling F, Gomez-DiCesare C, Hengerer A, Arnhart K, et al. Physicians' electronic health records use at home, job satisfaction, job stress and burnout. J Hosp Admin 2018 Jun 20;7(4):52. [doi: 10.5430/jha.v7n4p52]
49. Khairat S, Coleman C, Ottmar P, Bice T, Koppel R, Carson S. Physicians' gender and their use of electronic health records: findings from a mixed-methods usability study. J Am Med Inform Assoc 2019 Dec 1;26(12):1505-1514. [doi: 10.1093/jamia/ocz126] [Medline: 31504578]
50. Dastagir M, Chin H, McNamara M, Poteraj K, Battaglini S, Alstot L. Advanced proficiency EHR training: effect on physicians' EHR efficiency, EHR satisfaction and job satisfaction. AMIA Annu Symp Proc 2012;2012:136-143 [FREE Full text] [Medline: 23304282]
51. Chaiyachati KH, Shea JA, Asch DA, Liu M, Bellini LM, Dine CJ, et al. Assessment of inpatient time allocation among first-year internal medicine residents using time-motion observations. JAMA Intern Med 2019 Jun 1;179(6):760-767. [doi: 10.1001/jamainternmed.2019.0095] [Medline: 30985861]
52. Gilleland M, Komis K, Chawla S, Fernandez S, Fishman M, Adams M. Resident duty hours in the outpatient electronic health record era: inaccuracies and implications. J Grad Med Educ 2014 Mar;6(1):151-154 [FREE Full text] [doi: 10.4300/JGME-D-13-00061.1] [Medline: 24701327]
53. Robinson J, Martin S, Glorieux I, Minnen J. US Bureau of Labor Statistics Latest Numbers. 2011 Jun. The Overestimated Workweek Revisited URL: https://www.bls.gov/opub/mlr/2011/06/art3full.pdf [accessed 2020-01-01]
54. Rule A, Chiang M, Hribar M. Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods. J Am Med Inform Assoc 2020 Mar 1;27(3):480-490. [doi: 10.1093/jamia/ocz196] [Medline: 31750912]
55. The Lancet. Physician burnout: a global crisis. Lancet 2019 Jul 13;394(10193):93. [doi: 10.1016/S0140-6736(19)31573-9] [Medline: 31305255]

Abbreviations

EHR: electronic health record
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

Edited by G Eysenbach; submitted 14.04.20; peer-reviewed by I Shubina, MA Bahrami; comments to author 19.04.20; revised version received 22.04.20; accepted 27.04.20; published 15.07.20.

Please cite as:
Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, Ramamoorthi K, Zhang T, Jankowicz D
The Influence of Electronic Health Record Use on Physician Burnout: Cross-Sectional Survey
J Med Internet Res 2020;22(7):e19274
URL: https://www.jmir.org/2020/7/e19274
doi:10.2196/19274
PMID:32673234

©Tania Tajirian, Vicky Stergiopoulos, Gillian Strudwick, Lydia Sequeira, Marcos Sanches, Jessica Kemp, Karishini Ramamoorthi, Timothy Zhang, Damian Jankowicz. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adoption and Performance of Complementary Clinical Information Technologies: Analysis of a Survey of General Practitioners

Placide Poba-Nzaou[1], DBA, MSci, CRHA; Sylvestre Uwizeyemungu[2], DBA; Xuecheng Liu[3], PhD

[1]Department of Organization and Human Resources, University of Quebec in Montreal, Montreal, QC, Canada

[2]Accounting Department, University of Quebec in Trois-Rivières, Trois-Rivières, QC, Canada

[3]Statistics Department, 12M Statistical Consulting Services, Montreal, QC, Canada

Corresponding Author:
Placide Poba-Nzaou, DBA, MSci, CRHA
Department of Organization and Human Resources
University of Quebec in Montreal
315, Sainte-Catherine East
Montreal, QC, H2X 3X2
Canada
Phone: 1 514 987 3000 ext 7744
Email: poba-nzaou.placide@uqam.ca

Abstract

Background: The benefits from the combination of 4 clinical information systems (CISs)—electronic health records (EHRs), health information exchange (HIE), personal health records (PHRs), and telehealth—in primary care depend on the configuration of their functional capabilities available to clinicians. However, our empirical knowledge of these configurations and their associated performance implications is very limited because they have mostly been studied in isolation.

Objective: This study aims to pursue 3 objectives: (1) characterize general practitioners (GPs) by uncovering the typical profiles of combinations of 4 major CIS capabilities, (2) identify physician and practice characteristics that predict cluster membership, and (3) assess the variation in the levels of performance associated with each configuration.

Methods: We used data from a survey of GPs conducted throughout the European Union (N=5793). First, 4 factors, that is, EHRs, HIE, PHRs, and Telehealth, were created. Second, a cluster analysis helps uncover clusters of GPs based on the 4 factors. Third, we compared the clusters according to five performance outcomes using an analysis of variance (ANOVA) and a Tamhane T2 post hoc test. Fourth, univariate and multivariate multinomial logistic regressions were used to identify predictors of the clusters. Finally, with a multivariate multinomial logistic regression, among the clusters, we compared performance in terms of the number of patients treated (3 levels) over the last 2 years.

Results: We unveiled 3 clusters of GPs with different levels of CIS capability profiles: *strong* (1956/5793, 37.36%), *medium* (2764/5793, 47.71%), and *weak* (524/5793, 9.04%). The logistic regression analysis indicates that physicians (younger, female, and less experienced) and practice (solo) characteristics are significantly associated with a weak profile. The ANOVAs revealed a strong cluster associated with significantly high practice performance outcomes in terms of the quality of care, efficiency, productivity, and improvement of working processes, and two noncomprehensive medium and weak profiles associated with medium (equifinal) practice performance outcomes. The logistic regression analysis also revealed that physicians in the weak profile are associated with a decrease in the number of patients treated over the last 2 years.

Conclusions: Different CIS capability profiles may lead to similar equifinal performance outcomes. This underlines the importance of looking beyond the adoption of 1 CIS capability versus a cluster of capabilities when studying CISs. GPs in the strong cluster exhibit a comprehensive CIS capability profile and outperform the other two clusters with noncomprehensive profiles, leading to significantly high performance in terms of the quality of care provided to patients, efficiency of the practice, productivity of the practice, and improvement of working processes. Our findings indicate that medical practices should develop high capabilities in all 4 CISs if they have to maximize their performance outcomes because efforts to develop high capabilities selectively may only be in vain.

(J Med Internet Res 2020;22(7):e16300) doi:10.2196/16300

XSL•FO
RenderX

KEYWORDS

electronic health record; personal health record; health information exchange; telehealth; general practitioners; quality of care; efficiency; organizational productivity

Introduction

Background

Over the past several years, a consensus has emerged on the recognition of the potential of clinical information systems (CISs) to improve the health care delivered to patients and save lives [1,2].

Electronic health records (EHRs) are at the heart of the reform of health systems in many developed countries [3] as well as middle-income countries such as Brazil [4] or India [5]. An EHR can be defined as "an electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards, and that can be created, managed, and consulted by authorized clinicians and staff across more than one healthcare organization" [6]. EHRs can "help address the dual problems of high costs and poor quality in health" [7]. For instance, EHRs can assist clinicians to improve the care provided to patients by promoting adherence to guidelines [8], improving medical practice management, saving time, and facilitating condition-specific queries, to name a few [9]. In the specific context of primary health care organizations or general practices, EHRs offer "a unique opportunity to collect a wide range of ecologically valid patient data to support understanding of disease burden and health trajectories over the life-course" [10].

Due to their potential benefits, previous decades have witnessed rapid growth in the adoption of EHRs in health care settings. However, despite considerable investments by governments, the adoption of EHRs in some primary care organizations has been slow, especially in small practices [11], and disparities have been observed in terms of benefits associated with CISs in primary care practices [12]. In addition, other researchers have underscored that the potential benefits of EHRs are limited when health information stored in the system is shared only within the host institution, which means that greater benefits will be realized if only health information is shared beyond the host institution [13,14] and the technology used to support such sharing is health information exchange (HIE). For this reason, the *Meaningful Use* program in the United States includes incentives for health care providers to participate in HIE [15]. Empirical evidence suggests that the exchange and sharing of patient data can decrease mortality, systemic costs, and utilization costs in the emergency department [13,16,17].

In addition to the abovementioned CISs (EHR and HIE), the personal health record (PHR) has recently been gaining attention because of its potential to support the transformation of health systems to a more patient-centered model of care [18] as well as its key role in patient engagement [19,20]. Indeed, patient engagement is recognized as a critical factor for improving the quality of care [19] and patient safety [19,21,22]. It is not surprising, then, that patient engagement measures and the sharing of health information between providers and patients are also key parts of the *Meaningful Use* program in the United States [18,23]. Although the role of patients in the health care process is being increasingly recognized, Krist and Woolf [24] have pointed out that much of the energy of the health information movement has been devoted to the use of health information by clinicians, even though patients' use of these technologies carries equal promise. One of the most effective ways to share electronic health data with patients is via PHRs.

Similarly, telehealth has been gaining attention because of its potential to reduce barriers to access health care and to save time and reduce costs for remote patients [25,26]. In addition, telehealth can "be clinically supportive and educative by facilitating contact with peers" and, in turn, education can enhance the quality of care provided to patients [27,28].

In conclusion, EHR, HIE, PHR, and telehealth can be considered the most important components of a modern and desirable CIS for both hospital and primary care practices. However, in previous research, these 4 CISs have been studied in isolation with little or no attention to their combination and associated implications for performance outcomes. A search in Medical Literature Analysis and Retrieval System Online (MEDLINE); March 2020) using the 4 terms, "*electronic health records*," "*health information exchange*," "*personal health record*," and "*tele-health*," in all fields yielded no hits. The same search performed in *Article title*, *Abstract*, *Keywords* in Scopus (March 2020) yielded no hits. We conclude that there is some evidence from which one can infer that the body of knowledge in the way capabilities associated with the 4 CISs empirically coexist to form *configurations* or *profiles*, and the associated implication for performance outcomes is very limited. Of note is the fact that most studies have investigated the 4 CISs in isolation, including their implications for performance outcomes. For instance, several current reviews have found ambivalent or no significant relationship between EHRs and performance outcomes in primary care settings [29,30]. For their part, Black et al [29] concluded that "there is a large gap between the postulated and empirically demonstrated benefits of eHealth technologies [such as EHR, HIE, PHR, and telehealth]."

This paper takes a configurational perspective. In a broad sense, configuration is defined as "any multidimensional constellation of conceptually distinct characteristics that commonly occur together" [31].

Following Miller [32]; Fiss, Marx, and Cambre´ [33]; and Delery and Doty [34], we argued that studying the configurations of CIS adoption by primary care practices will not only allow us to take a holistic view of the adoption of CIS by these organizations as it is these patterns or profiles of 4 CISs rather than single isolated CISs that are related to performance, but it will also help to reveal insights that would have been otherwise difficult to obtain. In addition, configuration approaches help to investigate *equifinality*—defined as a concept in which different primary care practice profiles with different configurations of CIS adoption arrive at the same level of the

XSL•FO
RenderX

outcome measure in terms of practice performance outcomes [35].

Objectives

This study has 4 primary goals. The first is to identify predictors of the adoption of clinical systems by general practitioners (GPs) in Europe. The second is to identify and characterize GPs by uncovering typical profiles or patterns of the combination of 4 major CIS capabilities: EHRs, HIEs, PHRs, and telehealth. Consequently, this second objective is inductive in nature, empirically based, and taxonomic, dedicated to classification and subdivision [36,37]. The third objective is to identify physician and practice characteristics that predict cluster or profile membership. Finally, given that the ultimate objective of investing in CISs is to improve the quality of care provided to patients while decreasing cost, the fourth objective is to assess the variation in levels of performance outcomes associated with practice in each configuration or profile.

Methods

Data Source and Sample

We used a data set provided by the European Commission (EC) from the 2018 survey of European GPs. The objective of this study was to understand and measure the actual adoption and use of information and communication technology (ICT) and electronic health (eHealth) applications by general practitioners (GPs) in the 27 countries of the European Union (EU27) as well as changes in uptake over time. The 2018 survey of European GPs was a follow-up study of the 2013 survey, which included the EU27 plus 4 other countries (Croatia, Iceland, Norway, and Turkey).

Given that this 2018 survey of European GPs was a follow-up of the 2013 survey, eHealth is broadly defined, as in the previous study, as "the use of Information and Communication Technologies (ICT) across the whole range of health care functions" [38]. The 2018 survey used the same methodological approach and questionnaire as the previous study. Data collection was based on a survey in the EU27 using mixed methodology (web-based; web CATI, where CATI stands for computer-assisted telephone interviews; and face-to-face) and was conducted between January and June 2018 [38,39]. The data collection process was endorsed by the European Union of General Practitioners. A detailed description of the methodological approach is provided only in the 2013 survey report. The EC team took several measures to ensure the representativeness of the sample. The details are provided in the Technical Compendium of 121 pages published by the EC and available to the public [38]. The identification of the universe of the survey was based on the definition from the World Organization of Family Doctors Europe that characterizes GPs in European countries as "specialist physicians trained in the specialty of primary care who 'exercise their professional role by promoting health, preventing disease and providing cure, care, palliation and promoting patient empowerment and self-management'." The universe was composed of 425,622 GPs [39].

The questionnaire was composed of 3 parts: (1) GPs' sociodemographics, organizational settings, practice location, description of tasks, and workload; (2) ICT availability and use within a GP practice that is divided into 4 categories: EHRs, HIE, telehealth, and PHRs; and (3) attitudinal questions as well as questions related to motivations, perceived barriers, and impacts of ICT.

Following the previous 2013 survey approach, a final sample of 5793 GPs was randomly selected over the analyzed EU27, with an overall sampling error of plus or minus 1.30% [39].

Given that the 4 objectives of this study are related to CIS adoption and associated with the implications for performance outcomes, out of the initial sample of 5793 GPs, only the 5244 who had an EHR system and stored patient data electronically were considered for a subsequent analysis. Due to the presence of missing values in 3 variables (HIE, PHR, and telehealth), we applied a multiple imputation strategy. Among the 5244 subjects, 5022 (95.77%) subjects had complete data, 100 (1.91%) subjects had missing values on telehealth only, 72 (1.37%) subjects had missing values on PHR only, and the remaining 50 (<1%) subjects had other missing patterns.

We imputed the missing values using a multiple imputation procedure. On the one hand, multiple imputation methods perform better than single imputation ones. We selected the multiple imputation method of fully conditional specification with the Proc MI procedure in SAS software version 9.4 (SAS Institute) because of its flexibility in allowing us to define the multivariate model by a series of conditional models, one for each incomplete variable [40]. The 4 variables related to each CIS were chosen for the imputation model: EHR, HIE, telehealth, and PHR. For each of the 4 variables, we then chose regression because the regression model allows the specification of a minimum and a maximum value for imputed values on a variable-by-variable basis [41].

On the other hand, it is impossible to use multiple imputed data sets for cluster analysis as it produces different results of clusters, and it seems that there are no methods to combine these results. Therefore, we used the multiple imputation method as discussed but only chose *one* as the number of imputations. The final 4 variables used in the cluster analyses were the ones after imputations and then standardization: xc_EHR_imp_std, xc_HIE_imp_std, xc_TeH_imp_std, xc_PHR_imp_std.

Although the final 4 variables used for the cluster analysis were the ones with an imputation, we performed the same cluster analysis with the same variables without the imputation of missing values.

As stated earlier, the 549 GPs who did not store patients' data electronically and those who did not have an EHR were excluded from the analysis. The final sample in our analysis was 5793 – 549 = 5244 subjects. Table 1 compares the characteristics of the 549 and 5244 subsamples. Concerning the total sample (N=5244) composition in terms of countries, 8 countries represent a little less than 50% (49.77%) of the sample, each country accounting for about 5.54% and 7.15% of the sample, respectively. These countries include France, Italy,

XSL•FO
RenderX

Poland, Portugal, Romania, Spain, the United Kingdom, and the Netherlands.

Table 1. Characteristics of the respondents and their practices.

Variables, characteristics (ie, levels for categorical variables)	Nonsampled (n=549)	Sampled (n=5244)	Chi-square (*df*)	*t* test (*df*)	*P* value
Gender, n (%)					
Male	247 (45.0)	2652 (50.57)	62 (1)	N/A[a]	.01
Female	302 (55.0)	2592 (49.43)	62 (1)	N/A	.01
Age (years), mean (SD)	53.23 (10.35)	51.86 (10.82)	N/A	2.8 (5791)	.005
Years spent in general practice, mean (SD)	21.07 (11.13)	20.83 (11.22)	N/A	0.5 (5791)	.64
Professional status, n (%)					
Working in a health center	159 (29.0)	1547 (30.02)	103.0 (3)	N/A	<.001
Self-employed GPs[b] working alone	272 (49.5)	2013 (38.39)	103.0 (3)	N/A	<.001
Self-employed GPs working in a group practice	37 (7)	1226 (23.38)	103.0 (3)	N/A	<.001
Other	81 (15)	431 (8.2)	103.0 (3)	N/A	<.001
Size of the practice (number of physicians), n (%)					
Solo (1)	320 (58.3)	2155 (41.09)	67.4 (3)	N/A	<.001
Small (2-4)	58 (11)	920 (17.5)	67.4 (3)	N/A	<.001
Medium (5-9)	60 (11)	995 (19.0)	67.4 (3)	N/A	<.001
Large (10 or more)	111 (20.2)	1174 (22.39)	67.4 (3)	N/A	<.001
Location of workplace, n (%)					
Large city	204 (37.2)	1944 (37.07)	1.0 (2)	N/A	.62
Medium- to small-sized city	156 (28.4)	1402 (26.74)	1.0 (2)	N/A	.62
Rural town	189 (34.4)	1898 (36.19)	1.0 (2)	N/A	.62

[a]N/A: not applicable.

[b]GPs: general practitioners.

Selection of Clustering Variables and Measurements

As recommended by Aldenderfer and Blashfield [42] and Hair et al [37], the selection of clustering variables is theory driven. The selection of variables was based on the fact that they have a common theoretical foundation—they relate to the functional characteristics or capabilities of 1 of the 4 CISs (EHR, HIE, PHR, and telehealth). Indeed, according to the diffusion of innovation theory (DOI) [43], the characteristics of innovation, such as the functional capabilities of CISs are important factors in explaining its adoption.

All clustering variables were dichotomous (1: available and 0: not available). In total, 44 dichotomous variables were selected based on their theoretical relationship; that is, they relate to the functional characteristics of 1 of the 4 CISs (EHR, HIE, PHR, or telehealth). We also used 5 measures of performance. Four measures (quality of care provided to patients, efficiency of the practice, productivity of the practice, and improvement of working processes) were based on a 4-point Likert-type scale ranging from 0 (“strongly disagree”) to 3 (“strongly agree”), and 1 measure (number of patients over the last 2 years) of a categorical type with 3 categories (1: decrease, 2: remain the same, 3: increase)

A standardized average frequency or scaled frequency was computed for each group of variables, including CIS capabilities and performance, resulting in 8 scales. As presented in Table 2, in 7 out of 8 scales, the measurement of reliability is greater than 0.7 [44]. Although the reliability of telehealth was less than 0.7, we decided to keep this factor because Cronbach’s alpha is sensitive to the number of items in the scale and generally tends to underestimate the internal consistency reliability” [45] and “researchers can rely on less stringent reliability measurements for scales consisting of a few items” [46]. We then used standardized average frequencies as the clustering variables.

XSL•FO
RenderX

Table 2. Reliability of the scales.

Factors	Number of items	Cronbach alpha
Clinical system capabilities		
Electronic health record	19	.87
Health information exchange	15	.87
Telehealth	4	.59
Personal health record	6	.80
Performance measurement		
Quality of care provided to patients	6	.92
Efficiency of the practice	5	.89
Productivity of the practice	5	.83
Improvement of personal working practice	4	.80

The statistical analysis was performed in 3 parts: (1) cluster analysis, (2) ANOVAs With Tamhane T2 Post Hoc Tests, and (3) regression analysis.

Statistical Analysis 1: Cluster Analysis

As indicated earlier, we adopted a configurational approach that is taxonomic, based on cluster analysis [31,47]. General practices were classified to reveal patterns based on the available CIS capabilities related to EHR, HIE, PHR, and telehealth. Thereafter, we explored the influence of general practitioner and practice characteristics on the availability of the 4 components of CISs included in this study (EHR, HIE, PHR, and telehealth).

Broadly speaking, cluster analysis is a multivariate statistical technique that classifies items or objects (individuals, firms, or behaviors) from a given population into subgroups or clusters so that items that are classified in the same cluster are more similar to one another than they are to items in another cluster [37,48]. In doing so, the homogeneity of objects within each group is maximized, whereas the heterogeneity between the groups is maximized. Cluster analysis is a completely empirical method of classification that is inductive by nature [49]. Following previous research recommendations [48-50], our procedure was based on 2 steps that allowed the determination of the natural number of clusters in the data set, as it has been found to be more effective than other approaches in the ability to recover clusters. First, a hierarchical algorithm was employed to identify and define natural clusters and associated centroids. These centroids were then used as initial seeds in a nonhierarchical algorithm. Later, a discriminant function analysis was carried out [49] to validate the cluster solution. Several studies have used this technique to validate the results of cluster analysis [51,52].

It is important to remember that contrary to other statistical techniques such as regression analysis, which necessitate the satisfaction of the linearity assumption and have established rules for sample size calculation [53], no such assumptions are necessary for cluster analysis [54]. Given that this is a data mining technique, cluster analysis does not have *hard* sample size rules and does not need to satisfy parametric or even nonparametric statistical test assumptions [55]. Nonetheless, Hair et al [56] underscored that "the sample size must be large enough to provide sufficient representation of small groups within the population and represent the underlying structure." For their part, Formann [57,58] as well as Dolnicar [59] contend that there is a consensus that the minimal sample size for cluster analysis is 2^k observations (k=number of variables) to achieve sufficient power and confidence in the statistical analysis. In the same vein, Lowry et al [55], including Dolnicar [59], indicate that from a conservative perspective, the minimal sample must be no less than 5×2^k. In this study, k=4, which requires a minimum sample size of 80 to meet this criterion, which is obviously less than the sample size of this study (N=5244).

Determination of the Number of Clusters

The optimum number of clusters was determined by inspecting the dendrogram generated in the combination of the Ward minimum variance clustering algorithm and the squared Euclidean distance. This examination revealed that a 3-cluster solution would be optimal. To ascertain the reliability of the solution [48], the procedure was performed with different subsamples that were randomly selected (30%, 40%, and 60%). On the grounds of the preceding analysis, the 3-cluster solution was retained. It was the most meaningful and the one that best captured the patterns of the adoption of functional capabilities of EHR, HIE, PHR, and Telehealth among European GPs. The uncovered clusters formed significantly well-separated groups that had strong EHR, HIE, PHR, and telehealth functional capability adoption meanings. The 3 different patterns identified reveal how CIS capability priorities set at macrolevels (countries or EC), actually and empirically manifest at primary care levels.

It is worth recalling that cluster analysis was performed with and without the imputation of missing values. Of note is the fact that the results were similar with and without the imputation of missing values. As cluster analyses necessitate all variables to have nonmissing values, 740 observations were not classified in the analysis without the imputation of missing values. The two 3-cluster solutions were then compared to determine the degree of agreement among members of each cluster using Cohen kappa coefficient. The results reveal an almost perfect agreement between the two 3-cluster solutions with a kappa of 0.99 (kappa in the range of 0.80-1.00) [60].

XSL•FO
RenderX

Validation of the Cluster Solution and Multiple Discriminant Analysis

A multiple discriminant function analysis was performed following a cross-validation approach [37] to validate the 3-cluster solution and determine how the clusters differed on the 4 clustering variables portraying GPs based on EHR, HIE, PHR, and Telehealth functional capabilities [49]. Our discriminant analysis produced 2 functions with significant Wilks lambdas (P<.001). In addition, the hit ratios for the analysis and holdout samples were 95.90% and 94.70%, respectively. Two standard criteria were used to evaluate the accuracy of the hit ratios: the maximum chance criterion (C_{max}) and the proportional chance criterion (C_{pro}) [45]. The last authors [45] underscored C_{max} as the most conservative standard in that it will produce the highest standard of comparison. In this case, because the 3 clusters have different sizes, the C_{max} is equal to 52.70%. Hair et al [45] suggested that the classification accuracy should be at least one-fourth greater than the accuracy of the classification achieved by chance ($1.25 \times C_{pro}$): C_{pro}=43% and $1.25 \times C_{pro}$=53.75%. Overall, in a conservative view, both hit ratios are greater than maximum 1.25 (C_{pro} and C_{max}). In conclusion, the null hypothesis (that the percentage correctly classified was not significantly different from what would be classified by chance alone) was rejected. Finally, from the distribution of the GPs along with the clusters' centroids in the plane of the 2 computed discriminant functions (ie, the scatter plot of GPs in the space of the 2 discriminant functions with the clusters' centroids), we were able to observe a complete separation of each one of the centroids from another along the 2 discriminant axes [61]. The evidence suggests that the discriminant functions performed very well in separating the 3 groups.

Although our empirically derived taxonomy appears to be meaningful, its quality is discussed in light of Rich [62] framework on requirements for a valid taxonomy: breadth, meaning, depth, theory, quantitative measurement, completeness, logic, and recognizability. Following the call by Dayer et al [63] to adapt the framework when applied to taxonomies created through cluster analysis, the assessment is based on 5 of the 7 initial criteria.

Breadth

The selection of clustering variables from which GPs were grouped was theory driven [37,42]. As stated earlier, the 18 clustering variables were selected based on their having a common theoretical foundation, that is to say, they pertain to the functional characteristics of 1 of the 4 CISs (EHR, HIE, PHR, or Telehealth), and previous studies have confirmed the influence of these clustering variables on the adoption of technologies [43,64,65].

Meaning

The resulting taxonomy is built on the broad foundations of the DOI theory [43], which acknowledges the complexity of the adoption of innovation and suggests that GPs should not be considered a homogeneous group when investigating the adoption of CIS capabilities. This theory supports the need for classification of the GPs adopting CIS based on the capabilities they actually adopted. The resulting taxonomy highlights different clear priorities and routes to the adoption of the 4 major CIS capabilities.

Theory

The anchoring of the development of our taxonomy in the theory of DOI [43] provides a robust theoretical foundation that serves as a qualitative basis for GP grouping justification as well as variable selection and helps to describe and understand the adoption of the 4 major CISs (EHR, HIE, PHR, and Telehealth) at the medical practice level.

Quantitative Measurement

GPs were assigned to specific clusters or groups resulting from an inductive process based on empirical, multivariate data analysis as well as the application of ANOVA and post hoc analysis to enhance the validity of the results.

Recognizability

By deriving the taxonomy from the actual capabilities of the CIS circumscribed by each artifact (EHR, HIE, PHR, and telehealth technologies), we can claim that our taxonomy reflects the real world for both practitioners as well as theorists and depicts the actual landscape of EHR, HIE, PHR, and Telehealth adoption by GPs within the EU.

In a subsequent step, the 3 profiles (clusters) were compared according to 4 performance outcomes (quality of care provided to patients, efficiency of the practice, productivity of the practice, improvement of working processes) using an ANOVA and a Tamhane T2 post hoc test.

Statistical Analysis 2: ANOVAs With Tamhane T2 Post Hoc Tests

We used ANOVAs with Tamhane T2 post hoc tests to compare the 4 performance outcomes that were rated with a Likert scale: quality of care provided to patients, efficiency of the practice, productivity of the practice, and improvement of personal working practice.

Statistical Analysis 3: Regression Analysis

This analysis was performed in 4 steps. First, we used a regression model to identify the predictors of CIS adoption.

Second, for each of the following 6 characteristic variables, that is, 4 physician characteristics (gender, age, professional status, and years spent in general practice) and 2 practice characteristics (workplace location and practice size), a univariate multinomial logistic regression was conducted to analyze the effect of the characteristic variable on cluster membership.

Third, the multivariate multinomial logistic regression model was conducted with the 6 characteristic variables as independent variables and the 3-cluster solution as the outcome variable. This was used to analyze the effect of each characteristic variable on cluster membership, as shown in the first step, but controlling for the other 5 characteristic variables.

Fourth, the multiple multinomial logistic regression was performed to see if the cluster membership predicted the level of performance in terms of "the number of patients over the

XSL•FO
RenderX

past two (2) years" when controlling for the 6 characteristic variables.

Results

Characteristics of Adopters Versus Nonadopters of Electronic Storage of Patient Data

The logistic regression model indicates that compared with GPs who store patient data electronically, those who do not tend to be older in age, self-employed, working alone, with fewer years spent in general practice (Multimedia Appendix 1). Of note is the fact that the odds ratios (ORs) for age and *years spent in general practice* are close to 1.

Empirical CIS Profiles of GPs

As shown in Table 3, the ANOVA *F* test is highly statistically significant for all 4 factors or groups of functional capabilities. In addition, the Tamhane T2 post hoc multiple pairwise comparison test revealed significant differences between the means of all 4 factors or groups of functional capabilities across the 3 clusters [66]. It is important to recall that, contrary to other tests such as the Tukey Honestly Significant Difference Test, the Tamhane T2 post hoc test does not assume equal variances.

Table 3. Capabilities profile and analysis of variance of clinical information systems.

Variables	Cluster			ANOVA[a]	
	1	2	3	*F* test (*df*)	*P* value
Number of participants (n=5244), n (%)	1956 (37.30)	2764 (52.71)	524 (10.0)	N/A[b]	N/A
Clinical system capabilities[c]					
Electronic health record	*H:* 0.39[d]	*M:* 0.19[e]	*L:* −2.47[f]	5675.8 (2)	<.001
Health information exchange	*H:* 0.88[d]	*M:* −0.42[e]	*L:* −1.05[f]	2517.1 (2)	<.001
Telehealth	*H:* 0.62[d]	*M:* −0.37[e]	*M:* −0.33[f]	764.7 (2)	<.001
Personal health record	*H:* 0.93[d]	*M:* −0.55[e]	*M:* −0.56[f]	2727.4 (2)	<.001

[a]ANOVA: analysis of variance.

[b]N/A: not applicable.

[c,d,e,f]Within rows, different superscripts indicate significant (*P*<.05) pairwise differences between means on Tamhane T2 post hoc test. H=high; M=medium; L=low.

Cluster I

The strong profile (n=1956) is the second largest of the 3 clusters and accounts for approximately 37% of the sample. Statistically, this cluster scored *high* in the pairwise difference between the means of all 3 groups for all 4 CIS capabilities. GPs within this cluster have a strong CIS capability profile and pay a great deal of attention to all 4 CISs, namely EHR, HIE, PHR, and telehealth. Thus, cluster I is named the *strong* CIS capabilities profile.

Cluster II

The moderate profile (n=2764) is the largest of the 3 clusters and accounts for approximately 53% of the sample. This cluster scored *medium* in the pairwise difference between the means of all 4 CIS capabilities. GPs within this cluster have an equally moderate emphasis on all 4 CIS capabilities. Thus, cluster II is named the *moderate* CIS capabilities profile.

Cluster III

The weak profile (n=524) is the smallest of the 3 profiles and accounts for approximately 10% of the sample. This cluster scored *low* in the pairwise difference between the means for 2 of the 4 CIS capabilities (EHR and HIE), whereas scoring *medium* for Telehealth and PHR capabilities. GPs within this cluster exhibit a focus on Telehealth and PHR capabilities, albeit with moderate strength. Thus, cluster III is named the *weak* CIS capabilities profile.

Predictors of CIS Profile Membership

Univariate and multivariate logistic regression analyses were performed (Multimedia Appendix 2) with CIS profiles as the outcome and 6 independent variables of physician and practice characteristics: gender, age, physician professional status, workplace location, years spent in general practice, and practice size). Of note is the fact that the outcome variable has 3 CIS profiles, and the odds in the Multimedia Appendix are to the *strong* profile as the reference.

Physician Gender

The multivariate model indicated that female GPs were more likely than their male counterparts to be in the *weak* (OR 1.52, 95% CI 1.24-1.87) or *moderate* profiles (OR 1.18, 95% CI 1.04-1.33) when controlling for other GPs' individual (age, physician status, and years spent in general practice) and practice (workplace location and practice size) characteristics.

Physician Professional Status

The multivariate logistic model indicated that the GPs working in a health center were less likely than the self-employed GPs working alone to be in the *weak* profile (OR 0.46, 95% CI 0.24-0.87) or *moderate* profile (OR 0.65, 95% CI 0.42-0.99) when controlling for other physicians and practice characteristics.

Similarly, self-employed GPs working in a group practice were found to be less likely than the self-employed GPs working

XSL•FO
RenderX

alone to be in the *weak* profile (OR 0.44, 95% CI 0.22-0.87). However, there was no association between the 2 professional status and membership in the *moderate* profile. The same result is obtained when comparing GPs working neither with a health center nor as self-employed GPs in a group practice with the self-employed GPs working alone in relation to membership with any of the nonstrong profiles.

Physician Age

The multivariate model indicates no association between physician age and membership in the *weak* profile. However, the multivariate model indicates that older GPs are less likely to be in the *moderate* profile, with the OR decreasing to 0.97 (95% CI 0.96-0.98) when controlling for other characteristics.

Years Physicians Spent in General Practice

The multivariate model indicates that senior GPs are less likely to be in the *weak* profile, with the OR decreasing to 0.97 (95% CI 0.96-0.99) when controlling for other characteristics.

The multivariate model also indicates that GPs with more years of practice are more likely to be in the *moderate* profile, with the OR increasing to 1.02 (95% CI 1.00-1.03) for each additional year spent in practice when controlling for other characteristics.

Workplace Location

The multivariate model indicates that GPs within a practice located in a medium- to small-sized city or in a rural town are less likely than those located in a large city to be in the *weak* profile, with OR 0.741 (95% CI 0.58-0.95) and OR 0.60 (95% CI 0.48-0.77), respectively. No significant relationship was found between GPs more or less likely to be in the *moderate* profile when comparing workplaces either between medium- to small-sized cities and large cities or between rural towns and large cities.

Practice Size

First, no association was found for membership to the *moderate* profile when comparing small practice groups with those working in a solo practice.

Second, between medium and solo, GPs working in medium-sized practice groups are less likely to be in the *weak* profile and the *moderate* profile than those in a solo practice, with OR 0.34 (95% CI 0.18-0.66) and OR 0.46 (95% CI 0.30-0.71), respectively.

Finally, between large and solo, GPs working in large practice groups are also less likely to be in the *weak* profile and the *moderate* profile than those in solo practice, with OR 0.37 (95% CI 0.19-0.69) and OR 0.37 (95% CI 0.24-0.56), respectively.

Comparison of CIS Profiles According to Practice Performance

As stated earlier, we used 5 measures of performance, including 4 based on a 4-point Likert-type scale (quality of care provided to patients, efficiency of the practice, productivity of the practice, improvement of working processes) and 1 variable of categorical type that is composed of 3 levels (1: decrease, 2: remain the same, 3: increase) related to the number of patients treated over the last 2 years.

The results in Table 4 indicates that the *strong* profile outperforms the other 2 in terms of the quality of care provided to patients, efficiency of the practice, productivity of the practice, and improvement of working processes. Surprisingly, however, the *moderate* and *weak* profiles do not differ from one another in terms of the quality of care provided to patients, efficiency of the practice, and productivity of practice, whereas the *moderate* profile outperforms the *weak* profile concerning improvement of working processes.

Table 4. Clinical information systems profiles and practice performance.

Variables	Cluster			ANOVA[a]	
	1	2	3	*F* test (*df*)	*P* value
Number of participants (n=5244), n (%)	1956 (37.30)	2764 (52.71)	524 (10.0)	N/A[b]	N/A
Performance[c]					
Quality of care provided to patients	*H*: 0.18[d]	*M*: −0.09[e]	*M*: −0.20[e]	53.7 (2)	<.001
Efficiency of the practice	*H*: 0.14[d]	*M*: −0.07[e]	*M*: −0.15[e]	33.3 (2)	<.001
Productivity of the practice	*H*: 0.16[d]	*M*: −0.09[e]	*M*: −0.13[e]	41.0 (2)	<.001
Working processes improvement	*H*: 0.18[d]	*M*: −0.06[e]	*L*: −0.37[f]	73.1 (2)	<.001

[a]ANOVA: analysis of variance.

[b] N/A: not applicable.

[c,d,e,f]Within rows, different subscripts indicate significant (P<.05) pairwise differences between means on the Tamhane T2 post hoc test. H=high; M=medium; L=low.

A multinomial logistic regression model was used to test the association between the 3 profiles and the evolution of the number of patients treated over the past 2 years (a categorical-type variable with 3 categories, ie, 1: decrease, 2: remain the same, 3: increase) by controlling 6 characteristic variables with *stable* level as the reference for the 3-category outcome variable (Multimedia Appendix 3). This model indicates that physicians in the *weak* profile are more likely to have experienced a decrease in the number of patients treated than those in both the *strong* and *moderate* profiles, with

XSL•FO
RenderX

OR 1.67 (95% CI 1.20-2.33) and OR 1.82 (95% CI 1.33-2.48), respectively.

Discussion

Over the past several years, scholars and policy makers have agreed on the unsustainable nature of the increasing trends of health care spending and investing in health information technologies is seen as a viable option in dealing with this threat. As a result, most countries of the Organisation for Economic Co-operation and Development have begun promoting and investing in CISs and making them one of their top priorities. In this context, 4 CISs have emerged as the most important: EHR, HIE, PHR, and telehealth. Although our knowledge of the 4 CISs has been advanced by several studies that have investigated their adoption and associated performance outcomes, the majority did so by considering the 4 CISs in isolation, which implies that our understanding of this complex phenomenon is still limited.

Using data collected by the EC through a survey of 5793 GPs conducted throughout the EU, this study sought to improve our understanding of the adoption of 4 CISs and the implications for performance outcomes. To the best of our knowledge, this study is one of the first to investigate the empirical configurations of the capabilities of the 4 most important CISs (EHR, HIE, PHR, and telehealth) in general practice settings. At the same time, we believe that the findings of this study provide several interesting insights for medical informatics research by confirming, extending, or challenging previous results, in addition to having important normative implications.

First, consistent with previous studies on the adoption of EHR [67,68] or HIE [69] in practice settings or ambulatory care, our results confirm that CISs are less prevalent among older GPs, working alone in solo practices. However, we found no association with gender, and our study revealed 2 surprising results that call for further investigations: (1) GPs with more years of experience are more likely to adopt electronic data storage. Of note is the fact that the prediction of *electronic storage of patients' data* by age and *years spent in general practice* was both weak, with ORs close to 1.0 of 0.97 and 1.02, respectively, and (2) GPs within a practice located in a medium- to small-sized city or a rural town are less likely than those in a large city to be in the *weak* profile.

Second, after measuring the capabilities associated with each of the 4 CISs, we empirically uncovered 3 theoretically meaningful and significantly well-separated configurations of profiles of CIS adoption by GPs. This result empirically confirms that CIS capabilities as organizational elements correlate in an understandable and stable way [31,70], and only a fraction of the theoretically conceivable configurations is viable and apt to be observed empirically [31].

Third, among the 3 profiles, one, the *strong* profile, outperforms the other two and leads to significantly high performance in terms of the quality of care provided to patients, efficiency of the practice, productivity of the practice, and improvement of working processes. Given that the *strong* profile scored high (H) and higher than the other 2 profiles on all 4 CIS capabilities, the contrary would have been very worrying. Similarly, when compared in terms of *improvement of working processes* as a practice performance outcome, the *moderate* cluster that scored medium on all 4 CIS capabilities outperforms the *weak* profile. Although it seems obvious to expect the *moderate* profile to outperform the *weak* profile in certain ways, given the statistically significant differences in 2 CIS capabilities (of 4), the question remains as to why the expected performance gap manifests itself in exactly one performance indicator and why this indicator is *improvement of working process* and not the *quality of care provided to patients* or the *productivity of the practice* or the *efficiency of the practice*. Again, as expected, both the *strong* and *moderate* profiles outperform the *weak* profile in the number of patients treated over the past 2 years, but surprisingly, our results revealed no significant difference between the *strong* and *moderate* profiles (Multimedia Appendix 3).

Fourth, 2 counterintuitive pictures emerged from our results. First, when scrutinizing the *weak* profile, it can be noticed that this group displays a profile that seems to be reversed in terms of CIS capabilities. In fact, it exhibits a *low* score for EHR and HIE capabilities while exhibiting a medium score for telehealth and PHR. One would expect GPs to first build strong capabilities for EHR and HIE before considering the adoption of telehealth and PHR. Second, GPs in the *weak* profile deploy an overall set of CIS capabilities that seems to be inferior to the *moderate* profile, yet achieve *equifinal* performance outcomes [31,35]. More specifically, both profiles exhibit a *medium* score on 3 performance outcomes: *productivity of the practice*, *efficiency of the practice*, and *quality of care*. Following Gruber et al [71], who obtained similar results when linking small firms' capabilities to performance, we contend that these findings also suggest that configurations that lead to relatively higher performance outcomes for general practices in terms of the quality of care, efficiency of the practice, and productivity are not necessarily the inverse of those that lead to lower performance outcomes. It is important to remember that a general practice can be understood as a small firm.

This research contributes to the configuration literature by responding to a call for further empirical research on equifinality [35].

From a practical viewpoint, we contend that the resulting profiles of European GPs will assist policy makers to make sense of the general practice adoption of the 4 major CISs. As indicated in our study, GPs have been separated into "discrete and relatively homogeneous groups" [31] with different emphasis on EHR, HIE, PHR, and Telehealth capabilities, and unveiled associated practice performance consequences regarding five indicators. In particular, our results indicate that for a practice to maximize performance outcomes, it should develop high capabilities in all 4 CISs, because efforts to develop high capabilities on a selective subset may only be in vain, without any significant performance outcome to the point of being equivalent to developing inferior capabilities, at least concerning *quality of care provided to patients*, *efficiency of the practice* and *productivity of the practice*. Thus, policy makers should continue their efforts to stimulate the adoption of the 4 CISs among general practices while raising awareness

XSL•FO
RenderX

of the importance of achieving a *comprehensive profile*. In addition, given the laggardness of solo practices, it could be advised to define specific initiatives targeting this category of general practice.

By investigating the configuration of the 4 most important CISs and the associated implications for performance outcomes, this study explores a topic that has received limited attention until now. Hence, in interpreting our results, one should keep in mind some limitations. First, generalizability may be limited because our sample is composed of only European GPs. Second, there are intrinsic limitations due to the use of secondary data. In fact, we used a data set that was not collected to meet the specific objectives of this study. Third, even though the results of the testing instrument reported in this study indicate high reliability for most scales, one out of the 4 scales measuring CIS capability (telehealth) has a reliability less than 0.6. In addition, even though the instrument has substantial face validity, it has not been subjected to formal psychometric assessment.

Given the paucity of studies that investigate the empirical configurations of the 4 CISs (EHR, HIE, PHR, and Telehealth) at either primary care practice or hospital settings, we encourage other researchers to build upon our results and investigate the configuration of these 4 CISs and the associated implications for performance outcomes in other regions, including hospital settings.

Authors' Contributions

The first author contributed to the conception and design of the study and drafted the first version of the manuscript. The third author performed the statistical calculations using SAS software. All authors reviewed and approved this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General practitioners' characteristics associated with the Adoption of electronic storage of patient data.
[DOCX File , 19 KB - jmir_v22i7e16300_app1.docx]

Multimedia Appendix 2
Results of the logistic regression of general practioners and practice characteristics by cluster.
[DOCX File , 24 KB - jmir_v22i7e16300_app2.docx]

Multimedia Appendix 3
Multinomial logistic regression model to test association between profiles and evolution of the number of patients over the past 2 years, by controlling 6 characteristic variables.
[DOCX File , 18 KB - jmir_v22i7e16300_app3.docx]

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. Institute of Medicine, Board on Health Care Services, Committee onPatient Safety and Health Information Technology. Health IT and Patient Safety: Building Safer Systems for Better Care. Washington, DC: National Academies Press; 2012.
3. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007] [Medline: 18657471]
4. Puttini R, Toffanello A, Chaim R, Alves G, Rotzsch J, Carvalho EO, et al. Semantic Framework for Electronic Health Records. In: Proceedings of the 11th International Conference on Semantic Computing. 2017 Presented at: ICSC'17; January 30-February 1, 2017; San Diego, CA, USA. [doi: 10.1109/icsc.2017.98]
5. Powell AC, Ludhar JK, Ostrovsky Y. Electronic health record use in an affluent region in India: findings from a survey of Chandigarh hospitals. Int J Med Inform 2017 Jul;103:78-82. [doi: 10.1016/j.ijmedinf.2017.04.011] [Medline: 28551005]
6. The National Alliance for Health Information Technology Report to the Office of the National Coordinator for Health Information Technology on Defining Key Health Information Technology Terms. Health IT Answers. 2008. URL: https://www.hitechanswers.net/ [accessed 2019-04-14] [WebCite Cache ID 77cxc0fpv]
7. Mehta N, Vakharia N, Wright A. EHRs in a web 2.0 world: time to embrace a problem-list wiki. J Gen Intern Med 2014 Mar;29(3):434-436 [FREE Full text] [doi: 10.1007/s11606-013-2652-5] [Medline: 24129858]
8. Wolfe L, Chisolm MS, Bohsali F. Clinically excellent use of the electronic health record: review. JMIR Hum Factors 2018 Oct 5;5(4):e10426 [FREE Full text] [doi: 10.2196/10426] [Medline: 30291099]
9. Benefits of EHRs. Office of the National Coordinator for Health Information. 2018. URL: https://www.healthit.gov/topic/health-it-basics/benefits-ehrs [accessed 2020-06-12]

XSL•FO
RenderX

10. O'Donnell A, Kaner E, Shaw C, Haighton C. Primary care physicians' attitudes to the adoption of electronic medical records: a systematic review and evidence synthesis using the clinical adoption framework. BMC Med Inform Decis Mak 2018 Nov 13;18(1):101 [FREE Full text] [doi: 10.1186/s12911-018-0703-x] [Medline: 30424758]
11. Ramaiah M, Subrahmanian E, Sriram RD, Lide BB. Workflow and electronic health records in small medical practices. Perspect Health Inf Manag 2012;9:1d [FREE Full text] [Medline: 22737096]
12. Mack D, Zhang S, Douglas M, Sow C, Strothers H, Rust G. Disparities in primary care EHR adoption rates. J Health Care Poor Underserved 2016 Feb;27(1):327-338 [FREE Full text] [doi: 10.1353/hpu.2016.0016] [Medline: 27587942]
13. Miller AR, Tucker C. Health information exchange, system size and information silos. J Health Econ 2014 Jan;33:28-42. [doi: 10.1016/j.jhealeco.2013.10.004] [Medline: 24246484]
14. Adler-Milstein J, Lin SC, Jha AK. The number of health information exchange efforts is declining, leaving the viability of broad clinical data exchange uncertain. Health Aff (Millwood) 2016 Jul 1;35(7):1278-1285. [doi: 10.1377/hlthaff.2015.1439] [Medline: 27385245]
15. Health Information Exchange. Office of the National Coordinator for Health Information. 2017. URL: https://www.healthit.gov/topic/health-it-basics/health-information-exchange [accessed 2019-04-14] [WebCite Cache ID 77cx2ovlB]
16. McCullough J, Parente S, Town R. Health Information Technology and Patient Outcomes: The Role of Organizational and Informational Complementarities. National Bureau of Economic Research. 2011. URL: https://www.nber.org/papers/w18684 [accessed 2020-06-24]
17. Rudin RS, Motala A, Goldzweig CL, Shekelle PG. Usage and effect of health information exchange: a systematic review. Ann Intern Med 2014 Dec 2;161(11):803-811. [doi: 10.7326/M14-0877] [Medline: 25437408]
18. Key Considerations for HIE-Based Personal Health Records. Office of the National Coordinator for Health Information. 2015. URL: https://www.healthit.gov/sites/default/files/phrkeyconsiderations.pdf [accessed 2020-06-12]
19. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
20. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
21. Kim JM, Suarez-Cuervo C, Berger Z, Lee J, Gayleard J, Rosenberg C, et al. Evaluation of patient and family engagement strategies to improve medication safety. Patient 2018 Apr;11(2):193-206. [doi: 10.1007/s40271-017-0270-8] [Medline: 28795338]
22. Schwappach DL. Review: engaging patients as vigilant partners in safety: a systematic review. Med Care Res Rev 2010 Apr;67(2):119-148. [doi: 10.1177/1077558709342254] [Medline: 19671916]
23. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
24. Krist AH, Woolf SH. A vision for patient-centered health information systems. J Am Med Assoc 2011 Jan 19;305(3):300-301 [FREE Full text] [doi: 10.1001/jama.2010.2011] [Medline: 21245186]
25. Hepburn SL, Blakeley-Smith A, Wolff B, Reaven JA. Telehealth delivery of cognitive-behavioral intervention to youth with autism spectrum disorder and anxiety: a pilot study. Autism 2016 Feb;20(2):207-218 [FREE Full text] [doi: 10.1177/1362361315575164] [Medline: 25896267]
26. Speyer R, Denman D, Wilkes-Gillan S, Chen Y, Bogaardt H, Kim J, et al. Effects of telehealth by allied health professionals and nurses in rural and remote areas: a systematic review and meta-analysis. J Rehabil Med 2018 Feb 28;50(3):225-235 [FREE Full text] [doi: 10.2340/16501977-2297] [Medline: 29257195]
27. Lucas JAM, Day K, Honey M. Clinician's perceptions of telehealth for emergency care on the West Coast of New Zealand: findings of a descriptive study. Emerg Nurs 2016:6-10 [FREE Full text]
28. Mahadevan S, Muralidhar K, Shetty D. Tele-education service using telemedicine network in healthcare industry. Telemed J E Health 2012 Nov;18(9):699-702. [doi: 10.1089/tmj.2011.0278] [Medline: 23046243]
29. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of ehealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
30. Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. BMC Med Inform Decis Mak 2012 Feb 24;12:10 [FREE Full text] [doi: 10.1186/1472-6947-12-10] [Medline: 22364529]
31. Meyer AD, Tsui AS, Hinings CR. Configurational approaches to organizational analysis. Acad Manage J 1993 Dec 1;36(6):1175-1195. [doi: 10.2307/256809]
32. Miller D. Toward a new contingency approach: the search for organizational gestalts. J Manag Stud 1981;18(1):1-26. [doi: 10.1111/j.1467-6486.1981.tb00088.x]
33. Fiss P. Configurational theory and methods in organizational research: introduction. In: Marx C, Cambre B, editors. Configurational Theory and Methods in Organizational Research. Bingley, UK: Emerald; 2013.
34. Delery JE, Doty DH. Modes of theorizing in strategic human resource management: tests of universalistic, contingency, and configurational performance predictions. Acad Manage J 1996 Aug;39(4):802-835. [doi: 10.5465/256713]

XSL•FO
RenderX

35. Fiss PC. A set-theoretic approach to organizational configurations. Acad Manage Rev 2007 Oct;32(4):1180-1198. [doi: 10.5465/amr.2007.26586092]
36. Fiss PC. Case studies and the configurational analysis of organizational phenomena. In: Ragin C, Bryne D, editors. The SAGE Handbook of Case-Based Methods. Thousand Oaks, CA: Sage Publications; 2009.
37. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: With Readings. Englewood Cliffs, NJ: Prentice Hall; 2010.
38. Benchmarking Deployment of eHealth among General Practitioners (2013). Digital Health News. 2013. URL: https://www.ehealthnews.eu/download/publications/3880-benchmarking-deployment-of-ehealth-among-general-practitioners-2013 [accessed 2019-04-14]
39. Benchmarking Deployment of eHealth Among General Practitioners. Publications Office of the European Union. 2018. URL: https://op.europa.eu/en/publication-detail/-/publication/d1286ce7-5c05-11e9-9c52-01aa75ed71a1/language-en# [accessed 2019-12-20]
40. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007 Jun;16(3):219-242. [doi: 10.1177/0962280206074463] [Medline: 17621469]
41. Horton NJ, Lipsitz SR. Multiple imputation in practice. Am Stat 2001 Aug;55(3):244-254. [doi: 10.1198/000313001317098266]
42. Aldenderfer M, Blashfield R. Cluster Analysis. Beverly Hills: Sage Publications; 1984.
43. Rogers E. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
44. Nunnally J. Psychometric Theory. Second Edition. New York, USA: McGraw-Hill; 1978.
45. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: With Readings. London, UK: Pearson Education Limited; 2014.
46. Al-Omiri M, Drury C. A survey of factors influencing the choice of product costing systems in UK organizations. Manag Account Res 2007 Dec;18(4):399-424. [doi: 10.1016/j.mar.2007.02.002]
47. Fiss PC. Building better causal theories: a fuzzy set approach to typologies in organization research. Acad Manag J 2011 Apr;54(2):393-420. [doi: 10.5465/amj.2011.60263120]
48. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. Strateg Manag J 1996 Jun;17(6):441-458. [doi: https://doi.org/10.1002/(SICI)1097-0266(199606)17:6%3C441::AID-SMJ819%3E3.0.CO;2-G]
49. Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. J Mark Res 2018 Dec 13;20(2):134-148. [doi: 10.1177/002224378302000204]
50. Milligan GW, Sokol LM. A two-stage clustering algorithm with robust recovery characteristics. Educ Psychol Meas 2016 Jul 2;40(3):755-759. [doi: 10.1177/001316448004000320]
51. Kwon J, Johnson ME. Security practices and regulatory compliance in the healthcare industry. J Am Med Inform Assoc 2013 Jan 1;20(1):44-51 [FREE Full text] [doi: 10.1136/amiajnl-2012-000906] [Medline: 22955497]
52. Repschlaeger J, Erek K, Zarnekow R. Cloud computing adoption: an empirical study of customer preferences among start-up companies. Electron Mark 2013 Feb 10;23(2):115-148. [doi: 10.1007/s12525-012-0119-x]
53. Bartlett JE, Kotrlik JW, Higgins CC. Organizational research: determining appropriate sample size in survey research. Inf Technol J 2001;19(1):43-50.
54. Everitt B. Cluster Analysis. Second Edition. London, UK: Heineman Educational Books; 1980.
55. Lowry PB, Moody GD, Gaskin J, Galletta DF, Humpherys SL, Barlow JB, et al. Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: do expert journal assessments add value? MIS Q 2013 Apr 4;37(4):993-1012. [doi: 10.25300/misq/2013/37.4.01]
56. Hair J, Black W, Babin B, Anderson R, Tatham R. Multivariate Data Analysis. London, UK: Pearson Education Limited; 2013.
57. Formann A. Die Latent-Class-Analyse: Einfuhrung in die Theorie und Anwendung. Weinheim: Belz; 1984.
58. McCrimmon AW, Schwean VL, Saklofske DH, Montgomery JM, Brady DI. Executive functions in Asperger's syndrome: an empirical investigation of verbal and nonverbal skills. Res Autism Spectr Disord 2012 Jan;6(1):224-233. [doi: 10.1016/j.rasd.2011.05.003]
59. Dolnicar S. Using cluster analysis for market segmentation - typical misconceptions, established methodological weaknesses and some recommendations for improvement. Australas J Mark Res 2003;11(2):5-12. [doi: 10.1016/S1441-3582]
60. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
61. Tabachnick B, Fidell L. Using Multivariate Statistics. Fifth Edition. Boston, MA: Allyn & Bacon/Pearson Education; 2007.
62. Rich P. The organizational taxonomy: definition and design. Acad Manage Rev 1992 Oct;17(4):758-781. [doi: 10.5465/amr.1992.4279068]
63. Dayer AA, Allred SB, Stedman RC. Comparative analysis and assessment of forest landowner typologies. Soc Nat Resour 2014 Sep 15;27(11):1200-1212. [doi: 10.1080/08941920.2014.933931]

XSL•FO
RenderX

64. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
65. Greenhalgh T, Robert G, Bate P, Macfarlane F, Kyriakidou O. Diffusion of Innovations in Health Service Organisations: A Systematic Literature Review. New Jersey, US: John Wiley & Sons; 2008.
66. Hochberg Y, Tamhane A. Multiple Comparisons Procedures. New York, USA: Wiley; 1987.
67. Xierali IM, Phillips RL, Green LA, Bazemore AW, Puffer JC. Factors influencing family physician adoption of electronic health records (EHRs). J Am Board Fam Med 2013;26(4):388-393 [FREE Full text] [doi: 10.3122/jabfm.2013.04.120351] [Medline: 23833153]
68. des Roches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, et al. Electronic health records in ambulatory care--a national survey of physicians. N Engl J Med 2008 Jul 3;359(1):50-60. [doi: 10.1056/NEJMsa0802005] [Medline: 18565855]
69. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
70. Donaldson L. In Defence of Organisation Theory. Cambridge, UK: Cambridge University Press; 1986.
71. Gruber M, Heinemann F, Brettel M, Hungeling S. Configurations of resources and capabilities and their performance implications: an exploratory study on technology ventures. Strateg Manag J 2010 May 18;31(12):1337-1356. [doi: 10.1002/smj.865]

Abbreviations

ANOVA: analysis of variance
CIS: clinical information system
DOI: diffusion of innovation
EC: European commission
eHealth: electronic health
EHR: electronic health record
EU: European Union
GP: general practitioner
HIE: health information exchange
ICT: information and communication technology
OR: odds ratio
PHR: personal health record

Edited by G Eysenbach; submitted 17.09.19; peer-reviewed by M Jones, I Mircheva; comments to author 11.11.19; revised version received 12.04.20; accepted 14.05.20; published 23.07.20.

Please cite as:
Poba-Nzaou P, Uwizeyemungu S, Liu X
Adoption and Performance of Complementary Clinical Information Technologies: Analysis of a Survey of General Practitioners
J Med Internet Res 2020;22(7):e16300
URL: http://www.jmir.org/2020/7/e16300/
doi:10.2196/16300
PMID:32706715

©Placide Poba-Nzaou, Sylvestre Uwizeyemungu, Xuecheng Liu. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 23.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Information Overload in Emergency Medicine Physicians: A Multisite Case Study Exploring the Causes, Impact, and Solutions in Four North England National Health Service Trusts

Laura Sbaffi[1], BSc, MSc, MA, DPhil; James Walton[2], BSc, MBChB, MRCEM, FRCEM; John Blenkinsopp[3], BA (Hons); Graham Walton[4], BSc, MA, MBA, DPhil

[1]Information School, University of Sheffield, Sheffield, United Kingdom

[2]Leeds Teaching Hospitals Trust, Leeds, United Kingdom

[3]North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees, United Kingdom

[4]Centre for Information Management, Loughborough University, Loughborough, United Kingdom

Corresponding Author:
Laura Sbaffi, BSc, MSc, MA, DPhil
Information School
University of Sheffield
211 Portobello Street
Regent Court
Sheffield, S1 4DP
United Kingdom
Phone: 44 114 2222686
Email: L.Sbaffi@sheffield.ac.uk

Abstract

Background: Information overload is affecting modern society now more than ever because of the wide and increasing distribution of digital technologies. Social media, emails, and online communications among others infuse a sense of urgency as information must be read, produced, and exchanged almost instantaneously. Emergency medicine is a medical specialty that is particularly affected by information overload with consequences on patient care that are difficult to quantify and address. Understanding the current causes of medical information overload, their impact on patient care, and strategies to handle the inflow of constant information is crucial to alleviating stress and anxiety that is already crippling the profession.

Objective: This study aims to identify and evaluate the main causes and sources of medical information overload, as experienced by emergency medicine physicians in selected National Health Service (NHS) trusts in the United Kingdom.

Methods: This study used a quantitative, survey-based data collection approach including close- and open-ended questions. A web-based survey was distributed to emergency physicians to assess the impact of medical information overload on their jobs. In total, 101 valid responses were collected from 4 NHS trusts in north England. Descriptive statistics, principal component analysis, independent sample two-tailed *t* tests, and one-way between-group analysis of variance with post hoc tests were performed on the data. Open-ended questions were analyzed using thematic analysis to identify key topics.

Results: The vast majority of respondents agreed that information overload is a serious issue in emergency medicine, and it increases with time. The always available culture (mean 5.40, SD 1.56), email handling (mean 4.86, SD 1.80), and multidisciplinary communications (mean 4.51, SD 1.61) are the 3 main reasons leading to information overload. Due to this, emergency physicians experience guideline fatigue, stress and tension, longer working hours, and impaired decision making, among other issues. Aspects of information overload are also reported to have different impacts on physicians depending on demographic factors such as age, years spent in emergency medicine, and level of employment.

Conclusions: There is a serious concern regarding information overload in emergency medicine. Participants identified a considerable number of daily causes affecting their job, particularly the traditional culture of emergency departments being always available on the ward, exacerbated by email and other forms of communication necessary to maintain optimal, evidence-based practice standards. However, not all information is unwelcome, as physicians also need to stay updated with the latest guidelines on conditions and treatment, and communicate with larger medical teams to provide quality care.

(J Med Internet Res 2020;22(7):e19126) doi:10.2196/19126

XSL•FO
RenderX

KEYWORDS

emergency medicine; information overload; physicians; national health care system

Introduction

Background

The concept of information overload has existed long before the term was first defined in the 20th century. The writer of Ecclesiastes 12:12 in the third century before Christ was complaining that "of making books there is no end." In the first century, Anno Domini, the Roman philosopher Seneca noted how "the abundance of books is distraction" [1]. Hence, information overload has been experienced by people long before the appearance of today's digital gadgets and social media in the work environment and at home. Most people appreciate the freedom that comes with choice and the many channels of media that saturate the information world. However, a 2016 Pew Research Center survey [2] showed that around 20% of the American population experienced overload and unease in looking for information. According to this study, those with more access to the internet are more likely to express worries about information overload and report difficulty in finding the information they need. Moreover, many institutions such as government agencies, schools, banks, and health care services require a lot of data and information gathering from their customers to process their requests, which contributes to stress and anxiety [2]. In the United Kingdom, a recent survey conducted by TeleWare on 2000 employees revealed that 36% of employees admitted that they had wasted a lot of their working day attempting to resolve an issue where they have forgotten valuable information. A similar number (34%) explained that forgetting information has led them to deal ineffectively with customers, suppliers, or clients. Around a quarter have missed important deadlines (26%) or let their colleagues down (25%) because they did not have the necessary information in the front of their mind [3].

The idea of having every piece of information at hand in any given situation is undoubtedly tempting and, in critical working situations, mandatory [4]. However, this is becoming increasingly unrealistic because of the large volume of data and information. An understanding of the current situation may be gathered by setting information overload in the theoretical framework provided by Floridi [5], whose view entails the definition of 3 main periods in the advancement of humanity: *prehistory*, before recorded information; *history*, when society was aided by recorded information; and *hyperhistory*, where society is shaped and dependent on information and communication technologies. The spreading of the *onlife* human condition, whereby lives are lived both online and offline in an *infosphere*, mirrors the progression of society through hyperhistory. This drastic shift in how people perceive their existence has raised a number of issues, including information overload in a more contemporary, Information and Communication Technology (ICT)-centered lens.

This study aimed to explore the impact of information overload on emergency medicine physicians (EMPs) working across 4 National Health Service (NHS) trusts in north England. The focus was to explore physicians' views on the impact of information overload, including the extent of the impact, whether it had increased, and whether any preventative measures were taken to address information overload. The cross-disciplinary research team included an EMP, 2 academics, and an NHS clinical effectiveness adviser who oversaw the study.

Literature Review

Information Overload in the Workplace

The concept of information overload became generally known in the early 1970s, after the publication of Toffler's book on society's deep organizational changes destined to cause shock and disorientation in people [6]. This initial idea of information overload refers to the difficulty a person can have understanding an issue and making errors caused by the presence of too much information. A later interpretation of the concept, in the 1990s, saw information overload defined as "occurring when the information processing demands on an individual's time to perform interactions and internal calculations exceed the supply or capacity of time available for such processing" [7].

Recent research has demonstrated how there cannot be a unifying definition of the concept, as the context in which it is used changes all the time [8,9]. There is however an agreement that information overload occurs when the information received becomes a burden rather than a help for users [10]. A survey conducted in 2016 by the Chartered Institute of Management Accountants showed how information overload was the main contributor to poor decision making in business globally [11]. Together with weakened decision making, information overload can also lead people to feel unnecessarily overwhelmed and stressed. Recent research has shown that people presenting with large amounts of information while performing search tasks tend to experience a sense of being overwhelmed and even display an increased heart rate, which leads to increased stress levels [12]. In the workplace environment, information overload can lead to a lack of engagement and loss of productivity [13-15]. The increase in mobile technologies has exacerbated this issue. The overload effects associated with the use of mobile information technologies in the workplace have become increasingly dominant and can translate into a considerable reduction in job satisfaction [16,17]. This is mostly because of the perception of immediacy that people have in connection with such devices. It is assumed that emails, text messages, and notifications should be read and answered quickly [18]. Several studies have demonstrated that training in an email or a social media protocol can be effective as a means to reduce information overload [9,19]. Interestingly, people of all ages and backgrounds are affected by this problem in every aspect of their working (and personal) life [9]. However, the same authors stated how the literature tends to focus on the workplace as a whole rather than on job specifications and particular skills required in certain occupations. The same view is shared by Kalman [20], who stated that "the workplace pressures attributed to information overload are very real, the concept itself is vague

XSL•FO
RenderX

and poorly defined.". It is however worth noting that recent research has shown how information overload has the potential to enhance people's cognitive abilities and help them process more information in less time, particularly if the right visualization tools (eg, charts and tables instead of text) are applied in the correct contexts [21]. There are benefits in assimilating larger amounts of information, and the human mind can adapt in time to the increased influx, but it is also bound to manifest side effects such as psychological disturbances and neurological problems along the way [22].

Information Overload and Health Care Professionals

When health care professionals experience information overload, at least one extra layer of concern must be taken into account: the consequence on patient care. Several studies have looked at the effects of information overload on nurses [23,24]; however, few have focused on information overload as experienced in general by health care professionals [25-27] and even less have approached the impact on patient care. Laker et al [28] argued how it could prevent practitioners from seeing the most important details while examining patients. Nevertheless, it is not the *right* information that is overwhelming, as that, if readily available at the point of care, would lead to effective patient care. It is the *background noise* information and other confounding factors (eg, multiple sources of information and conflicting or irrelevant information) that cause disruptions [29].

Klerings et al [27] agreed with this view and suggested solutions to overcome information overload and manage both health care information and patient care effectively. They mentioned different solutions based on the main overload trigger mechanisms, such as "technological solutions; creation or adaption of specific content types; improving health literacy; and creating or strengthening the roles of human intermediaries."

Most likely, the situation is destined to get worse in this era of information expansion. As Yamamoto explained in 2016 [30], health care professionals are already overloaded by the information currently available, but the circumstances will worsen because of 3 key aspects: (1) the need to stay up to date from a technological and knowledge-based point of view; (2) the diversification of medical needs resulting in the growth of the medical practice; and (3) the rapid and constant changes in the health care environment. Digital technologies are proving to be both a curse and a blessing. On the one hand, they represent yet another source of—often unwanted information [31]—but, on the other hand, they can support physicians in the detection of adverse events and patients at risk of deterioration with the potential to improve health outcomes [32]. This is only achievable if, when planning and implementing digital information solutions in hospitals and clinics, factors such as the provision of clinical training, resources to cope with the additional workload, and optimization of algorithms to reduce superfluous alerts are taken into consideration [32].

In particular, emergency physicians are exposed to further burdens with the need for fast decision making [33], interruptions [34], multitasking [34], and highly variable workloads [35]. Emergency medicine is "a form of medical treatment and aid which seeks to treat unexpected, acute illnesses that pose grave threats to life" [36]. The urgent dimension typical of emergency departments requires doctors to be constantly vigilant, highly reactive, and mentally lucid throughout often very long work shifts [37]. This highly pressurized environment leads to occupational stress, often with negative consequences for doctors' mental and physical health and work efficiency, which can lead to mistakes in diagnoses and patient treatments [38]. Chronic workplace stress has been identified by the World Health Organization (WHO) as the leading cause of *burnout syndrome* [39]. According to the WHO, the syndrome is characterized by the following 3 main aspects: feelings of exhaustion, feelings of negativism toward one's job, and reduced professional effectiveness [39]. A considerable body of literature has demonstrated that there is a higher prevalence of burnout among emergency physicians compared with other medical professions and specialties [36,40-42]. It is also evident that information overload plays a determinant role in exacerbating burnout in the medical profession [43,44]. However, only 1 study, conducted by Walton [45] on hospital doctors and pharmacists, has considered the impact of information overload beyond burnout on hospital medical staff. The complex setting in which emergency physicians act and operate requires a better understanding of the impact of information overload on their practice and its implications for both patient care, work efficiency, and work/life balance. To the authors' knowledge, no empirical research has been conducted on information overload as experienced by EMPs specifically and according to different roles, age, and level of seniority; therefore, this paper hopes to serve as a pathbreaker for future studies.

Aim and Objectives

This study aimed to identify and assess the main causes and sources of medical information overload as experienced by EMPs. Specifically, the research objectives (ROs) were as follows:

1. RO1: Determine the extent to which EMPs experience medical information overload.
2. RO2: Establish the contributory factors to information overload and evaluate their impact on physicians' working lives.
3. RO3: Identify the coping mechanisms adopted by EMPs to respond to information overload.
4. RO4: Explore whether EMPs at different career stages respond to information overload in different ways.
5. RO5: Determine the impact of physicians' information overload on patient care.

Methods

Data Collection

The assessment of EMPs' experience with information overload was undertaken using a survey-based primary data collection tool. This was a multicenter study covering 4 NHS trusts in north England, which had approached the staff in their emergency departments and asked them to complete a web-based survey over a period of 3 months in the summer of 2019. All eligible doctors from the trusts were sent an email inviting them to participate in the study. The opening page of

XSL•FO
RenderX

the survey included an information sheet and a consent form. Participants were not able to proceed with the survey unless they had ticked the consent option. The study followed the NHS Health Research Authority protocol and was granted both the NHS Research Ethics Committee and the University of Sheffield ethics approvals. The survey, designed using *LimeSurvey GmbH*, took approximately 10 min to complete and was divided into 2 main sections: the first related to demographics and the second addressing emergency physicians' views on medical information overload, opening with a definition of information overload to guarantee the homogeneity of interpretation. This section was composed of 42 Likert scale–like questions, comprising 2 overarching questions and 4 main aspects of information overload: (1) impact rating of individual information overload sources, (2) impact of information overload on specific work/life aspects, (3) preventative measures to manage information overload, and (4) useful sources of medical information. This part included statements such as "I feel that information overload significantly impacts on my work" and "E-mails I receive are a major contributing factor to my feelings of information overload" rated on a scale from 1 (completely disagree) to 7 (completely agree). There were also 7 open-ended questions to allow the respondents to elaborate on their opinions and perceptions. Most of the statements in the survey were extracted from a literature review conducted by Hall and Walton [8] and from a subsequent survey that one of the authors of this paper (GW) administered to hospital pharmacists and doctors in 2013 [45]. The remaining statements have been informed by 2 other authors (JW and JB) who are directly involved in emergency medicine and are aware of the current main sources of information overload in this sector. In total, the survey invite was sent to 263 physicians; 164 respondents participated with 101 recording full responses (38% response rate), which were then selected for analysis, as 63 records only contained partial demographic information and were deemed unusable.

Data Analysis

All data were saved in an IBM SPSS version 26 file for analysis. The dataset was initially inspected for errors and out-of-range values for each variable. The demographics were then produced to profile the sample. Next, descriptive statistics were performed for the Likert-style questions in the main section of the survey; means and standard deviations were calculated for each of the statements. A principal component analysis (PCA) was performed on the survey statements related to the sources of information overload. Finally, independent samples *t* tests and one-way between-group analyses of variance (ANOVAs) with Hochberg GT2 post hoc tests were conducted to compare mean scores in relation to respondents' demographics (ie, the current level of employment, age, and the number of years in emergency medicine). The analyses performed in terms of gender and NHS trust did not show any differences between groups; hence, they are not reported here.

In addition, the open-ended questions and closing remarks were manually coded independently by 2 members of the research team using the 6-step approach to thematic analysis by Braun and Clarke [46].

Results

Sample Demographic Profile

This section provides a summary of the demographic characteristics of the respondents. The sample shows a relatively balanced distribution between doctors working as junior trainees (31/101, 30.7%), senior trainees (28/101, 27.7%), and consultants (29/101, 28.7%), whereas 12.9% (13/101) reported other types of working levels (Table 1). In terms of gender, 42.6% (43/101) were male and 56.4% (57/101) were female (only 1/101, 1%, did not indicate their gender). As for age, the majority (70/101, 69.3%) were between 20 and 39 years, with 30.7% (31/101) between 40 and 59 years. In terms of the period since respondents became EMPs, two-third (68/101, 67.4%) had been emergency physicians for 10 years or less. Finally, across the 4 NHS Foundation Trusts surveyed, 39.6% (40/101) of respondents were based at Leeds Teaching Hospitals, followed by Bradford Teaching Hospitals (24/101, 23.7%) and Airedale (23/101, 22.8%), whereas 13.9% (14/101) were based on North Tees and Hartlepool.

XSL•FO
RenderX

Table 1. Demographic data from the respondents to the survey.

Demographic characteristics	Values, n (%)
Age (years)	
20-29	24 (23.8)
30-39	46 (45.5)
40-49	18 (17.8)
50-59	13 (12.9)
Gender	
Male	43 (42.6)
Female	57 (56.4)
Prefer not to say	1 (1.0)
How long have you been an emergency medicine physician for? (years)	
<1	21 (20.8)
1-4	25 (24.8)
5-10	22 (21.8)
11-15	16 (15.8)
16-20	9 (8.9)
>20	8 (7.9)
What is your current level?	
Junior trainee[a]	31 (30.7)
Senior trainee[b]	28 (27.7)
Consultant	29 (28.7)
Other[c]	13 (12.9)
What NHS[d] trust do you work for?	
North Tees and Hartlepool NHS Foundation Trust	14 (13.9)
Bradford Teaching Hospitals NHS Foundation Trust	24 (23.7)
Leeds Teaching Hospitals NHS Trust	40 (39.6)
Airedale NHS Foundation Trust	23 (22.8)

[a]Foundation year, VTS, and ACCS.

[b]ST4-ST7, staff grade, and junior clinical fellow.

[c]Including advanced clinical practitioner and associate specialist.

[d]NHS: National Health Service.

RO1 and RO2: Emergency Physicians' Views on Medical Information Overload

Table 2 summarizes the responses to the Likert scale statements listed in the second section of the survey in terms of mean and standard deviation. The table reports the 2 general questions asked at the beginning of the survey, and the remaining statements were divided across the 4 main information overload aspects covered in decreasing order of importance. The results suggest that there is overall agreement among emergency physicians that information overload has an impact on their work (mean 5.10, SD 1.25), and the problem has been increasing in time (mean 5.63, SD 1.34).

In terms of the sources that have the most impact on medical information overload, there is a wide range of mean values. Doctors listed the 24/7 culture (mean 5.40, SD 1.56), email handling (mean 4.86, SD 1.80), and, in equal measure, multidisciplinary communications (mean 4.51, SD 1.61) and local clinical guidelines (mean 4.50, SD 1.52). On the other hand, the information derived from patients (mean 3.29, SD 1.62), social media (mean 3.15, SD 1.88), and drug companies (mean 2.36, SD 1.50) does not pose particular concerns.

XSL•FO
RenderX

Table 2. Emergency clinicians' perspectives on information overload.

Statements[a]	VMean (SD)
To what extent does medical information overload impact your work?	5.10 (1.25)
To what extent do you agree that medical information overload has increased for you at work?	5.63 (1.34)
Impact of each of the following causes of medical information overload	
24/7 culture (ie, the *always available* workplace)	5.40 (1.56)
Email	4.86 (1.80)
Multidisciplinary communication	4.51 (1.61)
Local clinical guidelines	4.50 (1.52)
Info from employer/manager	4.43 (1.67)
Evidence-based practice	4.22 (1.57)
National clinical guidelines	4.21 (1.52)
Info to employer/manager (eg, as part of reporting activity)	4.14 (1.67)
Journal papers	3.41 (1.67)
Patients bringing information	3.29 (1.62)
Social media	3.15 (1.88)
Drug company information	2.36 (1.50)
How does medical information overload impact on the following aspects?	
Guideline fatigue	6.08 (1.56)
Stress and tension	5.25 (1.37)
Longer working hours	4.87 (1.70)
Impaired decision making	4.85 (1.57)
Imprecise clinical judgments	4.84 (1.45)
Tiredness/illness	4.81 (1.54)
Decrease in social life	4.26 (1.80)
What preventative measures do you adopt to combat information overload?	
Prioritization	6.07 (1.13)
Email handling skills (ie, ability to discriminate by importance)	5.39 (1.50)
Effective time management	5.30 (1.22)
Effective information management	5.23 (1.30)
Software/hardware	3.40 (1.81)
Information technology point of need delivery	3.13 (1.82)
Medical librarian	1.85 (1.32)
What sources do you find most useful when in need of medical information?	
Local clinical guidelines	6.13 (1.03)
National clinical guidelines	5.80 (1.09)
Colleagues	5.69 (1.34)
Evidence-based practice	5.14 (1.37)
Multidisciplinary communication	4.41 (1.49)
Journal papers	3.54 (1.39)
Patients bringing information	2.67 (1.56)
Social media	2.62 (1.65)
Medical librarian	2.60 (1.60)
Drug company information	2.17 (1.29)

XSL•FO
RenderX

[a]All statements were rated on a Likert scale from 1 to 7, and the values reported here are in decreasing order of importance.

Views on the consequences of information overload on different aspects of physicians' daily job and life show a more compacted set of values (less than 2 points difference between the top and bottom entries), indicating that all listed aspects are negatively impacted. Guideline fatigue is the leading consequence of information overload (mean 6.08, SD 1.56), followed by stress and tension (mean 5.25, SD 1.37). However, longer working hours, inaccurate decision making and clinical judgment, and chronic tiredness leading to illness also returned high mean values. Physicians also complained that information overload is detrimental to their personal and social lives (mean 4.26, SD 1.80).

RO3: Coping Mechanisms Adopted by Emergency Medicine Physicians to Address Information Overload

When it comes to the adoption of preventive measures to overcome information overload, prioritization is by far the most common approach (mean 6.07, SD 1.13), closely followed by email handling (mean 5.39, SD 1.50) and time (mean 5.30, SD 1.22) and information (mean 5.23, SD 1.30) management. Conversely, *external* support such as information technology (IT) and ICT and clinical librarians are not considered particularly effective.

In situations of actual need for medical information, both local (mean 6.13, SD 1.03) and national (mean 5.80, SD 1.09) clinical guidelines are the 2 primary sources, which interestingly highlight the struggle doctors regularly experience. On the one hand, such guidelines are crucial to maintaining a critical level of medical knowledge, but, on the other hand, because of their sheer volume and constant updating, they represent the main cause of medical information overload. In this context, colleagues are also an important source of easily accessible and reliable medical information while on the ward (mean 5.69, SD 1.34). Patient information (mean 2.67, SD 1.56), social media (mean 2.62, SD 1.65), and drug companies (mean 2.17, SD 1.29) are considered irrelevant sources because of their unqualified or biased nature. However, medical librarians have also scored a low mean value (mean 2.60, SD 1.60), perhaps because their services were not considered to be of immediate relevance to emergency physicians' very fast case turnover.

PCA with direct oblimin rotation with Kaiser normalization was performed on the 12 statements addressing the main causes of information overload. The Kaiser-Meyer-Olkin value was 0.807, exceeding the recommended value of 0.6 [47], and the Bartlett test of sphericity reached statistical significance (P<.001), supporting the factorability of the correlation matrix. The analysis identified the presence of 3 factors explaining 42.2%, 12.2%, and 9.6% of the variance, respectively. Table 3 reports the factors and Cronbach α values calculated to assess the reliability of the scale.

Table 3. Exploratory factor analysis of factors relative to causes of information overload.

Factors/statements	Cronbach α
Factor 1: the dutiful physician	.800
24/7 *always available* culture	
Emails	
Information from employer/manager	
Information to employer/manager	
Multidisciplinary communications	
Factor 2: the informed physician	.842
Local clinical guidelines	
National clinical guidelines	
Evidence-based practice	
Factor 3: the current physician	.785
Journal papers	
Social media	

The first factor, named *the dutiful physician*, includes the potential sources of information overload that cannot be ignored or overseen as they form part of physicians' daily administrative tasks that must be undertaken in addition to direct patient care. These tasks are demanding and time-consuming. Regular reporting to and from line managers or employers and multidisciplinary communications are crucial to the smooth running of emergency departments and provide a more coordinated approach to patient care discharge planning [48]. Most of such communications are exchanged via email, which becomes the main conduit of information overload, and are carried out at all times, hence exacerbating further the *always available* culture.

The second factor, named *the informed physician*, relates to physicians' need to stay up to date with the changes in guidelines, both at the local and national level and evidence-based practice on how health care professionals should care for patients with specific conditions, and which are based on best evidence practice. These guidelines help the standardization of care across NHS trusts and can be modified

XSL•FO
RenderX

to adjust to local needs to provide a cost-effective approach to care [49].

The third factor, named *the current physician*, includes additional sources of information overload that physicians experience on a regular basis in the form of notifications and which are mostly derived from scientific publications (eg, notifications of new articles, subscription invites, and electronic book updates) and social media. The respondents' comments on this aspect have shown how there appears to be an expectation that physicians, trainees in particular, cluster in online study groups and follow relevant Twitter handles such as #MedEd or #FOMed.

RO4: Information Overload Perceptions by Age and Years Spent in Emergency Medicine

Independent sample *t* tests have been undertaken on the bank of Likert scale statements in terms of age of respondents and their experience in emergency medicine. All statistically significant results are reported in Tables 4 and 5. In terms of age (Table 4), younger physicians appeared less worried about their reporting activities (mean 3.89, SD 1.02) than older ones (mean 4.65, SD 1.11). Conversely, they are more affected by the loss in social activities because of information overload (mean 4.52, SD 1.31) as opposed to more experienced physicians (mean 3.68, SD 1.05). In addition, younger physicians reported a higher rate of impaired decision making (mean 5.19, SD 1.40) and imprecise clinical judgments (mean 5.12, SD 1.18) than their older colleagues (mean 4.07, SD 1.14 and mean 4.23, SD 1.37), who are still concerned about their performance but have greater experience to draw upon. Younger physicians also seem to rely more on local clinical guidelines (mean 6.28, SD 1.08) than older ones (mean 5.81, SD 1.23) to reduce decision fatigue.

Table 4. Independent samples *t* tests on the age of respondents.

Questions	*t* test (*df*)	*P* value	Mean difference	Age (20-39 years), mean (SD)	Age (40-59 years), mean (SD)	Total mean
Impact of info to employer/manager (eg, as part of reporting activity) on medical information overload	−2.109 (94)	.04	0.75	3.89 (1.02)	4.65 (1.11)	4.14
How does medical information overload impact on decrease in social life?	2.098 (51.011)	.04	0.84	4.52 (1.31)	3.68 (1.05)	4.26
How does medical information overload impact on impaired decision making?	3.304 (50.699)	.002	1.12	5.19 (1.40)	4.07 (1.14)	4.85
How does medical information overload impact on imprecise clinical judgments?	2.762 (49.767)	.008	0.89	5.12 (1.18)	4.23 (1.37)	4.84
What preventative measures do you adopt to combat information overload? Email handling skills (ie, ability to discriminate by importance)	−2.055 (70.489)	.04	0.61	5.19 (1.22)	5.81 (1.18)	5.39
What sources do you find most useful when in need of medical information? Local clinical guidelines	2.198 (60.325)	.03	0.47	6.28 (1.08)	5.81 (1.23)	6.13

In terms of years of experience in emergency medicine (Table 5), the only dissimilarity seems to relate to the impact and handling of work emails. Physicians with more than 10 years of experience identify the number of emails received as one of the main causes of information overload (mean 5.39, SD 1.19) as opposed to physicians with less experience who scored a mean of 4.60 (SD 1.26) and hence attributed more importance to email handling skills as a way to manage the problem (mean of 6.03 and SD 1.15 against a mean of 5.06 and SD 1.13 reported for less experienced physicians).

XSL•FO
RenderX

Table 5. Independent samples *t* tests on the years spent in emergency medicine.

Questions	*t* test (*df*)	*P* value	Mean difference	Years of experience (0-10 years), mean (SD)	Years of experience (more than 10 years), mean (SD)	Total mean
Impact of emails on information overload	−2.134 (65.084)	.04	0.79	4.60 (1.26)	5.39 (1.19)	4.86
What preventative measures do you adopt to combat information overload? Email handling skills (ie, ability to discriminate by importance)	−3.166 (96)	.002	0.97	5.06 (1.13)	6.03 (1.15)	5.39

RO4: Information Overload Perceptions by Level of Employment

To conclude the quantitative assessment of the data collected, one-way between-group ANOVAs by the level of employment were also performed on the survey statements (Table 6). Across the 3 main levels populating emergency departments, senior trainees and consultants are those acknowledging the increased information overload with time the most, possibly because of their longer service in hospitals. However, in relation to other aspects, senior trainees seem to experience the most difficulties. For example, in terms of the aspects of their work and life impacted by information overload, they complain about a decrease in social life and diminished decision-making abilities.

Table 6. One-way between-group analysis of variance on current level of employment.

Questions	*F* test	*P* value	Junior trainee, mean (SD)	Senior trainee, mean (SD)	Consultant, mean (SD)	Total mean
To what extent do you agree that medical information overload has increased for you at work?	5.798	.004	5.07[a,b] (1.01)	6.07[a] (1.02)	6.03[b] (1.22)	5.63
How does medical information overload impact on decrease in social life?	3.249	.04	4.50 (1.27)	4.79[c] (1.06)	3.72[c] (1.36)	4.26
How does medical information overload impact on impaired decision making?	4.331	.02	5.17 (1.14)	5.36[c] (1.14)	4.29[c] (1.40)	4.85
What preventative measures do you adopt to combat information overload? Email handling skills (ie, ability to discriminate by importance)	9.223	<.001	5.47 (1.16)	4.67[c] (1.18)	6.10[c] (1.01)	5.39
What sources do you find most useful when in need of medical information? Social media	9.246	<.001	2.03[a] (1.21)	3.59[a,c] (1.02)	2.23[c] (1.19)	2.62

[a]Junior trainee versus senior trainee.

[b]Junior trainee versus consultant.

[c]Senior trainee versus consultant.

In addition, whereas for consultants, the ability to effectively manage work emails is crucial to handle, at least partially, information overload (mean 6.10, SD 1.01), senior trainees are the least concerned about this aspect (mean 4.67, SD 1.18). Finally, although the use of social media as a source of medical information is not a very popular choice among emergency physicians in general, senior trainees use them significantly more than the other 2 groups, with a mean of 3.59 (SD 1.02), as opposed to 2.03 (SD 1.21) for junior trainees and 2.23 (SD 1.19) for consultants.

Qualitative Analysis of Open-Ended Questions

The survey contained several open-ended questions and, although they did not form part of the compulsory sections, still attracted a considerable number of spontaneous and insightful comments.

XSL•FO
RenderX

In terms of the length of time physicians have been experiencing information overload, 39.6% (40/101) claimed that the phenomenon had increased in the last 5/6 years; 10.9% (11/101) confirmed that it started before that, about 10 to 15 years ago; and 14.9% (15/101) did not specify a date but listed aspects such as progressing in their career, noting how the pressure shifted from night shifts only to an all-time problem and overload worsening once the email was firmly established within the NHS. Around 3.0% (3/101) felt that they did not have enough experience to judge the problem, and the remaining respondents did not comment.

The remaining comments revolved around 2 main themes (Table 7), specifically, the ways in which medical information overload impacts physicians' work and various preventative measures to overcome information overload. The 5 subthemes identified within each of the 2 main themes are discussed below.

Table 7. Overview of the themes and subthemes emerging from open-ended questions.

Theme and subthemes	Strands
Ways in which medical information overload impacts on clinicians' working life	
1. Positive impact	
	Reverse views
2. Direct causes leading to reduced patient care	
	Impact on patients
	Distraction and forgetfulness
	Confusion
	Delayed decision making
	Tiredness, stress, and anxiety
3. Indirect causes leading to reduced patient care	
	Lack of time
	Email handling
	Information and communication technology systems
	Breadth of information
	Guidelines and evidence-based practice
	Ability to stay up to date
	Seniority
Preventative measures to combat information overload	
4. Personal attitudes	
	Avoidance
	Therapy
	Work-life balance
5. Workload management	
	Email and social media management
	Task management
	Time management
	Prioritization
	Delegation

RO5: Ways in Which Medical Information Overload Impacts Physicians' Working Life

1. Positive Impact

A total of 3 respondents, all of them consultants, admitted that information overload was indeed present in their working life but did not represent a hindrance to their performance (*reverse views* subtheme):

> *It doesn't. You do have be to discerning about where you get information from and use trusted sources though.*

XSL•FO
RenderX

> *Increase in research is interesting and gives more confidence in decision-making. I find this is helpful rather than a hindrance.*
>
> *The more information I have the better I can do my JOB!*

The majority of respondents however did not agree with this view and listed a series of impeding aspects, which can be divided into direct and indirect causes leading to a reduction in patient care.

2. Direct Causes Leading to Reduced Patient Care

Among the direct causes, concerns for the well-being and quality of care of patients (*impact on patients* strand) were the most mentioned and most concerning for the physicians:

> *On a busy shift on the shop floor when there are competing priorities for my attention, increasingly feel that this means if I try to see my own patients then I either give them poor care or lose control of the department.*
>
> *Slow decision-making, increased risk of errors.*

This continuous feeling of apprehension and an impending sense of loss of control are accompanied by *distraction and forgetfulness*, which was the second most reported issue (strand) caused by information overload and can be summarized in the quote below:

> *I have a significantly higher background level of anxiety at work due to concerns about giving poor advice or forgetting something important due to the volume of information I am trying to process. I find the non-shop floor information much less of a problem.*

Other strands, such as *confusion* with the information received, *delayed decision making* and *tiredness, stress, and anxiety* were mentioned in equal measure as contributing to a potentially reduced patient's quality of care:

> *Toward the end of the shift my answers take longer to arrive at, I feel like I am even more risk averse and at times feel like I am can become short with staff asking for help.*
>
> *During busy environments if documents/protocols contain too much information I may struggle to pick up salient points.*

3. Indirect Causes Leading to Reduced Patient Care

Among the indirect causes of information overload, leading to a reduced quality of patient care, *lack of time* to review the information properly was reported as one of the most serious problems:

> *Huge amount of information coming in and little time to process and implement any actions needed.*
>
> *Not enough time to document all relevant information consistently.*

Email handling and problems with *ICT systems* were reported as common, overpowering, and frustrating:

> *Email mainly - almost overwhelming.*
>
> *Vast numbers of irrelevant or barely relevant emails containing information which will be forgotten immediately as it concerns rare circumstances or conditions never before experienced (and therefore poorly understood).*
>
> *Multiple IT systems with different passwords and access. Constant emails - usually not relevant. Feel overwhelmed by the day to day IT so I don't want to access online resources to broaden my knowledge.*

Moreover, 3 additional strands, namely *guidelines and evidence-based medicine*, *ability to stay up-to-date*, and *breadth of information*, are related to the need to be informed that physicians experience on a daily basis [50] and are difficult to address in an environment overloaded with what seems like irrelevant information:

> *I feel I am unable to keep up to date with the multitude of clinical guidelines from NICE, the college...only exacerbate by looking at Tweeter!*
>
> *The number of guidelines/policies that now exist both locally and nationally - it is hard to keep track of all of these.*

Finally, *seniority* was also mentioned as a hindering factor contributing to information overload, as more junior staff and nurses (and patients) tend to direct their queries and doubts to the more experienced physicians:

> *More difficult out of hours/weekends when I am the only senior decision-maker in the department. If working with a good registrar this can lessen the impact but if working with a less experienced middle grade all questions/information comes though me.*
>
> *Questions from juniors and other colleagues.*
>
> *Lots of questions asked throughout shift from nurses to other doctors.*

Preventative Measures to Combat Information Overload

When it comes to addressing and solving the problem of information overload in effective ways, respondents suggested solutions that can be classified into 2 subthemes: *personal attitudes* and *workload management strategies*.

4. Personal Attitudes

As part of the individual attitudes observed, some physicians simply ignored the problem (*avoidance* strand). Some other physicians adopted a very pragmatic approach and had no difficulties recognizing their own limits:

> *When I am at work and feeling overloaded, I step away from the situation.*
>
> *Learn to ignore.*
>
> *Acceptance that you can't do everything.*

Psychological stress is increasingly recognized within the medical profession, particularly in emergency medicine [51], and some respondents prefer to rely on formal support in the form of *therapy*:

XSL•FO
RenderX

> *Therapy - I have had CBT sessions and NLP sessions to help me deal with the stress and anxiety information overload gives me and to gain strategies regarding how to cope with this.*
>
> *I take little steps and with the help of my therapist I try not to compare myself negatively to others.*

Others preferred to self-manage and find a more balanced relationship between working and personal life (*work-life balance* strand), although often long shifts and complex working patterns make this option fairly difficult to attain:

> *My days off are for fun and time away - I do not work on days off. I avoid seeing colleagues on these days who talk about work when they are not there.*
>
> *I feel a lot of the 'critical information' I need for my day to day job is limited - especially while trying to revise for exams, the need to switch off and rest is very important.*

5. Workload Management

Physicians have also indicated a number of strategies to manage information overload that are more strictly linked to practical workload management and dealing effectively with emails and social media (*email and social media management* strand). This appeared to be a priority for many respondents:

> *When I am not at work or feeling overloaded I ignore my e-mails and mute my social media accounts.*
>
> *Nil work emails on phone mute work WhatsApp groups when on leave no longer active on Facebook at all due to general overload but also work based groups.*

This was followed by strategies to address *task and time management* in an effective and sustainable way:

> *Keeping a to-do- list booklet.*
>
> *Summarising things. Writing things down properly helps consolidate my head.*

Finally, *prioritization* and *delegation* were also mentioned as effective ways of managing information overload, particularly at times of crisis:

> *Prioritisation and having some "IT free" time is vital.*
>
> *Delegation of clinical supervision.*

Discussion

RO1: Determine the Extent to Which Emergency Medicine Physicians Experience Medical Information Overload

Both quantitative and qualitative data have shown that emergency physicians are deeply affected by an ever-increasing amount of medical information impacting every aspect of their job. The 24/7 (or *always available*) culture places itself firmly at the heart of the problem, as identified by Walton [45], and leads to problems with fatigue and, consequently, with performance, productivity, and safe decision making [52]. All these aspects have been highlighted by physicians in this study as being compromised not just by the very fast-moving pace typical of their job, but by the constant and relentless information received and the information requested from them. What emerged very clearly from this study is the difficulty in striking a balance between the need to maintain an acceptable and current level of knowledge and the ability to sift and discriminate through large amounts of irrelevant information. There are valid arguments supporting both the benefits of information overload and its drawbacks. A small outlier group of respondents composed of 3 consultants claimed how there was no such thing as information overload as a negative attribute of their job. They see information as an opportunity to learn, as observed by Falschlunger [21], and increase their confidence in their decisions, in line with the *freedom of choice* described by Brehm and Brehm [53] and De Angelis [54]. However, the majority of physicians perceived information overload as a negative concept hindering their ability to process information effectively and accurately [55]. The migration toward *big data* handling by health care professionals because of the information expansion phenomenon described by Yamamoto [30] is also reflected in this study as physicians are faced with multiple evolving environments. The increasing number of emails and social media notifications received daily and regular updates in policies and evidence-based practice, coupled with new or updated IT systems to keep abreast of, will all contribute to feelings of inadequacy and frustration. Unless controlled and harnessed, such feelings, in the long term, can lead to *burnout syndrome*, which is very common among all types of health care professionals [43,44]. It is difficult to establish when physicians have started experiencing the burden of information overload in their careers. It appears that consultants and people working in emergency departments the longest have noticed a steady increase, made sharper since the ready access to internet-based resources and emails started over 20 years ago. This is in line with findings by Roetzel [56], who reported how the research around the 1980s and, particularly, the 1990s coincided with the rise and spread of information technologies. However, for most, the problem started to take shape with the advent of mobile platforms and the deployment of electronic health record systems in NHS trusts in 2005 [57].

RO2: Establish the Contributory Factors to Information Overload and Evaluate Their Impact on Physicians' Working Life

The leading causes of information overload in emergency physicians are multiple, and it is often difficult to discriminate among them because of their interlinked nature. The PCA conducted on the survey statements revealed 3 main factors contributing to information overload. The main factor leading to the concept of *the dutiful physician* is the administrative tasks physicians must carry out as part of their daily routine, such as reporting, responding to emails, and communicating with the other physicians forming a patient's care team. This study has reported an increased burden because of email handling compared with Walton's [45] research on hospital doctors, leading to the conclusion that, in the last 6 years, since that study was published, the number of emails has significantly increased as part of physicians' daily duties. It is also clear that administrative tasks are one of the main causes of physicians' burnout syndrome [58] because of their persistent and repetitive

XSL•FO
RenderX

nature. However, physicians are also aware of the importance of carrying out such tasks for the smooth running of the ward and the good care of patients [26]. A second set of causes deeply affecting physicians' perceptions of information overload is linked to the consultation of local and national clinical guidelines. Ironically, such guidelines, as important as they are to guarantee patients' consistency of care, are also a burden to health care professionals because of their frequent updates and notifications. On the one hand, physicians must remain current (ie, *the informed physician*) and follow the policies and directives reported in the guidelines to guarantee the quality of patient care, but on the other hand, the pressure of being up to date and making sure nothing is overlooked plays a considerable role in increasing physicians' stress [59]. A third factor contributing to information overload entails the number of notifications and alerts received mostly via social media, but also from journal editors and publishers and regarding new medical approaches, clinical trials, discoveries, and case studies (*the current physician*). These sources have registered a significant increase since Walton's study [45] in 2014, likely because of the current much wider diffusion of social media platforms in all aspects of society. Once again, physicians experience an internal dispute about when to start ignoring such alerts. Maggio and Artino [59] believe that medical librarians could help battle physicians' information overload and help them navigate through the various systems and social media, but they are not utilized as much as they should, a conclusion corroborated by this study, where librarians score among the least useful sources of medical information (Table 2). However, regardless of the causes of information overload, the consequences can be summarized as directly or indirectly leading to reduced patient care, which will be discussed further below.

RO3: Identify the Coping Mechanisms Adopted by Emergency Medicine Physicians to Respond to Information Overload

To help mitigate information overload, physicians have devised strategies for managing and filtering available medical information. The results of this study have shown similar results and led to the conclusion that information overload prevention and mitigation are left entirely in the hands of individuals and their personal views and preferences. A number of actions are taken from a pragmatic point of view (workload management) quite regularly, but with varying degrees of success. Among the most common approaches, email and social media handling skills are highly desirable, considering the steadily growing amount of communication, including patient information, passing through these tools [60]. Strictly linked to this is also the ability to prioritize tasks (including emails and social media), appearing very high in the list of possible solutions, particularly in situations of multiple concurrent demands [33], which are common practices in emergency departments. Prioritization is a skill developed with time and experience [61], in situations of high demand and variable workload [34], and is particularly popular in emergency departments. Task, time, and information management are other measures frequently adopted by physicians to overcome medical information overload and can take the form of simple actions, such as keeping daily and weekly to-do lists, maintaining written records, timing actions, summarizing information, or key articles to help consolidate knowledge. Although most of the above points were identified by Walton [45], in this study, their importance is stressed further, and new approaches have emerged, including those dictated by personal attitudes. These play an important role in determining how people deal with information overload. This study has highlighted a few approaches, although not all adopted with the same conviction. A small group of respondents admitted that the easiest way is to ignore circumstances or demands when feeling overloaded, which is not necessarily a positive action to undertake, particularly in emergency care. This is symptomatic of the difficulty and pressure physicians are subject to and is a common self-preservation mechanism triggered by overwhelming situations [62]. Other people have mentioned formal therapy and a better work-life balance in which working and private life are kept separate, and personal time is seen as an opportunity to regain strength and stability. These approaches are not new, nor specific to emergency physicians, as they are common to most workplaces [11,19], but it would be advisable to integrate them with formal training on the job and regular sessions on developing coping skills.

RO4: Explore Whether Emergency Medicine Physicians at Different Career Stages Respond to Information Overload in Different Ways

This part of the study offers a new viewpoint of medical information overload drawn from physicians' experiences at different times in their careers and highlights deeply diverse perceptions and needs.

The study revealed an overall homogeneity of perceptions of medical information overload, in the sense that most respondents agreed on its increasing presence, size, and impact. However, some specific aspects showed differences in responses depending on the seniority of the physician. For example, people who have been in emergency care for 10 years or longer appeared to resent the increased number of emails received and attributed more importance to the ability to handle them than people who have been working in emergency care for less than ten years. This is also in line with the independent sample *t* tests conducted on the age of the respondents and with the ANOVA tests conducted on the level of employment. However, younger physicians and trainees appeared to be more worried about the loss in social life, impaired decision making, and imprecise clinical judgment than older colleagues and consultants. The literature demonstrates that junior clinicians have higher levels of anxiety because of uncertainty than senior clinicians [63,64], and this is reproduced in this study because of the additional stress of information overload. In particular, it appears that senior trainees are the most concerned about this problem and are the most affected by it, particularly with respect to impaired decision making. A recent study on Canadian emergency medicine residents [65] revealed how junior trainees are properly supervised and supported, but when they then become senior trainees, they lose most of that support and are left to fend for themselves. In addition, senior trainees are at a very demanding stage in their career, where they are preparing for independent practice [66] and are entrusted with more complex decisions

XSL•FO
RenderX

than ever before. This is bound to contribute to an increased sense of unease and frustration when it comes to information overload.

RO5: Determining the Impact of Physicians' Information Overload on Patient Care

Compromising the quality of patient care represents the ultimate consequence of medical information overload and rests at the forefront of physicians' thoughts. The qualitative responses in this study reported on the physicians' perceptions of the problem and showed how slow decision making because of information overload and frequent interruptions can lead to substandard care and errors in diagnoses and prescriptions. In addition, remembering patient diagnostics and medical history becomes difficult throughout a shift because of the large amount of information acquired about other cases and other confounding factors [28]. The fast turnover of emergency departments implies that there can often be a queue of people waiting to speak to a physician, who frequently does not manage to action one of these requests before others come in, hence compromising patient safety [32]. Among the effects of information overload mentioned by the respondents, distraction and tiredness play a key role. Physicians experience a higher background level of anxiety at work because of concerns about giving poor advice or forgetting something important because of the volume of information they are trying to process. This is exacerbated by long shifts, which can leave physicians both mentally and physically drained [43,44]. In critical situations, finding the right information fast is essential, but information overload can prevent this. Some respondents have been struggling with the correct information they should follow, as sometimes a lot of vague or unclear guidelines or advice can be worse than none at all, whereas others complained that documents and protocols contained so much information that it was difficult to discern salient points. In addition to this, other less obvious causes can still lead to poor patient care, particularly the ubiquitous nature of digital technologies, which provoke conflicting feelings among health care professionals. They are perceived as intrusive and time-consuming, but also as an important resource of clinical information, particularly when dealing with emergencies [31]. A recent study by Laker et al [28] introduced the concept of *emphasis framing* to aid health care professionals toward a more focused approach to decision making. This cognitive process could help in identifying the most relevant components of any piece of information while discarding the rest. For example, physicians could be directed toward article abstracts rather than full papers or be alerted to an incorrectly entered value for a medication prescription. The study reported an increase in clinical decision-making time through this approach, but also a significantly improved quality of medical decisions. Although a lot of work needs to be done to properly assess the real benefits of emphasis framing and how to implement it in real-world workflows, it is a research direction worth exploring further.

Conclusions

This study reported on the high and ever-increasing medical information overload in emergency medicine departments in the UK NHS trusts. Among the many factors listed by physicians as influencing information overload, the historical emergency medicine culture of being constantly available throughout a shift (24/7 culture), email, and multidisciplinary communications have been mentioned as being the most severe. However, different levels of employment, experience in emergency medicine, and age also seem to influence physicians' perceptions of the leading causes of information overload. Nevertheless, some information is critical to guarantee high standards of patient care, particularly those derived from clinical guidelines.

This study advances previous research on information overload in 4 distinct ways: (1) by identifying the current causes affecting UK emergency medicine while also recognizing key information sources; (2) by offering insights on how information overload affects physicians in relation to demographic factors such as age, years spent in emergency medicine, and level of employment (ie, seniority); (3) by identifying mechanisms that emergency physicians put in place to manage information overload; and finally (4), by categorizing influencing factors in terms of their impact on patient care and reports on preventative measures currently in use to combat information overload.

However, this study is also subject to the following limitations: (1) although survey is an effective method to receive responses from a large number of participants and to explore relationships between variables, it might be less effective to answer complex questions such as user perceptions of information overload, contributing factors, and coping mechanisms, particularly when limited previous studies were conducted on this topic and with this particular user group; (2) the sample size is limited, although studies conducted on health care professionals are usually affected by this problem; (3) some of the variables listed as potential sources of information overload can be difficult to uniquely discriminate from each other (eg, emails may include clinical guidelines, employer information, and patient information), but the survey wanted to offer both information content aspects and information source channels; and (4) the study was only able to assess physicians' *perceptions* on the impact that information overload has on the quality of patient care. To attain an objective measure of such impact, different indicators would have been needed to allow for a direct correlation between patients' health outcomes and physicians' information overload.

This study has identified further research areas that could be worth taking forward to provide a more in-depth and more informed insight into information overload. The 24/7 culture, multidisciplinary interactions, and the pressure of following relevant guidelines are aspects specific to the emergency medicine profession and must be addressed to relieve physicians. Further investigation could determine the influence of professional boundaries on such aspects. A consistent thread in the findings was the contribution made by email in increasing information overload. This was an interesting outcome where it could be assumed that social media as a communication channel was taking the place of email. This would point to the value of a more detailed exploration, specifically on email and information overload. An area of concern that this study established was the interrelationship between clinical guidelines and information overload. By establishing how clinical guidelines make this contribution, it is possible to determine

XSL•FO
RenderX

how presentation, style, and format could be adapted to make them more accessible.

NHS trusts in the United Kingdom should consider some of the strategies adopted by the physicians participating in this study to reduce the burden of information overload (eg, prioritization) and try to embed routine training or guidance on time and information management skills in staff development programs.

Conflicts of Interest

None declared.

References

1. Blair A. Too Much To Know: Managing Scholarly Information Before the Modern Age. Yale: Yale University Press; 2010.
2. Horrigan J. Information Overload. Pew Research Center. 2016. URL: https://www.pewresearch.org/internet/2016/12/07/information-overload/ [accessed 2020-03-11]
3. Information Recall: The Secret Weapon You Didn't Know You Needed. TeleWare. URL: https://www.teleware.com/wp-content/uploads/2018/04/TeleWare_ReCall_White_Paper.pdf [accessed 2020-02-03]
4. Dadashi N, Golightly D, Sharples S. Seeing the woods for the trees: the problem of information inefficiency and information overload on operator performance. Cogn Tech Work 2017 Nov 23;19(4):561-570 [FREE Full text] [doi: 10.1007/s10111-017-0451-1]
5. Floridi L. The Fourth Revolution: How the Infosphere is Reshaping Human Reality. Oxford, UK: Oxford University Press; 2014.
6. Toffler A. Future Shock. New York, USA: Bantam Books; 1970.
7. Schick A, Gordon L, Haka S. Information overload: a temporal approach. Account Organ Soc 1990 Jan 1;15(3):199-220 [FREE Full text] [doi: 10.1016/0361-3682(90)90005-f]
8. Hall A, Walton G. Information overload within the health care system: a literature review. Health Info Libr J 2004 Jun;21(2):102-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2004.00506.x] [Medline: 15191601]
9. Benselin J, Ragsdell G. Information overload: the differences that age makes. J Libr Inf Sci 2016 Jul 8;48(3):284-297 [FREE Full text] [doi: 10.1177/0961000614566341]
10. Jackson T, Farzaneh P. Theory-based model of factors affecting information overload. J Glob Inf Manag 2012 Dec;32(6):523-532 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2012.04.006]
11. Joining the Dots: Decision Making for a New Era. Chartered Global Management Accountant. 2016. URL: https://www.cgma.org/content/dam/cgma/resources/downloadabledocuments/joining-the-dots-report.pdf [accessed 2020-03-11]
12. Abdullah N, Mustapar N. Heart rate reading patterns of moment of information overload during online information searching. J ICT 2019 Jul;18(3):233-251 [FREE Full text] [doi: 10.32890/jict2019.18.3.1]
13. Borkovich D, Middle G. Information overload revisited: infinite organizational threats. Issues Info Sys 2018;19(4):150-161 [FREE Full text]
14. Hwang M, Hong J, Tai K, Chen J, Gouldthorp T. The relationship between the online social anxiety, perceived information overload and fatigue, and job engagement of civil servant LINE users. Gov Inf Q 2020 Jan;37(1):101423 [FREE Full text] [doi: 10.1016/j.giq.2019.101423]
15. Whelan E, Islam A, Brooks S. Applying the SOBC paradigm to explain how social media overload affects academic performance. Comput Educ 2020 Jan;143:103692 [FREE Full text] [doi: 10.1016/j.compedu.2019.103692]
16. Yin P, Ou CX, Davison RM, Wu J. Coping with mobile technology overload in the workplace. Internet Res 2018 Oct 2;28(5):1189-1212. [doi: 10.1108/intr-01-2017-0016]
17. Joyce S, Shand F, Bryant RA, Lal TJ, Harvey SB. Mindfulness-based resilience training in the workplace: pilot study of the internet-based resilience at work (RAW) mindfulness program. J Med Internet Res 2018 Sep 11;20(9):e10326 [FREE Full text] [doi: 10.2196/10326] [Medline: 30206055]
18. Szóstek A. 'Dealing with my emails: latent user needs in email management. Comput Hum Behav 2011 Mar 1;27(2):723-729 [FREE Full text] [doi: 10.1016/j.chb.2010.09.019]
19. Jackson T, Burgess A, Edwards J. A Simple Approach to Improving Email Communication: Going Back to Basics. CORE – Aggregating the World's Open Access Research Papers. 2006. URL: https://dl.acm.org/doi/pdf/10.1145/1132469.1132493 [accessed 2020-03-11]
20. Kalman Y. Why do we blame information for our overload? In: Work Pressures. New York, USA: Routledge; 2016.
21. Falschlunger L, Lehner O, Treiblmaier H. InfoVis: The Impact of Information Overload on Decision Making Outcome in High Complexity Settings. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, USA URL: https://aisel.aisnet.org/sighci2016/3
22. Kumar A, Faiq M, Pandey S, Pareek V, Mochan S, Kumar P, et al. Addictive influences and stress propensity in heavy internet users: a proposition for information overload mediated neuropsychiatric dysfunction. Curr Psychiatr Rev 2018 Feb 20;13(4):293-300 [FREE Full text] [doi: 10.2174/1573400513666170728155836]

XSL•FO
RenderX

23. Melby L, Toussaint P. Supporting Operating Nurses' Collaborative Work: Preventing Information Overload and Tailoring Information Access. In: Proceedings of the IEEE Symposium on Computer-Based Medical Systems. 2009 Presented at: CBMS'09; August 3-4, 2009; Albuquerque, New Mexico, USA URL: https://doi.org/10.1109/CBMS.2009.5255364 [doi: 10.1109/cbms.2009.5255364]
24. Wisner K, Lyndon A, Chesla CA. The electronic health record's impact on nurses' cognitive work: an integrative review. Int J Nurs Stud 2019 Jun;94:74-84. [doi: 10.1016/j.ijnurstu.2019.03.003] [Medline: 30939418]
25. Marbouti M, Maurer F. Social Media Use During Emergency Response – Insights from Emergency Professionals. In: Proceedings of the Conference on e-Business, e-Services and e-Society. 2016 Presented at: I3E'16; September 13-15, 2016; Swansea, United Kingdom URL: https://doi.org/10.1007/978-3-319-45234-0_49 [doi: 10.1007/978-3-319-45234-0_49]
26. Walter SR, Raban MZ, Dunsmuir WT, Douglas HE, Westbrook JI. Emergency doctors' strategies to manage competing workload demands in an interruptive environment: an observational workflow time study. Appl Ergon 2017 Jan;58:454-460. [doi: 10.1016/j.apergo.2016.07.020] [Medline: 27633242]
27. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Z Evid Fortbild Qual Gesundhwes 2015;109(4-5):285-290. [doi: 10.1016/j.zefq.2015.06.005] [Medline: 26354128]
28. Laker L, Froehle C, Windeler J, Lindsell C. Quality and efficiency of the clinical decision-making process: information overload and emphasis framing. Prod Oper Manag 2017 Oct 2;27(12):2213-2225 [FREE Full text] [doi: 10.1111/poms.12777]
29. Walter S, Dunsmuir W, Westbrook J. Studying interruptions and multitasking in situ: the untapped potential of quantitative observational studies. Int J Hum Comput Stud 2015 Jul 1;79:118-125 [FREE Full text] [doi: 10.1016/j.ijhcs.2015.01.008]
30. Yamamoto Y. Healthcare and the roles of the medical profession in the big data era. Japan Med Assoc J 2016 Sep;59(2-3):125-139 [FREE Full text] [Medline: 28299246]
31. Fernandes CO, Miles S, Lucena CJ, Cowan D. Artificial intelligence technologies for coping with alarm fatigue in hospital environments because of sensory overload: algorithm development and validation. J Med Internet Res 2019 Nov 26;21(11):e15406 [FREE Full text] [doi: 10.2196/15406] [Medline: 31769762]
32. Connell A, Black G, Montgomery H, Martin P, Nightingale C, King D, et al. Implementation of a digitally enabled care pathway (part 2): qualitative analysis of experiences of health care professionals. J Med Internet Res 2019 Jul 15;21(7):e13143 [FREE Full text] [doi: 10.2196/13143] [Medline: 31368443]
33. Diniz V, Borges M, Gomes J, Canós J. Decision making support in emergency response. In: Encyclopedia of Decision Making and Decision Support Technologies. New York, USA: IGI Global; 2008:184-191.
34. Westbrook JI, Raban MZ, Walter SR, Douglas H. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. BMJ Qual Saf 2018 Aug;27(8):655-663 [FREE Full text] [doi: 10.1136/bmjqs-2017-007333] [Medline: 29317463]
35. Levin S, France D, Hemphill R, Jones I, Chen K, Rickard D, et al. Tracking workload in the emergency department. Hum Factors 2006;48(3):526-539. [doi: 10.1518/001872006778606903] [Medline: 17063967]
36. Liadova A, Korkiya E, Mamedov A, Panich N. The burnout among emergency physiciansvidence from Russia (sociological study). Man India 2017;97(15):495-507 [FREE Full text]
37. Schooley B, Hikmet N, Tarcan M, Yorgancioglu G. Comparing burnout across emergency physicians, nurses, technicians, and health information technicians working for the same organization. Medicine 2016 Mar;95(10):e2856 [FREE Full text] [doi: 10.1097/MD.0000000000002856] [Medline: 26962780]
38. Ellahi A, Mushtaq R. Doctors at risk of job burnout, diminishing performance and smoking habits. J BSA Med Soc Group 2012;6(3):36-47.
39. Burn-out an 'Occupational Phenomenon': International Classification of Diseases. World Health Organization. 2019. URL: https://www.who.int/mental_health/evidence/burn-out/en/ [accessed 2020-03-09]
40. Shanafelt TD, Mungo M, Schmitgen J, Storz KA, Reeves D, Hayes SN, et al. Longitudinal study evaluating the association between physician burnout and changes in professional work effort. Mayo Clin Proc 2016 Apr;91(4):422-431. [doi: 10.1016/j.mayocp.2016.02.001] [Medline: 27046522]
41. Lemaire JB, Wallace JE. Burnout among doctors. Br Med J 2017 Jul 14;358:j3360. [doi: 10.1136/bmj.j3360] [Medline: 28710272]
42. Nmadu A, Omole V, Oyefabi A, Usman N, Igboanusi C, Gobir A, et al. Physician burnout: an overview. J Res Basic Clin Sc 2019 Mar 28;1(1):111-114 [FREE Full text]
43. Hauck S, Gabbard GO. Institutional factors in the medical burnout epidemic. Braz J Psychiatry 2019;41(2):191-192 [FREE Full text] [doi: 10.1590/1516-4446-2018-0340] [Medline: 30942324]
44. Furlow B. Information overload and unsustainable workloads in the era of electronic health records. Lancet Respir Med 2020 Mar;8(3):243-244. [doi: 10.1016/S2213-2600(20)30010-2] [Medline: 32135094]
45. Walton G. Knowledge Management in the Age of Information Overload. In: Proceedings of the 19th Congress of the European Association of Hospital Pharmacists. 2014 Presented at: EAHP'14; April 8, 2014; Barcelona, Spain URL: https://www.managementstudyguide.com/knowledge-management-in-age-of-information-overload.htm
46. Braun V, Clarke V. What can 'thematic analysis' offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
47. Kaiser H. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36 [FREE Full text] [doi: 10.1007/bf02291575]

XSL•FO
RenderX

48. Hellier C, Tully V, Forrest S, Jaggard P, MacRae M, Habicht D, et al. Improving multidisciplinary communication at ward board rounds using video enhanced reflective practice. BMJ Qual Improv Rep 2015;4(1):e [FREE Full text] [doi: 10.1136/bmjquality.u206968.w2801] [Medline: 26734342]
49. Penney G. Adopting and adapting clinical guidelines for local use. Obstet Gynecol 2007 Jan;9(1):48-52 [FREE Full text] [doi: 10.1576/toag.9.1.048.27296]
50. Russell-Rose T, Chamberlain J. Expert search strategies: the information retrieval practices of healthcare information professionals. JMIR Med Inform 2017 Oct 2;5(4):e33 [FREE Full text] [doi: 10.2196/medinform.7680] [Medline: 28970190]
51. Aisling M, Aisling D, David C. An assessment of psychological need in emergency medical staff in the northern health and social care trust area. Ulster Med J 2016 May;85(2):92-98 [FREE Full text] [Medline: 27601762]
52. Banks S, Landon LB, Dorrian J, Waggoner LB, Centofanti SA, Roma PG, et al. Effects of fatigue on teams and their role in 24/7 operations. Sleep Med Rev 2019 Dec;48:101216. [doi: 10.1016/j.smrv.2019.101216] [Medline: 31630015]
53. Brehm S, Brehm J. Psychological Reactance: A Theory of Freedom and Control. New York, USA: Academic Press; 2013.
54. de Angelis M. Choice contexts and social communication. Micro Macro Mark 2017(3):435-456. [doi: 10.1431/88004]
55. Pilli L, Mazzon J. Information overload, choice deferral, and moderating role of need for cognition: empirical evidence. Rev Adm Empres 2016 Jan 1;51(1):36-55 [FREE Full text] [doi: 10.5700/rausp1222]
56. Roetzel P. Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. Bus Res 2018 Jul 6;12(2):479-522 [FREE Full text] [doi: 10.1007/s40685-018-0069-z]
57. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in 'early adopter' hospitals. Br Med J 2011 Oct 17;343:d6054 [FREE Full text] [doi: 10.1136/bmj.d6054] [Medline: 22006942]
58. Ozkula G, Durukan E. Burnout syndrome among physicians: the role of socio-demographic characteristics. J Psychiatry Neurosci 2017 Jun 15;1(30):136-144. [doi: 10.5350/dajpn2017300207]
59. Maggio LA, Artino AR. Staying up to date and managing information overload. J Grad Med Educ 2018 Oct;10(5):597-598 [FREE Full text] [doi: 10.4300/JGME-D-18-00621.1] [Medline: 30377482]
60. Roland D, Brazil V. Top 10 ways to reconcile social media and 'traditional' education in emergency care. Emerg Med J 2015 Oct;32(10):819-822. [doi: 10.1136/emermed-2015-205024] [Medline: 26253148]
61. Garner J, Murray S, Parisaei M. Prioritisation on labour ward. Obstet Gynaecol Reprod Med 2019 Oct;29(10):292-297 [FREE Full text] [doi: 10.1016/j.ogrm.2019.06.004]
62. Fink G. Stress, definitions, mechanisms, and effects outlined: lessons from anxiety. Handbook of Stress Series 2016 Jan;1:3-11 [FREE Full text] [doi: 10.1016/b978-0-12-800951-2.00001-7]
63. Li CJ, Syue YJ, Tsai TC, Wu KH, Lee CH, Lin YR. The impact of emergency physician seniority on clinical efficiency, emergency department resource use, patient outcomes, and disposition accuracy. Medicine (Baltimore) 2016 Feb;95(6):e2706 [FREE Full text] [doi: 10.1097/MD.0000000000002706] [Medline: 26871807]
64. Bovier PA, Perneger TV. Stress from uncertainty from graduation to retirement--a population-based study of Swiss physicians. J Gen Intern Med 2007 May;22(5):632-638 [FREE Full text] [doi: 10.1007/s11606-007-0159-7] [Medline: 17443371]
65. Kilbertus S, Pardhan K, Zaheer J, Bandiera G. Transition to practice: evaluating the need for formal training in supervision and assessment among senior emergency medicine residents and new to practice emergency physicians. CJEM 2019 May;21(3):418-426. [doi: 10.1017/cem.2019.8] [Medline: 30841941]
66. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: implications for assessment and entrustment. Med Educ 2018 Apr 19;52:970-980 [FREE Full text] [doi: 10.1111/medu.13588] [Medline: 29676054]

Abbreviations

ANOVA: analysis of variance
EMP: emergency medicine physician
ICT: Information and Communication Technology
IT: information technology
NHS: National Health Service
PCA: principal component analysis
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Eysenbach; submitted 04.04.20; peer-reviewed by X Lu, Y Zhang, R Guo; comments to author 11.05.20; revised version received 22.05.20; accepted 11.06.20; published 27.07.20.

Please cite as:
Sbaffi L, Walton J, Blenkinsopp J, Walton G
Information Overload in Emergency Medicine Physicians: A Multisite Case Study Exploring the Causes, Impact, and Solutions in Four North England National Health Service Trusts
J Med Internet Res 2020;22(7):e19126
URL: http://www.jmir.org/2020/7/e19126/
doi:10.2196/19126
PMID:32716313

©Laura Sbaffi, James Walton, John Blenkinsopp, Graham Walton. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Patient Experience and Healthcare Quality of Dental Care Using Patient Online Reviews in the United States: Mixed Methods Study

Ye Lin[1], MSc; Y Alicia Hong[2], PhD; Bradley S Henson[1*], DDS, PhD; Robert D Stevenson[1*], DDS; Simon Hong[1*], MPH, DMD; Tianchu Lyu[3], MPH; Chen Liang[3], PhD

[1]College of Dental Medicine, Western University of Health Sciences, Pomona, CA, United States

[2]College of Health and Human Services, George Mason University, Fairfax, VA, United States

[3]Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[*]these authors contributed equally

Corresponding Author:
Chen Liang, PhD
Arnold School of Public Health
University of South Carolina
Suite 347
915 Greene Street
Columbia, SC, 29208
United States
Phone: 1 803 777 8139
Email: cliang@mailbox.sc.edu

Abstract

Background: Over the last two decades, patient review websites have emerged as an essential online platform for doctor ratings and reviews. Recent studies suggested the significance of such websites as a data source for patients to choose doctors for healthcare providers to learn and improve from patient feedback and to foster a culture of trust and transparency between patients and healthcare providers. However, as compared to other medical specialties, studies of online patient reviews that focus on dentists in the United States remain absent.

Objective: This study sought to understand to what extent online patient reviews can provide performance feedbacks that reflect dental care quality and patient experience.

Methods: Using mixed informatics methods incorporating statistics, natural language processing, and domain expert evaluation, we analyzed the online patient reviews of 204,751 dentists extracted from HealthGrades with two specific aims. First, we examined the associations between patient ratings and a variety of dentist characteristics. Second, we identified topics from patient reviews that can be mapped to the national assessment of dental patient experience measured by the Patient Experience Measures from the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Dental Plan Survey.

Results: Higher ratings were associated with female dentists (t_{71881}=2.45, P<.01, g=0.01), dentists at a younger age ($F_{7, 107128}$=246.97, P<.001, g=0.11), and those whose patients experienced a short wait time ($F_{4, 150055}$=10417.77, P<0.001, g=0.18). We also identified several topics that corresponded to CAHPS measures, including discomfort (eg, painful/painless root canal or deep cleaning), and ethics (eg, high-pressure sales, and unnecessary dental work).

Conclusions: These findings suggest that online patient reviews could be used as a data source for understanding the patient experience and healthcare quality in dentistry.

(J Med Internet Res 2020;22(7):e18652) doi:10.2196/18652

KEYWORDS

dental care; healthcare quality; consumer health informatics; patient online reviews; patient review websites; natural language processing

XSL•FO
RenderX

Introduction

Over the last few years, patient review websites have gained increasing interest among health consumers, academic communities, and healthcare providers [1,2]. A tremendous amount of online patient reviews were shared with the public through patient review websites, becoming a common source of information for patients choosing a doctor. Accordingly, a growing body of literature in online patient review studies has been developed using public health informatics methods to leverage the distribution and determinants of data for informing public health and public policy, coinciding with the Research Framework of Infodemiology and Infoveillance [3].

Despite debates on whether patient-generated reviews would be useful to improve healthcare quality [4-6], online reviews could serve as a valuable data source for understanding patient experience and the patient-provider relationship, and present unique values in improving dental care. Online reviews play an increasingly important role in health consumers' decision-making for choosing dental services [7,8]. A recent national survey reported that 59% of respondents recognized the importance of patient review websites, 35% selected a doctor based on good ratings, and 37% avoided selecting a doctor based on bad ratings. Nearly 90% of the respondents rated listings of accepted health insurance on review sites as high importance [2]. Online patient review studies also found strong associations between ratings and doctor characteristics, such as gender, age, years of practice, online presence, and medical education and training, with variations at specialties [1,9].

Online reviews are also rich in data on the patient experience, a critical measure of healthcare quality [10]. The notion of patient experience includes the entire scope of interactions, from appointment scheduling and access to information to communications with clinicians, cost, and payment. Positive patient experience is associated with better prevention and treatment adherence, safer healthcare outcomes, and better utilization of healthcare resources [11-15]. Incorporating data from the patient experience is also consistent with the principles of patient-centered care [10], as well as the goal of public reporting and the performance-based payment model set by the Centers for Medicare and Medicaid Services (CMS) [16].

Online patient reviews can inform better decision-making among patients and can be used to improve healthcare quality. As such, many studies on online patient reviews in a variety of medical specialties have been reported [1]. Nevertheless, analyses of online reviews of dentists are sparse. We only identified one study of online patient reviews in dentistry conducted in Germany [17]. In the United States, studies of dental patient reviews are missing. To address the literature gaps, we conducted a study with two aims. The first aim was to characterize the online reviews of dental patients in the United States by analyzing data extracted from patient review websites. In particular, we examined the association between dentist characteristics and patient ratings.

Our second aim was to understand to what extent online reviews can inform assessments of the patient experience by identifying semantic mentions of the patient experience in patient reviews. Dental patient experience is traditionally assessed with the CAHPS (Consumer Assessment of Healthcare Providers and Systems) Dental Plan Survey, administered by CMS [18]. Using online reviews for patient experience assessment is feasible but challenging. For instance, information extraction from online patient reviews is labor-intensive because the reviews are written in free text. Content analysis frequently used in social science research could not be applied to this study, in which there were hundreds of thousands of reviews to assess. The identification of information relevant to the patient experience is also necessary to ensure research validity, yet there are no guidelines or empirical studies on how to identify patient experience content in online reviews. Patient reviews cover a variety of topics, but not all of them are related to the patient experience. For example, topics of patient satisfaction should not be identified as patient experience. The patient experience is comprised of the clinical encounters and has been frequently used as an indicator of patient safety [19], whereas patient satisfaction relates to whether patients' expectations about healthcare are met [20]. To overcome these challenges, we incorporated natural language processing [21] and human expert evaluation to identify relevant topics from online patient reviews. Natural language processing has extensive applications in health informatics and medical internet research, in which it has vastly improved the efficiency of processing free-text data using advanced statistical methods and automated computing [3]. Human expert evaluation mitigates the problem of misprocessed data typically produced by such a computerized method and improves the interoperability of the data analysis and results.

Methods

Data Acquisition

Publicly accessible online review data were obtained from HealthGrades, a well-known patient review website in the United States. Among many other sites, we focused on HealthGrades for two reasons. First, HealthGrades is widely used by patients who receive healthcare services from a full range of medical specialties in the United States. Second, HealthGrades provides a well-organized sitemap structure that facilitates data extraction. We analyzed data from a single review site because the data structure and measures of dentist demographics and performance vary, which will hinder data consolidation. In addition, data from multiple patient review sites would have little impact on representation and generalizability of this study because an active dentist typically has profiles on all popular patient review sites.

Online reviews for 204,751 dentists were extracted. This census approximates but does not fully match the workforce statistics (199,486 working dentists as of 2018) reported by the American Dental Association (ADA) [22] because some profiles are for recently inactive dentists and misclaimed or inappropriately captured profiles. The data contain the following attributes: state, city, specialty, gender, age, language, education, number of reviews, ratings, reviews, and wait time.

XSL•FO
RenderX

Data Preprocessing

There were 41 dental service listings in HealthGrades. We categorized these services into the 10 dental specialties defined by the ADA, general dentistry, and others (ie, unidentified and miscellaneous), resulting in 8 specialties (ie, dental anesthesiology, endodontics, oral and maxillofacial pathology, oral and maxillofacial surgery, orthodontics and dentofacial orthopedics, pediatric dentistry, periodontics, and prosthodontics), general dentistry, and others for downstream analyses. Public health dentistry and oral and maxillofacial radiology were excluded as they have only one entry at most. The data extraction was completed in September 2019. The study was identified as a nonhuman study by the Institutional Review Board of the University of South Carolina.

Statistical Analyses

We employed statistical analyses to assess the associations between ratings and dentists' characteristics using R Project for Statistical Computing. We used descriptive statistics to calculate proportions and mean distributions. Based on reported online reviews studies in general medical specialties, we hypothesized that (1) female dentists, (2) young dentists, and (3) short wait times would be associated with higher overall ratings. We also hypothesized that specialties are associated with overall ratings. An independent sample t-test was used to test whether ratings differ by gender. Analysis of Variance (ANOVA) was used to test whether ratings differ by specialties, age, and wait time, respectively. We used Hedges *g* to approximate the effect size because of unbalanced sample sizes in comparison groups.

Text Mining

Semi-automated natural language processing was used to identify concepts related to patient experience with limited human labor required. Figure 1 is a diagram of text mining procedures. We used Python for text processing and computation. Built on our pilot study of analyzing patient-generated reviews [23], we first extracted reviews from online sites, followed by standard data cleaning procedures, including tokenization and removal of stop words.

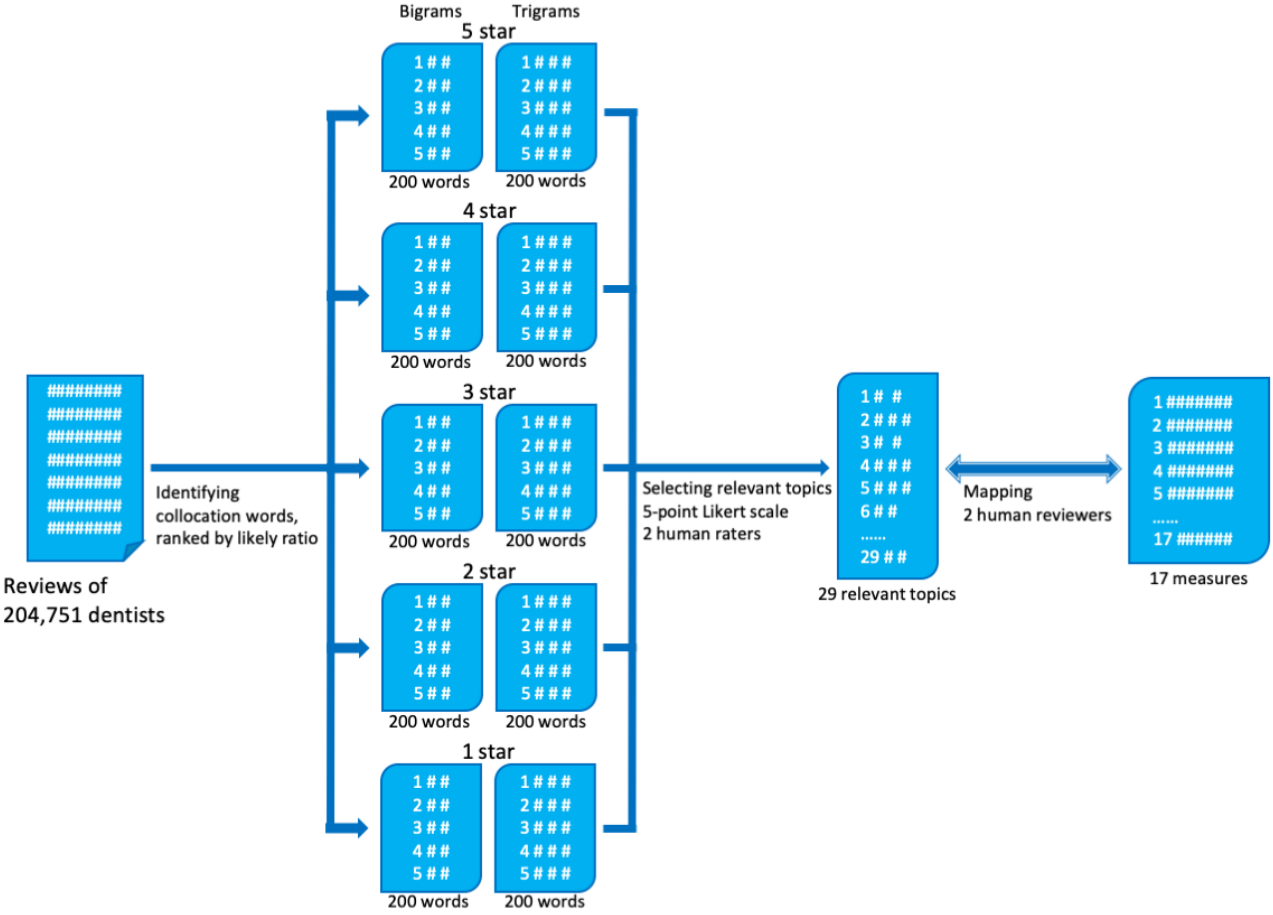
We then calculated bigram and trigram collocations. Collocations are habitual expressions of multiple words. In this study, we ranked using the "likelihood ratio" method [24], top 200 bigram and trigram collocations, respectively, from reviews associated with every rating category (ie, 1, 2, 3, 4, 5).

We observed a number of the collocations that are irrelevant to concepts of dental care and patient experience (eg, "phone call" and "many years ago") but were still ranked top 200 by likelihood ratio. Therefore, two raters (YL and CL) independently picked collocations related to patient experience using a 4-point Likert scale ("definitely relevant," "somewhat relevant," "somewhat irrelevant," "definitely irrelevant"). Inter-rater reliability was assessed using Cohen kappa. Two raters discussed on the collocations that received contrary opinions (relevant vs irrelevant) until a consensus was reached.

Next, we mapped patient-experience-related concepts onto a total of 17 composite measures from the Patient Experience Measures from the CAHPS Dental Plan Survey. Table 1 shows the 17 measures categorized in three dimensions. Two reviewers (RS and JT), who are professors of dental medicine and licensed dentists, independently completed the mapping procedures. A consensus was reached after a discussion of the initial mapping results.

XSL•FO
RenderX

Figure 1. Diagram of text mining procedures.



XSL•FO
RenderX

Table 1. Composite measures from the Patient Experience Measures from the CAHPS Dental Plan Survey.

Dimensions, items	Composite measures
Care from dentists and staff	
Q6	How often did your regular dentist explain things in a way that was easy to understand?
Q7	How often did your regular dentist listen carefully to you?
Q8	How often did your regular dentist treat you with courtesy and respect?
Q9	How often did your regular dentist spend enough time with you?
Q11	How often did the dentists or dental staff do everything they could to help you feel as comfortable as possible during your dental work?
Q12	How often did the dentists or dental staff explain what they were doing while treating you?
Access to dental care	
Q13	How often were your dental appointments as soon as you wanted?
Q15	If you tried to get an appointment for yourself with a dentist who specializes in a particular type of dental care (such as root canals or gum disease) in the last 12 months, how often did you get an appointment as soon as you wanted?
Q16	How often did you have to spend more than 15 minutes in the waiting room before you saw someone for your appointment?
Q17	If you had to spend more than 15 minutes in the waiting room before you saw someone for your appointment, how often did someone tell you why there was a delay or how long the delay would be?
Q14	If you needed to see a dentist right away because of a dental emergency in the last 12 months, did you get to see a dentist as soon as you wanted?
Dental plan costs and services	
Q19	How often did your dental plan cover all of the services you thought were covered?
Q22	How often did the 800 number, written materials, or website provide the information you wanted?
Q27	How often did your dental plan's customer service give you the information or help you needed?
Q28	How often did your dental plan's customer service staff treat you with courtesy and respect?
Q20	Did your dental plan cover what you and your family needed to get done?
Q24	Did this information (from your dental plan) help you find a dentist you were happy with?

Results

Descriptive Statistics

Table 2 shows the demographics of these dentists. Among the collected data, 154.683 (75.55%) dentists received at least one rating or review. The number of reviews ranged from 1 to 1789. There were more male dentists than females with a percentage difference of 85.87%. The majority of dentists (78.55%) were general dentists, followed by dentists specialized in orthodontics and dentofacial orthopedics (9.23%), pediatrics (4.76%), and endodontics (3.81%). The majority of dentists (91.83%) did not specify any second languages other than English. Among those who indicated speaking a second language, Spanish, Hindi, Arabic, French, and Chinese were most common. There were 150,571 (73.54%) dentists who received an overall rating of ≥3 out of 5, and 50,068 (24.45%) of all dentists received no ratings.

XSL•FO
RenderX

Table 2. Dentist demographics.

	Count	Proportion (%)
Gender		
Female	58,309	28.48
Male	146,044	71.33
Unknown	398	0.19
Age		
Under 30	1585	0.77
30-39	28,736	14.03
40-49	36,715	17.93
50-59	29,006	14.17
60-69	29,585	14.45
70-79	11,716	5.72
80-89	1826	0.89
Over 89	157	0.08
Unknown	65,425	31.95
Specialty		
Dental Anesthesiology	338	0.17
Endodontics	7803	3.81
General Dentistry	160,831	78.55
Oral and Maxillofacial Pathology	166	0.08
Oral and Maxillofacial Radiology	1	0.00
Oral and Maxillofacial Surgery	975	0.48
Orthodontics and Dentofacial Orthopedics	18,891	9.23
Pediatric Dentistry	9743	4.76
Periodontics	1541	0.75
Prosthodontics	4249	2.08
Other	213	0.10
Unknown	0	0
Language		
Spanish	3972	1.94
Hindi	510	0.25
Arabic	471	0.23
French	467	0.23
Chinese	440	0.21
Russian	429	0.21
Farsi	338	0.17
Vietnamese	305	0.15
Korean	304	0.15
Portuguese	272	0.13
Unknown	188,025	91.83
Rating		

XSL•FO
RenderX

	Count	Proportion (%)
1-1.9	1344	0.66
2-2.9	2768	1.35
3-3.9	16,431	8.02
4-4.9	61,520	30.05
5	72,620	35.47
Unknown	50,068	24.45
Wait time		
Under 10 minutes	98,104	47.91
10-15 minutes	46,347	22.64
16-30 minutes	4395	2.15
31-45 minutes	880	0.43
Over 45 minutes	335	0.16
Unknown	54,690	26.71

Inferential Statistics

ANOVA showed no significant effect of specialties on ratings ($F_{9, 154673}$=58.74, P=1.37, g=0.06). However, the average rating was higher for female dentists (M=4.58) than male dentists (M=4.57) (t_{71881}=2.45, P<.01, g=0.01). See Figure 2.

We also found a significant effect of age on ratings ($F_{7, 107128}$=246.97, P<.001, g=0.11) with younger age associated with higher ratings. The Tukey Honestly Significant Difference (Tukey HDS) test showed a significant difference in ratings between each pair of age groups except “30-39” vs “Over 89” (P=.15, g=0.01), “Under 30” vs “30-39” (P=.10, g=0.01), “40-49” vs “Over 89” (P=.95, g=0), “50-59” vs “60-69” (P=1.00, g=0), “50-59” vs “Over 89” (P=1.00, g=0), “60-69” vs “Over 89” (P=1.00, g=0), and “80-89” vs “Over 89” (P=1.00, g=0). See Figure 3.

There was also a significant effect of wait time on ratings ($F_{4, 150055}$=10417.77, P<.001, g=0.26) with shorter wait times associated with higher ratings. Tukey HDS showed statistical significance in comparing the mean difference of ratings in each pair of wait times (all P<.001, g=[0.01, 0.18]). See Figure 4.

XSL•FO
RenderX

Figure 2. Overall ratings by gender.

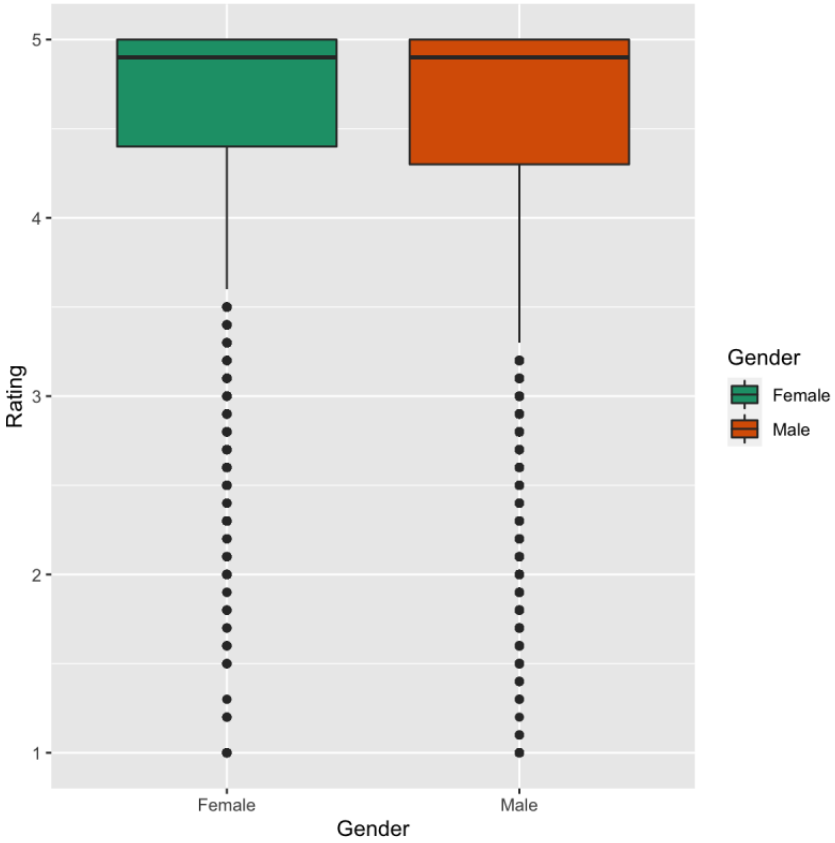
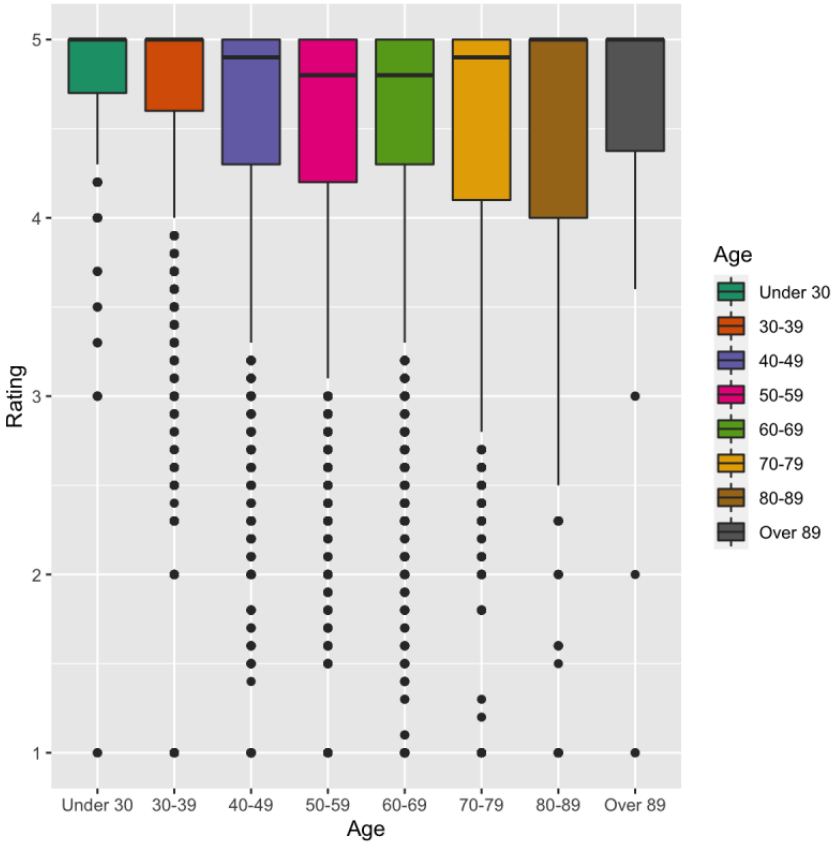
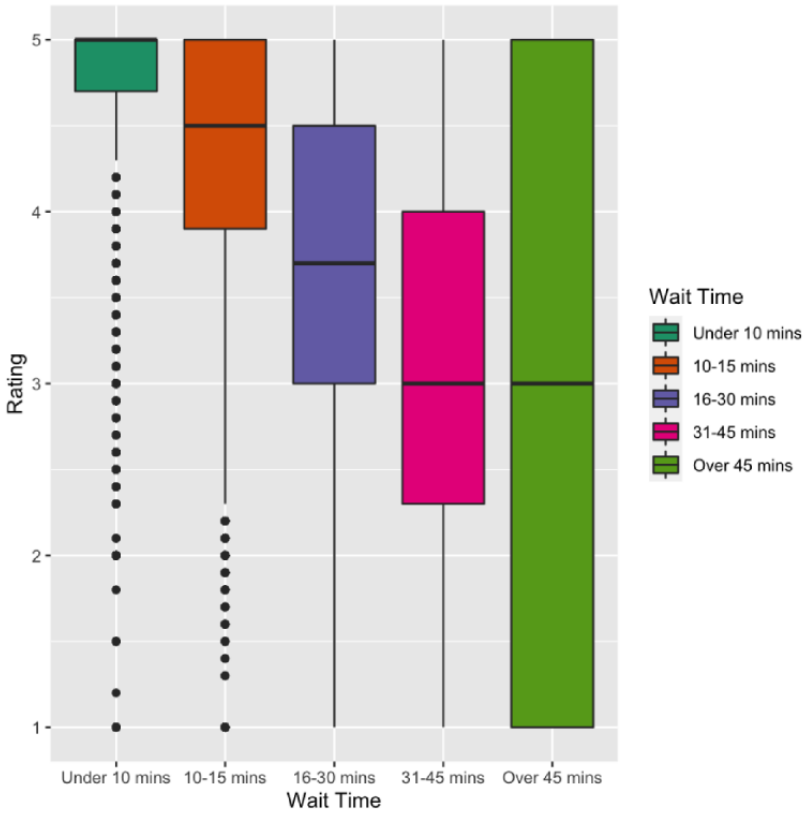


Figure 3. Overall ratings by age group.



XSL•FO
RenderX

Figure 4. Overall ratings by wait time.

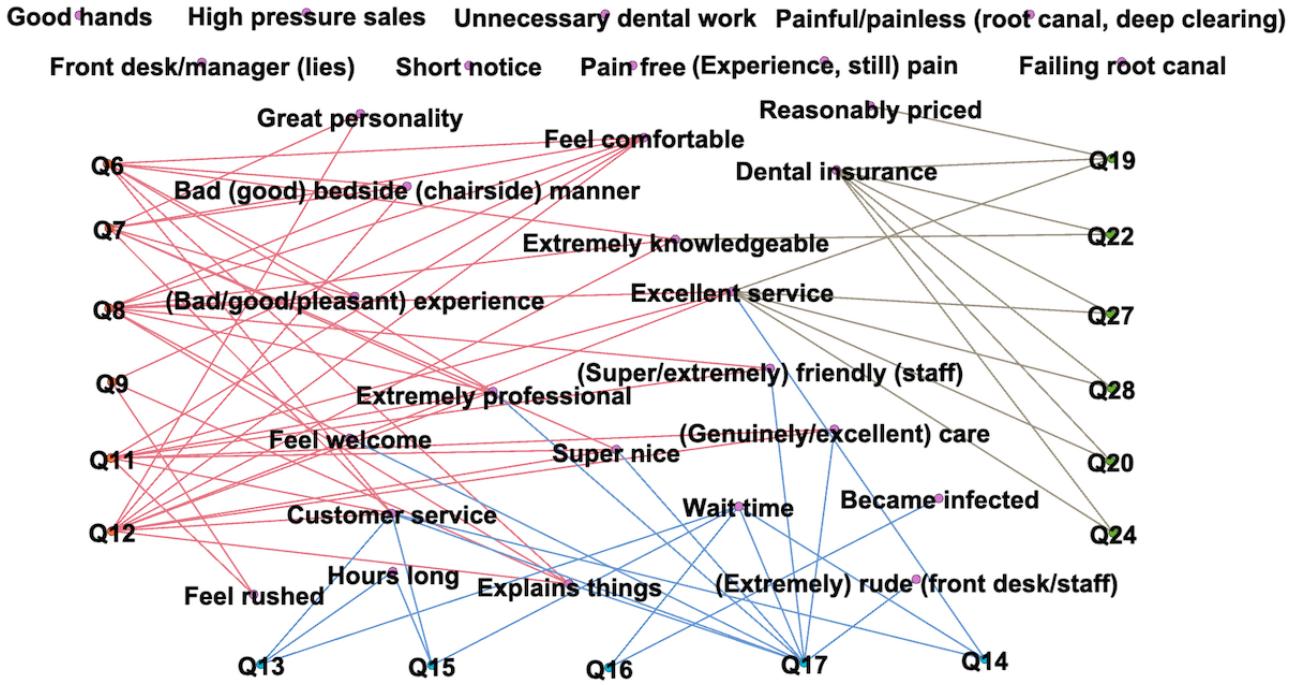


Identification of Concepts Relating to Dental Patient Experience

Cohen kappa (equally weighted) was 0.95 between the two raters who independently identified patient experience-related words and phrases from 2000 automatically extracted collocations. After discussion, they identified 29 words and phrases, then two other reviewers of dental experts independently mapped the 29 words and phrases onto the 17 composite measures in the Patient Experience Measures from the CAHPS Dental Plan Survey. Figure 5 shows a map of links. Each composite measure has 2-10 representing words and phrases. Out of the three dimensions of the patient experience, there were more topics representing “care from dentists and staff” and “access to dental care” compared to “dental plan costs and services.” Patients were more likely to discuss their experience with dentists and staff than health insurance providers. There were eight words and phrases related to the patient experience that did not correspond to any composite measures. Some of these topics were specific to dental care. For example, discomfort (eg, painful/painless root canal or deep cleaning) is a common type of feedback from dental patients. Topics relating to ethics (eg, high-pressure sales and unnecessary dental work) have received little attention in dentistry but merits further research.

XSL•FO
RenderX

Figure 5. Mapping of Patient Experience Measures from the CAHPS Dental Plan Survey and the words and phrases extracted from patient reviews.



Discussion

Principal Findings

Over the last few years, researchers have begun a systematic analysis of online patient reviews. In the United States, several empirical studies investigating online reviews of general healthcare services and specialties are well documented, but such studies have not been performed in dentistry [1]. To the best of our knowledge, this study represents the first study of online reviews of dental care in the United States. In particular, this study demonstrated that online patient reviews are an essential data source for studying dental patient experience. This study also characterized patient feedback to dental care, which can inform dental care quality improvement.

We found several factors associated with overall dentist ratings. In particular, female dentists were rated slightly higher than their male counterparts, although this should be interpreted with the understanding that the effect size is small (g=0.01), and the sample size was small for senior dentists. The gender and age differences we found in this study were similar to a study of online reviews of dentists in Germany [25] as well as reviews of surgeons [26,27], but such a statistical significance was not consistent in other studies. Earlier studies on patient-provider communication suggested that female healthcare providers engage in significantly more active partnership behaviors, positive talk, and building trust with patients [28]. Female providers were also found associated with lower mortality and readmission as compared to their male counterparts [29]. The literature on healthcare quality suggests that younger physicians in acute care hospitals report lower mortality rates [30]. Our data also showed that higher overall ratings were associated with shorter wait times (g=0.26). This finding has been reported and discussed in several studies of online patient reviews [1]. Patient review websites such as HealthGrades and RateMDs have included “wait time/punctuality” as a default measure for healthcare providers.

The proliferation of patient review websites represents a wealth of patient-experience data, but these data remain understudied. In this study, we identified unstructured descriptions of patient experience using a method integrating quantitative text mining and qualitative human evaluation. Our method recognizes the role of automated textual data analytics in harnessing information from online reviews, consistent with other recent studies [31-37], although some researchers argued the limitations of text mining because it involves limited supervision by human experts [31]. To minimize the impact of this lack of human oversight, we incorporated human evaluation procedures into the automated natural language processing effort. Our findings showed that patient reviews covered a full range of topics measured by the Patient Experience Measures from the CAHPS Dental Plan Survey and demonstrated a high level of correlation. Among the eight topics not corresponding to any measures from the survey, some topics may have provided a nuanced view of dental care as compared to patient survey responses [36]. These findings suggest that online patient reviews can be used to assess patient experience during dental care.

Limitations

Our study has the following limitations. First, data from Healthgrades, like data from any other patient review website, may be incomplete and biased. Not all dentist profiles have been claimed. Although authentication is required for dentist profile information, inaccuracies may still exist. The overall ratings may be biased because patients who are happy with health services are more likely to leave ratings and reviews [2]. Second, although there are good reasons for using data from a single review site, this limitation may still weaken the reliability of the study. Repeated measures and follow up studies are needed to evaluate the findings of the this study. Third, it is challenging to differentiate the descriptions of the patient

XSL•FO
RenderX

experience from patient satisfaction as patients often write about their satisfactory or unsatisfactory experience with emotions and personal preferences. In this study, we only analyzed words and phrases at a semantic level, while contextual information is limited. Third, text mining is efficient for big data analyses but falls short in domain-specific and context-based analyses, in which traditional qualitative approaches should be considered as an essential complement of text mining [31].

Despite these limitations, this study analyzed an extensive dataset and found an association between dentists and online patient reviews. The thematic analysis also identified themes of patient experience similar to those of CAHPS, suggesting that online patient reviews can inform improved quality in dental care.

Conclusions

This study demonstrated that PORs are a potential data source that can supply rich performance data from the patient perspective, based on which assessments of dental care quality and the patient experience is feasible. We also identified several factors associated with dentists' overall ratings, which could be used to inform dental care quality improvement.

Acknowledgments

We thank Dr Jenny S Tjahjono for the contribution to the expert evaluation. This study is supported by a seed grant from the University of South Carolina.

Conflicts of Interest

None declared.

References

1. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What Do Patients Say About Doctors Online? A Systematic Review of Studies on Patient Online Reviews. J Med Internet Res 2019 Apr 08;21(4):e12521 [FREE Full text] [doi: 10.2196/12521] [Medline: 30958276]
2. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
3. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
4. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009;338:b1030. [Medline: 19293223]
5. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ 2009 Mar 17;338:b1033. [doi: 10.1136/bmj.b1033] [Medline: 19293224]
6. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 07;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]
7. Farley DO, Elliott MN, Short PF, Damiano P, Kanouse DE, Hays RD. Effect of CAHPS performance information on health plan choices by Iowa Medicaid beneficiaries. Med Care Res Rev 2002 Sep;59(3):319-336. [doi: 10.1177/107755870205900305] [Medline: 12205831]
8. Bader JD, Shugars DA. Dental care plan assessment using the CAHPS Dental Plan Survey. J Public Health Dent 2012;72(4):295-301. [doi: 10.1111/j.1752-7325.2012.00337.x] [Medline: 22497638]
9. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
10. Committee on Quality of Health Care in America, Institute of Medicine Staff. Crossing the quality chasm: A new health system for the 21st century. Washington DC: National Academies Press; Jul 19, 2001.
11. Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev 2014 Oct;71(5):522-554 [FREE Full text] [doi: 10.1177/1077558714541480] [Medline: 25027409]
12. Browne K, Roseman D, Shaller D, Edgman-Levitan S. Analysis & commentary. Measuring patient experience as a strategy for improving primary care. Health Aff (Millwood) 2010 May;29(5):921-925. [doi: 10.1377/hlthaff.2010.0238] [Medline: 20439881]
13. Davies E, Shaller D, Edgman-Levitan S, Safran DG, Oftedahl G, Sakowski J, et al. Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. Health Expect 2008 Jun;11(2):160-176 [FREE Full text] [doi: 10.1111/j.1369-7625.2007.00483.x] [Medline: 18494960]
14. Friedberg MW, SteelFisher GK, Karp M, Schneider EC. Physician groups' use of data from patient experience surveys. J Gen Intern Med 2011 May;26(5):498-504 [FREE Full text] [doi: 10.1007/s11606-010-1597-1] [Medline: 21161419]
15. Goldstein E, Cleary PD, Langwell KM, Zaslavsky AM, Heller A. Medicare Managed Care CAHPS: A Tool for Performance Improvement. Health Care Financ Rev 2001;22(3):101-107 [FREE Full text] [Medline: 25372460]

XSL•FO
RenderX

16. CMS Issues final rule for first year of hospital value-based purchasing program. Centers for Medicare & Medicaid Services. 2011. URL: https://www.cms.gov/newsroom/fact-sheets/cms-issues-final-rule-first-year-hospital-value-based-purchasing-program [accessed 2020-06-04]
17. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
18. CAHPS Dental Plan Survey. Agency for Healthcare Research and Quality. 2011. URL: https://www.ahrq.gov/cahps/surveys-guidance/dental/index.html [accessed 2020-06-04]
19. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 03;3(1):e001570 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
20. What Is Patient Experience? Agency for Healthcare Research and Quality. 2016. URL: https://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html [accessed 2020-06-04]
21. Manning CD, Schutze H. Foundations of statistical natural language processing. Cambridge: MIT Press; 1999.
22. Workforce. American Dental Association. 2019. URL: https://www.ada.org/en/science-research/health-policy-institute/dental-statistics/workforce [accessed 2020-06-04]
23. Lin Y, Hong S, Liang C. Semantic Analysis of Online Dentist Review: Toward Assessing Safety and Quality of Dental Care. 2019 Presented at: International Conference on Human-Computer Interaction; 2019; Orlando FL p. 488-494. [doi: 10.1007/978-3-030-23525-3_66]
24. Dunning T. Accurate methods for the statistics of surprise and coincidence. Computational linguistics 1993;19(1):61-74. [doi: 10.5555/972450.972454]
25. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
26. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a Sports Medicine Surgeon's "Quality" in the Modern Era: an Analysis of Popular Physician Online Rating Websites. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
27. Kirkpatrick W, Abboudi J, Kim N, Medina J, Maltenfort M, Seigerman D, et al. An Assessment of Online Reviews of Hand Surgeons. Arch Bone Jt Surg 2017 May;5(3):139-144 [FREE Full text] [Medline: 28656160]
28. Roter D, Lipkin M, Korsgaard A. Sex differences in patients' and physicians' communication during primary care medical visits. Med Care 1991 Nov;29(11):1083-1093. [doi: 10.1097/00005650-199111000-00002] [Medline: 1943269]
29. Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians. JAMA Intern Med 2017 Feb 01;177(2):206-213 [FREE Full text] [doi: 10.1001/jamainternmed.2016.7875] [Medline: 27992617]
30. Tsugawa Y, Newhouse JP, Zaslavsky AM, Blumenthal DM, Jena AB. Physician age and outcomes in elderly patients in hospital in the US: observational study. BMJ 2017 May 16;357:j1797 [FREE Full text] [doi: 10.1136/bmj.j1797] [Medline: 28512089]
31. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
32. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
33. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
34. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
35. Hao H, Zhang K, Wang W, Gao G. A tale of two countries: International comparison of online doctor reviews between China and the United States. Int J Med Inform 2017 Mar;99:37-44. [doi: 10.1016/j.ijmedinf.2016.12.007] [Medline: 28118920]
36. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Aff (Millwood) 2016 Apr;35(4):697-705 [FREE Full text] [doi: 10.1377/hlthaff.2015.1030] [Medline: 27044971]
37. Agarwal AK, Mahoney K, Lanza AL, Klinger EV, Asch DA, Fausti N, et al. Online Ratings of the Patient Experience: Emergency Departments Versus Urgent Care Centers. Ann Emerg Med 2019 Jun;73(6):631-638. [doi: 10.1016/j.annemergmed.2018.09.029] [Medline: 30392737]

Abbreviations

ADA: American Dental Association

XSL•FO
RenderX

ANOVA: Analysis of Variance
CAHPS: Consumer Assessment of Healthcare Providers and Systems
CMS: Centers for Medicare and Medicaid Services
Tukey HDS: Tukey's Honestly Significant Difference

Edited by G Eysenbach; submitted 10.03.20; peer-reviewed by M Nomali, A Mavragani; comments to author 07.04.20; revised version received 21.04.20; accepted 14.05.20; published 07.07.20.

Please cite as:
Lin Y, Hong YA, Henson BS, Stevenson RD, Hong S, Lyu T, Liang C
Assessing Patient Experience and Healthcare Quality of Dental Care Using Patient Online Reviews in the United States: Mixed Methods Study
J Med Internet Res 2020;22(7):e18652
URL: https://www.jmir.org/2020/7/e18652
doi:10.2196/18652
PMID:32673240

©Ye Lin, Y Alicia Hong, Bradley S Henson, Robert D Stevenson, Simon Hong, Tianchu Lyu, Chen Liang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Causal Effect of Honorary Titles on Physicians' Service Volumes in Online Health Communities: Retrospective Study

Haiyan Yu[1], PhD; Yali Wang[1], BA; Jying-Nan Wang[2*], PhD; Ya-Ling Chiu[3*], PhD; Hang Qiu[4], PhD; Mingyue Gao[5], MSc

[1]Chongqing Key Laboratory of E-commerce and Modern Logistics, Chongqing University of Posts and Telecommunications, Chongqing, China

[2]College of International Finance and Trade, Zhejiang Yuexiu University of Foreign Languages, Shaoxing, China

[3]College of International Business, Zhejiang Yuexiu University of Foreign Languages, Shaoxing, China

[4]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

[5]Life Course Epidemiology and Biostatistics, Population, Policy, and Practice Programme, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Mingyue Gao, MSc
Life Course Epidemiology and Biostatistics, Population, Policy, and Practice Programme
Great Ormond Street Institute of Child Health, University College London
Guilford Street
London, WC1N 1EH
United Kingdom
Phone: 44 07410372022
Email: ming.gao.17@ucl.ac.uk

Abstract

Background: An OHC online health community (OHC) is an interactive platform for virtual communication between patients and physicians. Patients can typically search, seek, and share their experience and rate physicians, who may be involved in giving advice. Some OHC providers provide incentives in form of honorary titles to encourage the web-based involvement of physicians, but it is unclear whether the award of honorary titles has an impact on their consultation volume in an OHC.

Objective: This study is designed to identify the differential treatment effect of the incentive policy on the service volumes for the subgroups of treatment and control in an OHC. This study aims to answer the following questions: Does an honorary title for physicians impact their service volumes in an OHC? During the period of discontinuity, can we identify the sharp effect of the incentive award on the outcomes of physicians' service volumes?

Methods: We acquired the targeted samples based on treatment, namely, physicians with an honorary title or not and outcomes measured before and after the award of the 2 subgroups. A regression discontinuity design was applied to investigate the impact of the honorary titles incentive as a treatment in an OHC. There was a sharply discontinuous effect of treatment on physicians' online health service performance. The experimental data set consisted of 346 physicians in the treatment group (with honorary titles). Applying the propensity score matching method, the same size of physicians (n=346) was matched and selected as the control group.

Results: A sharp discontinuity was found at the time of the physician receiving the honorary title. The results showed that the parametric estimates of the coefficient were significantly positively (P<.001) associated with monthly home page views. The jump in the monthly volumes of home page views was much sharper than that of the monthly consultations.

Conclusions: The changes in the volumes of monthly consultations and home page views reflect the differential treatment effect of honorary titles on physicians' service volumes. The effect of the incentive policy with honorary titles is objectively estimated from both the perspective of online and offline medical services in an OHC. Being named with honorary titles significantly multiplied monthly home page views, yet it did not significantly impact monthly consultations. This may be because consultation capacity is limited by the physician's schedule for consultations.

(J Med Internet Res 2020;22(7):e18527) doi:10.2196/18527

XSL•FO
RenderX

KEYWORDS

causality; health information systems; organizational policy; physician-patient relations; remote consultation

Introduction

Background

Online health communities (OHCs) are an essential channel for creatively allocating health care resources among widely distributed patients in modern life. An OHC is usually referred to as an interactive platform between patients and physicians for virtual communication. Patients can search, seek, and share their experience of medical advice and rate the physician as social returns in an OHC. Meanwhile, physicians can share their medical knowledge and provide online medical services. An OHC is a convenient, real-time supplement for physician-patient interaction (PPI) without limitations on time and space [1]. Physicians' web-based involvement (ie, the service volumes) is a determinant factor of OHCs' success in health service delivery [2]. The platform management provider of an OHC often provides incentive policies (ie, honorary titles) to encourage physicians' web-based involvement. For example, 346 physicians on the *Good Doctor* platform were awarded the honorary title *2017 Annual Good Doctor* in January 2018. According to a survey of large OHC firms by the Towers Watson/National Business Group on Health [3], 69% reported that they offered wellness incentives and that the size of the incentives increased with time. However, it is still unclear whether the award of honorary titles has an impact on their consultation volume in an OHC.

With the quick development of OHCs, the physician-patient online interaction has gained more attention from scholars all over the world, especially regarding what incentive mechanisms and strategies can be designed to prompt sustained PPIs in OHCs [4]. In the bidirectional process of PPIs [5], patients seek information and make the decision to select a physician for consultation, and then, the physicians share their medical knowledge and provide medical services. Patients could also provide returns (rate, vote, and share experience online) for physicians. After receiving feedback and returns, physicians could balance their efforts for the subsequent PPI process. A good PPI can benefit both patients and physicians. It can provide patients with a truly information-based selection process and good outcome of the consultation process. In addition, a good PPI can provide physicians with returns that can also affect their reputation. Incentive policies have been widely developed to encourage PPI [3]. The high rating or award of good doctors online is thought to be a good indicator. This indicator not only represents the praise of the physician but can also predict the following process of PPI theoretically. Recently, there have been more studies focusing on the impact of incentive policies (high rating or honorary title of *Good Doctor*) in OHCs [1]. A previous study reported that physicians' online contributions and reputation were closely associated with patients' decision-making process when seeking medical consultation [6]. Due to the cross-sectional study design, the conclusion can only be explained in terms of association [7]. To evaluate the causal impact of incentive policy on physicians' consultation volumes, a sequential occurrence needs to be considered and analyzed in causal design, as appropriate.

In this study, we intend to estimate the causal effect of an incentive strategy on the PPI process of an OHC. Our analysis uses the information from the biggest OHC in China (*Good Doctor* website) [1], which provides reliable information on more than 10,000 hospitals and more than 640,000 physicians across the country. The specific treatment in our study is the award of the honorary title of *Annual Good Doctor*, which is assigned by a threshold of a weighted score summarizing the hospital level, the professional capital, number of votes, and the experience of health care service of patients. The dimensions of treatment included both temporal (pre- and post-treatment) and treatment (treated and untreated group) effects. The outcome contained 2 aspects [8]: the home page view frequency and the online consultation frequency. The home page view frequency mainly represents the physicians' reputation and effort from the patients' side. The online consultation frequency further includes the accessibility and efforts of physicians when chosen for online services. The 2 indicators together represent the performance of the continuous PPI process in OHCs.

Literature Review

Although the natural experiment design is widely adopted to evaluate the policy's causal effects in empirical research [9], the experiment remains difficult or impossible to implement because of ethical, political, and financial reasons. A large share of the empirical work on policy evaluation relies on observational data, in which policies are determined in a way other than through randomization assignment. Drawing the inference of a policy's causal effect based on observational data is quite challenging, especially for incentive policy evaluation in an OHC setting. Several methodological issues need to be overcome by adopting causal methods creatively because of the various weaknesses of classic methods.

The counterfactual impact of the treatment needs to be estimated in causal inference. *Treatment* is a general term referring to certain interventions of interest, for example, the incentive policy in our study. An important prerequisite is that the treated and untreated groups are comparable and balanced to draw unbiased causal effects, which can be achieved by randomization in an experimental setting. However, in an observational study, treated and untreated groups may differ in observed and unobserved characteristics, which can affect the assignment of treatment and outcome. Several techniques such as multiple linear models and extensions, widely adopted in previous studies to identify the association between rewards and physicians' contribution [1], only control for observed confounders. Propensity score matching (PSM) [10,11] mimics the randomization process, reduces the confounding on the treatment assignment, and reaches a balanced group sample (with simulated counterfactual control units) in an observational setting. With the propensity score defined [12] as a conditional probability of assignment to treatment based on covariates, PSM largely reduces the matching process from multiple dimensions

XSL•FO
RenderX

to a single dimension [13]. The matching process is also appropriate for temporal case matching [14] and the classification of medical cases [15].

To mimic the experimental design with observational data, both *treatment* and temporal effects need to be considered. With the treated samples, a simple comparison of outcomes between pre- and post-treatment could be contaminated by the effects of other events (except the treatment) that occurred during both periods. For example, the seasonable factor may affect the change in the outcomes of OHCs [1]. With both treated and control groups, the comparison of post-treatment outcomes could capture more than the treatment effect, even after controlling for observed confounding [16]. The reason lies in the difference in unobserved attributes between the treated and untreated groups. Difference-in-differences (DID) [17], combined with PSM to achieve a parallel trend assumption (2 groups would show a parallel trend if neither of them experienced the treatment effect), is a useful method to reduce the impact of extraneous factors and selection bias. DID compares the average change over time in the outcome of the treated group with that of the control (untreated) group [18].

In many practical cases, the treatment assignment is (partially) determined by a cutoff or a threshold. This advanced design, known as a sharp regression discontinuity design (RDD) [19], is an extension of DID. Despite comparing the observations of the pre- and post-treated outcomes available in both groups, RDD also shows a good capacity for causal inference when the treatment assignment is deterministic and discontinuous at the cutoff. Comparing the observations close to the cutoff (local treatment effect) would achieve local *randomization* near to the threshold.

This study aims to estimate the causal effect of an incentive strategy on the PPI process of an OHC. In accordance with the previous discussion on methodology, PSs were estimated as the predicted probabilities of treatment (being awarded honorary titles based on covariates). Score matching was then conducted to reach a comparable control sample with the treatment group. Considering both the temporal and treatment dimensions of the policy, the DID idea was applied to compare the average pre- and posttreatment changes in the treated group (with the honorary title) with those in the untreated group. If the relationship between the covariates and the potential outcomes is *smooth* around the threshold (in covariates), the discontinuity (sharp jump) created by the treatment can provide local randomization. RDD would then be implemented appropriately to evaluate the causal effect of treatment (receiving the honorary title) on both outcomes (home page view and online consultation frequency) at the threshold.

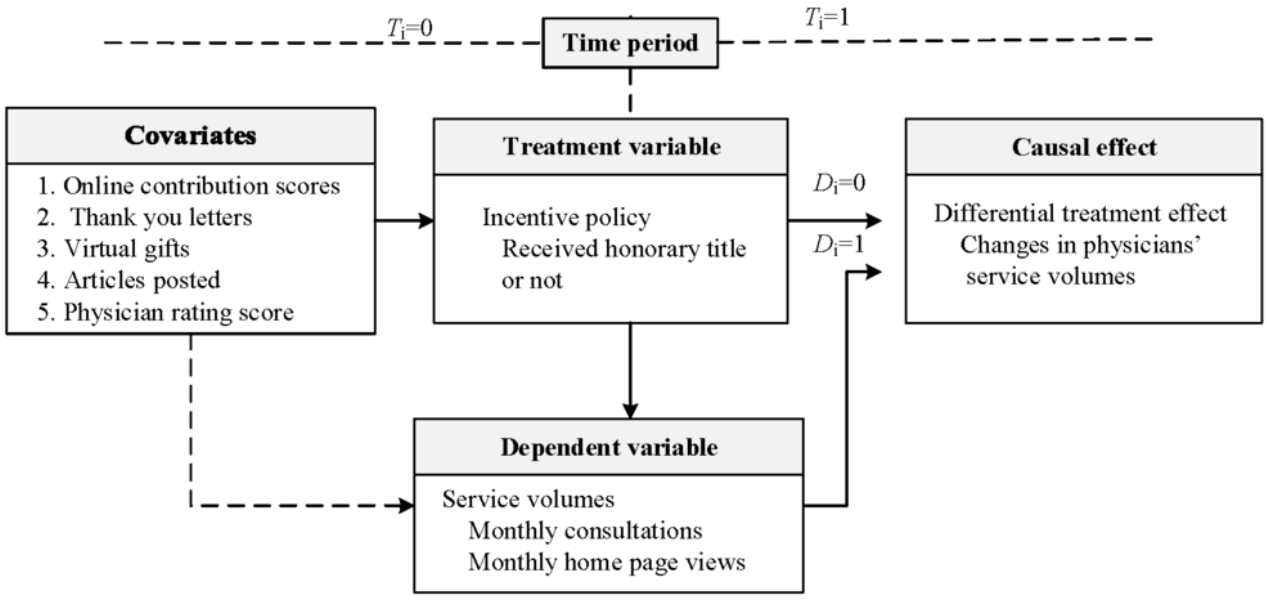
In summary, this study aimed to identify the average change in home page views and online consultation volume for physicians with the honorary title versus those without the honorary title. The investigation attempts to answer the following questions: (1) Does there exist the average treatment effect (ATE) of the honorary title on changes in outcomes (physicians' home page views and consultation volumes) in an OHC? (2) At the discontinuity of the *treatment* assignment, does the sharp effect of the incentive award exist on the outcomes, and can it be identified?

Methods

Research Models

This section demonstrates the research framework of this study. To investigate the differential treatment effect, the incentive policy was regarded as the treatment in the research design. The research framework was demonstrated to investigate the differential treatment effect of the incentive policy on physicians' service volumes, as illustrated in Figure 1. This service volume also reflected the patients' choice of physicians online.

Figure 1. Framework of the causal effect study on the outcome of physicians' service volumes, with D=1 indicating that physicians received honorary titles (in the treatment group), D=0 indicating the control group, T=1 indicating the postaward period, and T=0 indicating the preaward period.



XSL•FO
RenderX

First, 5 covariates were considered in this observational study. These variables are the physician rating score (PRS), the number of thank you letters, the number of virtual gifts, the number of online contribution scores, and the number of articles posted. The covariates represent the doctor's characteristics at a specific time point. Second, the honorary titles of the physician awarded from the OHC were viewed as the treatment. Within the causal inference mechanism, the objects of interest were those physicians assigned as the award recipient or not (D_i). This time factor helped in distinguishing the factors of the cause and effect among those variables. Third, to further investigate the dynamism of the effect on the changes in their service volumes, the time periods (T_i) were divided into those before awarding and after awarding. Fourth, the service volume is measured by 2 factors: the number of monthly consultations and the number of monthly home page reviews. These 2 factors reflect the online service (home page reviews) and the offline service (serviced patients of medical consultations), respectively. The number of patients that physicians serviced monthly (Patients #) and the number of their home page views monthly (Views #) [20] can be viewed as the proxies of the outcomes. Thus, these 2 proxy variables were set as the dependents. Moreover, the initial states of the numbers of those 5 covariates are measured as their cumulative before the examined period (June 25, 2017).

The definitions and measurements of all variables are presented in Table 1. $Patient_i\ (t)$ is measured as the number of online consultations (for patients) provided by the physician i in month t. $Views_i\ (t)$ is measured as the number of online home page views of the physician i in month t. The covariates were considered for case-control matching. PRS_i is measured as the PRS (by patients), which refers to the star scores listed on the OHC website. $Thank_i\ (t)$ is measured as the mean of the number of thank you letters of the physicians. $Gift_i\ (t)$ is measured as the mean of the number of gifts received by the physician i in month t. $Contr_i\ (t)$ is measured as the mean of the contribution score of physician i in the month t, which refers to the contribution scores listed on the website. $Aritcle_i\ (t)$ is measured as the mean of the number of physician articles. N is the number of physicians in the experimental data.

Table 1. Variable definitions and measurements.

Variables	Definitions	Measurements
Causal effect		
RDDeffect	Differential treatment effect	The causal effects of honorary titles incentives (treatment) in OHC[a] with RDD[b]
Dependent variable Y		
Patient $_i$ *(t)*	Number of monthly consultations	Number of online consultations (for patients) provided by the physician *i* in the month *t*
Views $_i$ *(t)*	Number of monthly home page views	Number of online home page views of the physician *i* in the month *t*
Treatment variable		
D_i	Receive honorary title or not	$D_i=1$ indicates physician *i* was titled as *2017 Good Doctor* in OHC, otherwise $D_i=0$
Time periods		
T_i	Preaward or postaward period	$T_i=0$ indicating the period before January 2018 (date of the honorary title), otherwise $T_i=0$
Covariates		
Contr $_i$ *(t)*	Contribution score	Total online contribution score calculated by OHC for physician *i* in the month *t*
Thank $_i$ *(t)*	Number of thank you letters	Total number of online thank you letters physician *i* received in the month *t*
Gift $_i$ *(t)*	Number of virtual gifts	Total number of online virtual gifts physician *i* received in the month *t*
Article $_i$ *(t)*	Number of articles posted	Number of articles posted by physician *i* in the month *t*
PRS_i *(t)*	PRS[c]	Mean of the rating scores by patients for physician *i* in the month *t*

[a]OHC: online health community.

[b]RDD: regression discontinuity design effect.

[c]PRS: physician rating score.

Data Collection

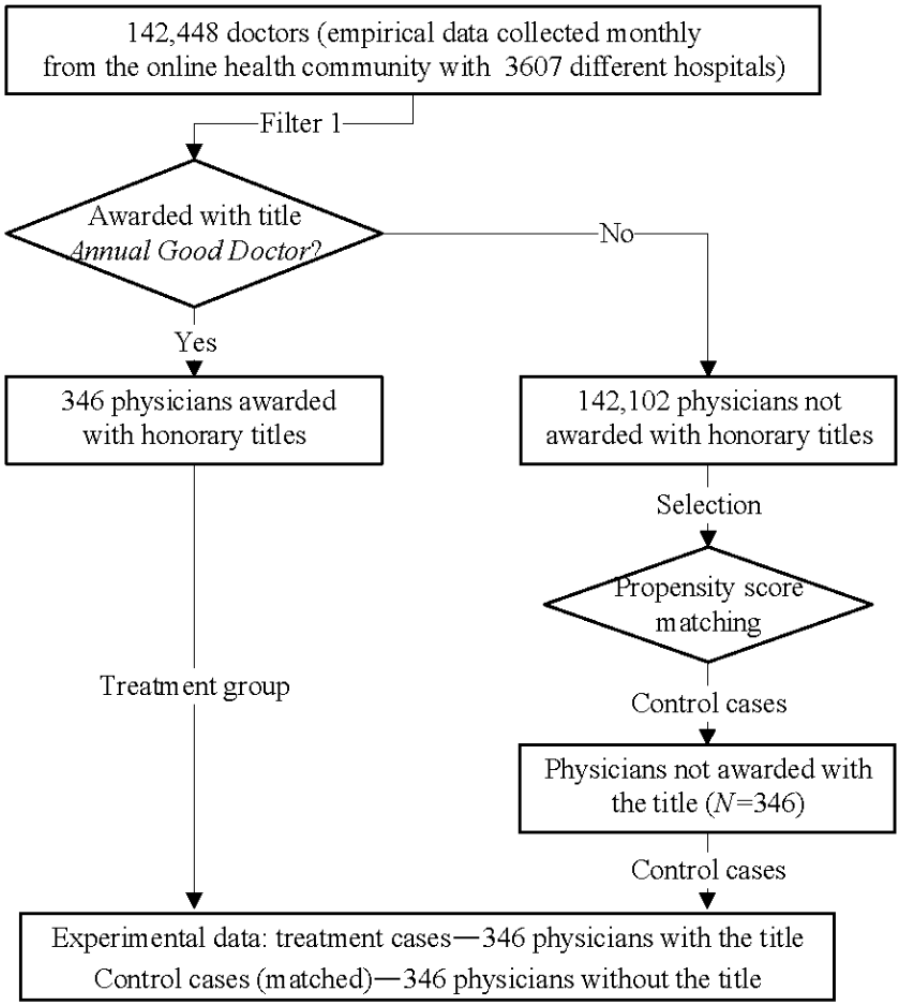
This study used existing records to conduct a retrospective study. The requirement for individual doctor consent was waived as the study did not impact clinical care and all data were deidentified. None of the data collected for the study were related to private information about the physicians.

Through the web crawler technology, a longitude data set from July 26, 2017, to June 26, 2018, was collected and filtered monthly from the Good Doctor website for the study. The collected data set contained the variables used in this study and other deidentified information. Among the online physicians, over 140,000 were involved and participated in OHC [1] with their personal profiles (including personal home pages). In January 2018, 346 of them were awarded the honorary title *2017 Annual Good Doctor.* Honorary titles were rated based on the number of consultations, the number of appointment referrals (patients), patient satisfaction with online services (the review

XSL•FO
RenderX

rating score), and other factors. The flowchart of data acquisition and filtering process is shown in Figure 2.

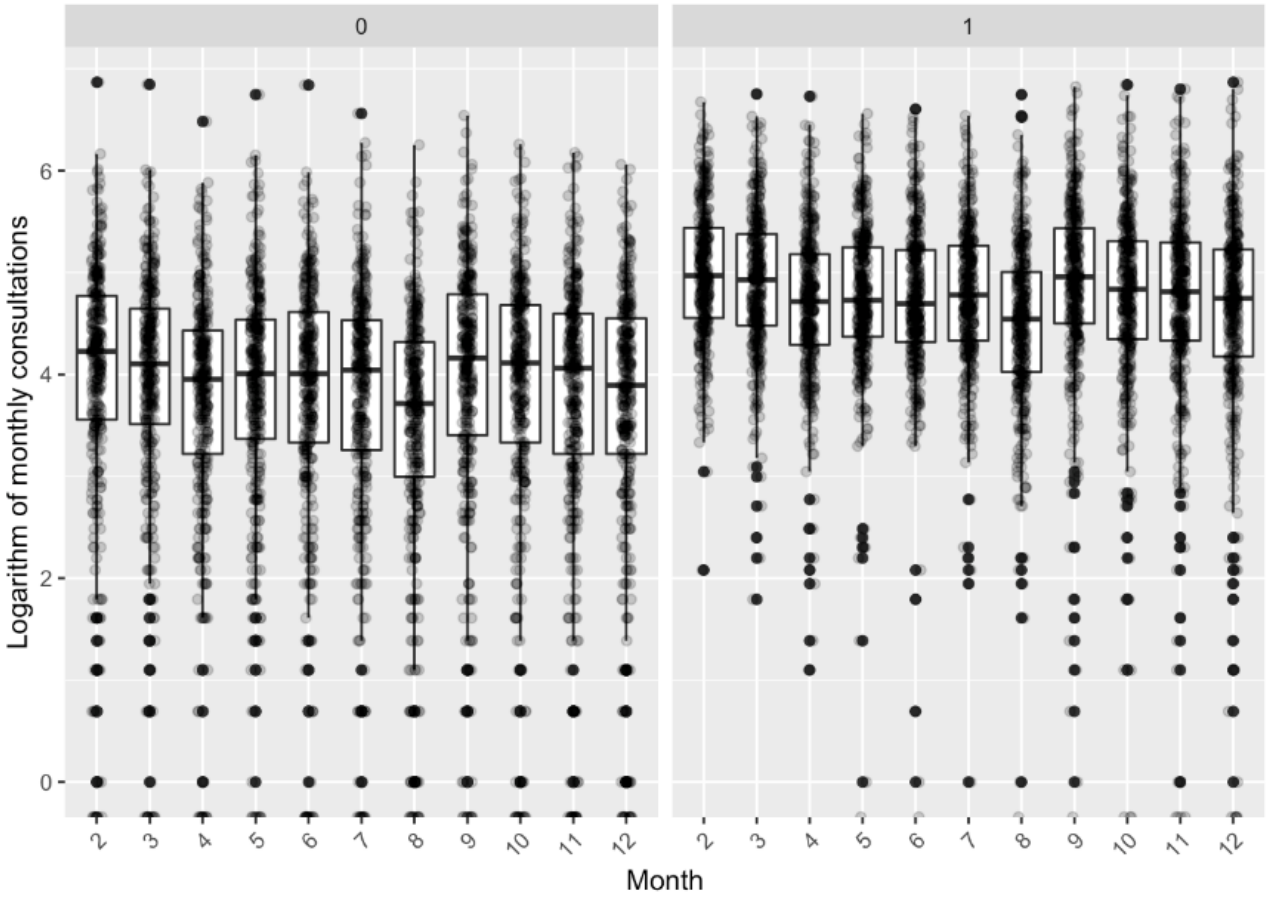
Figure 2. Data collection and preprocessing procedures.



During the preprocessing, outliers were removed from the original data. The design of the study is based on the propensity score matching (PSM) with 1:1 matching. The distribution of the logarithm of monthly consultations and that of the monthly home page views are illustrated in Figures 3 and 4. In these figures, 0 indicates the control group and 1 indicates the treatment group. The distribution results suggest that the causal effect cannot be estimated with the distributions directly.

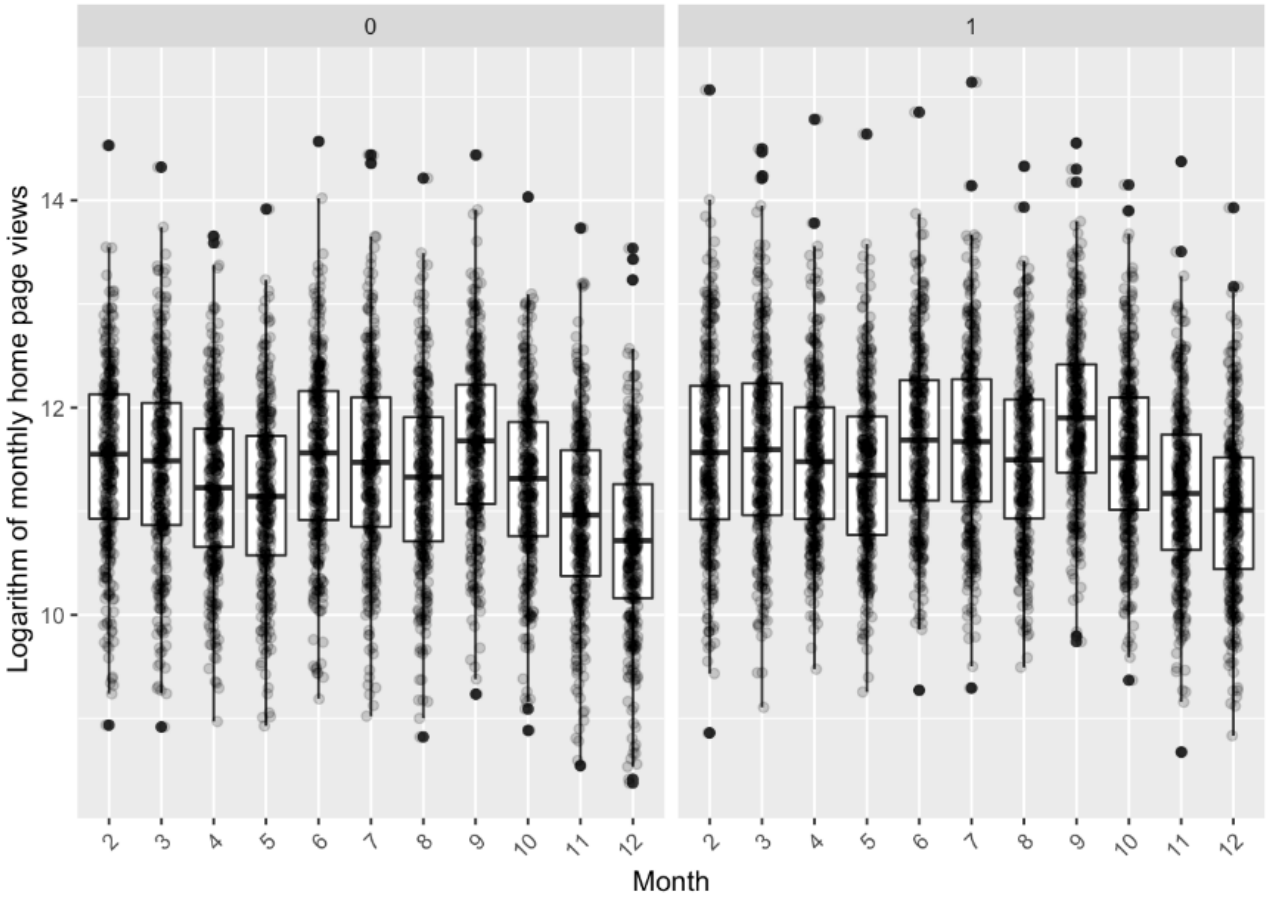
XSL•FO
RenderX

Figure 3. Distribution of the logarithm of monthly consultations, with 0 indicating the control group and 1 indicating the treatment group.



XSL•FO
RenderX

Figure 4. Distribution of the logarithm of monthly home page views, with 0 indicating the control group and 1 indicating the treatment group.



Estimation of RDD Effect

Through the RDD model, this study attempted to identify the ATE [21] of honorary titles for physicians on the changes in their service volumes. The theoretical contributions of this study not only lie in the design for estimating the RDD effect but also in combining it with the DID model through prediction of the counterfactual outcomes of the matching samples. The prediction of the RDD effect can be modeled as [22] follows:



Where  and  are the estimated expectation of the treatment group and the control group with the covariates x (ie, time period) and the outcomes y (ie, $Patient_i$ (t) and $Views_i$ (t)), respectively.

To derive a balanced control-treatment case data set, we trained the logistic regression to estimate the PS [23].



where β_0 is the coefficient of the constant term and β_j, j=1,..., 5, are the coefficients of the covariates, as detailed in Table 1. The error term  obeys a normal distribution with mean and variance .

We then matched control-treatment cases on pretreatment covariates with the PS. In the matching process, the scalar (N) can be preset for the number of matches needed (ie, the default value 1 is for one-to-one matching). More similar units are more likely to experience similar trends to meet the parallel path assumptions. Thus, the bias of the RDD effect can be reduced with such data sets of control-treatment pairs.

Results

In this section, we describe the results with descriptive statistics, the overlap assessment, and the differential treatment effect of honorary titles. Our findings provide empirical evidence that a regression discontinuity exists at the cutoff of the period.

Descriptive Statistics

The statistics of the empirical experimental data for each month are shown in Table 2. The state column illustrates the initial mean value of the experimental period, and the following 11 columns show the marginal changes per month. For example, PRS in August 2017 was 0.01, which means that it increased by 0.01 on average from the mean value (4.58) in the last month (July 2017). As the certification date of honorary titles (award) was January 20, 2018, the examined periods of treatment were from January 2018 to June 2018 (6 months).

From Table 2, the results show that for D=1 (award), monthly changes of preaward PRS were positive, whereas for D=0 (nonaward), there were negative changes. In terms of the thank you letters, the monthly data of D=1 (award) was higher than that of D=0 (nonaward). Similar trends were observed for virtual

XSL•FO
RenderX

gifts, posted articles, contribution scores, consulted patients, and home page views.

Table 2. Statistics of the empirical experimental data.

Variables	State[a]	T=0 (Monthly data[b])					T=1 (Monthly data)					
	July 2017	August 2017	September 2017	October 2017	November 2017	December 2017	January 2018	February 2018	March 2018	April 2018	May 2018	June 2018
D=1												
PRS[c]	4.58	0.01	0.008	0.008	0.007	0.002	0.006	–0.003	–0.003	0.002	–0.005	0.002
Thank[d]	139.3	6.6	6.4	5.8	4.8	5	5.4	4.1	4.4	5	6.2	6.8
Gift	711.4	25.4	22.2	26.1	20.7	17.3	19.6	19.6	18.1	16.6	14.6	14.2
Contr[e]	87370	3699	3684	3337	3464	3557	3823	3187	4351	4157	3960	3946
Article	63.77	1.01	1.19	1.05	0.87	1.07	1.42	1.49	1.29	1.05	1.57	0.89
Patient#[f]	5771	186	173	143	149	150	151	123	182	164	157	148
Views#[g]	4,099,352	173,977	184,098	145,770	128,537	180,300	181,598	148,503	214,388	153,022	105,504	86348
D=0												
PRS	4.493	–0.007	–0.005	–0.004	–0.001	–0.009	–0.003	–0.014	–0.009	–0.007	0.001	–0.02
Thank	121.8	4.2	4.1	3.3	2.8	3	3	2.4	2.4	0.5	5.98	3.62
Gift	611.5	14	10.9	14.8	10.3	9.2	9.6	9.9	8.6	8.2	8.1	6.9
Contr	78281	1515	1471	1331	1436	1475	1475	1200	1613	1457	1352	1500
Article	50.57	–2.02	0.61	0.52	0.45	0.55	0.47	0.37	0.44	0.65	0.2	0.31
Patient#[f]	5220	87	79	65	74	72	70	54	83	43	77	65
Views#[g]	4,882,225	143,432	141,268	107,423	101,340	151,260	153,008	120,576	164,884	30,649	115,264	65,165

[a]Initial state of the recorded data.

[b]Monthly change of the recorded data.

[c]PRS: physician rating score.

[d]Thank: Thank you letters.

[e]Contr: contribution score.

[f]Mean of the monthly consultations of the group.

[g]Mean of the monthly home page views of the group.

Overlap Assessment

The first step in analyzing the experimental data was to estimate the PSs using a logistic regression model with one main effect (on treatment) for each covariate. In the estimation of PS, the dispersion parameters for the binomial family were taken to be 1. With many covariates, it is difficult to examine the numeric diagnostics carefully for each covariate. As usual [24,25], graphical diagnostics are helpful for quickly assessing the covariate balance. Although the densities of raw treated and matched treated cases did not change, those of raw controls and matched controls changed significantly. The absolute standardized difference is defined as follows [26]:

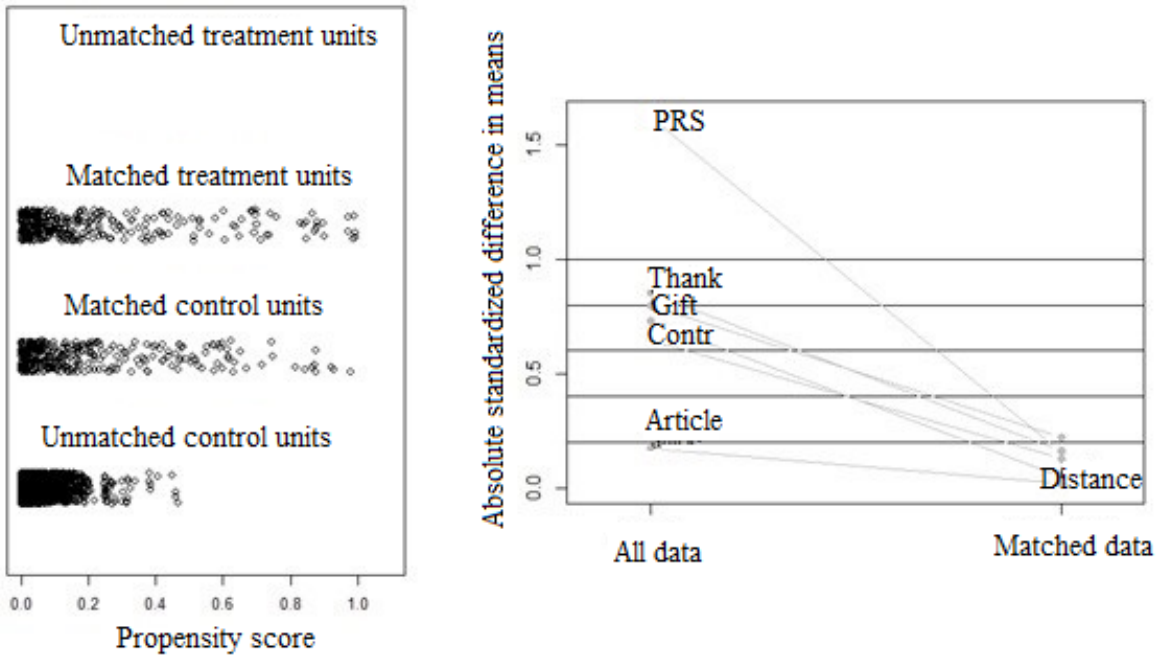


where *mean* (x_{treat}) and *mean* ($x_{control}$) denote the sample mean of the covariate in treated and control units (physicians), respectively, and s^2_{treat} and $s^2_{control}$ denote the sample variance of the covariate in treated and control units, respectively.

Figure 5 shows the weighted dots by their proportional size, which is also useful for stratification. Meanwhile, the absolute standardized difference is helpful for comparing the mean of continuous variables between the 2 groups, as illustrated in Figure 5 (right). The results show an adequate overlap of the PSs, with a good control match for each treated unit.

XSL•FO
RenderX

Figure 5. Distribution of propensity scores with experimental data. The left subfigure illustrates the distribution of propensity scores, and the right subfigure illustrates the absolute standardized difference in means of all data and matched data. Contr: Contribution score; PRS: physician rating score.



Differential Treatment Effect of Honorary Titles

This treatment effect illustrates the impact of honorary titles on the changes in physicians' service volumes. Two tests for the estimation of the impact of honorary titles were carried out with the panel data of 12 months. The period of control was investigated from July 2017 to December 2017, and the period of treatment was from January 2018 to June 2018. Among them, the first month of control was used to acquire the initial state of the system. The marginal quantity of the sequential periods was then acquired accordingly. The impacts of honorary titles on monthly consultations and home page views were analyzed using the panel data, as demonstrated in Table 3. The estimation of the number of monthly consultations before and after doctors receiving honorary titles is also visually illustrated in Figure 6. Similarly, this study also investigated the estimates of the number of monthly home page views, and the results are shown in Figure 7.

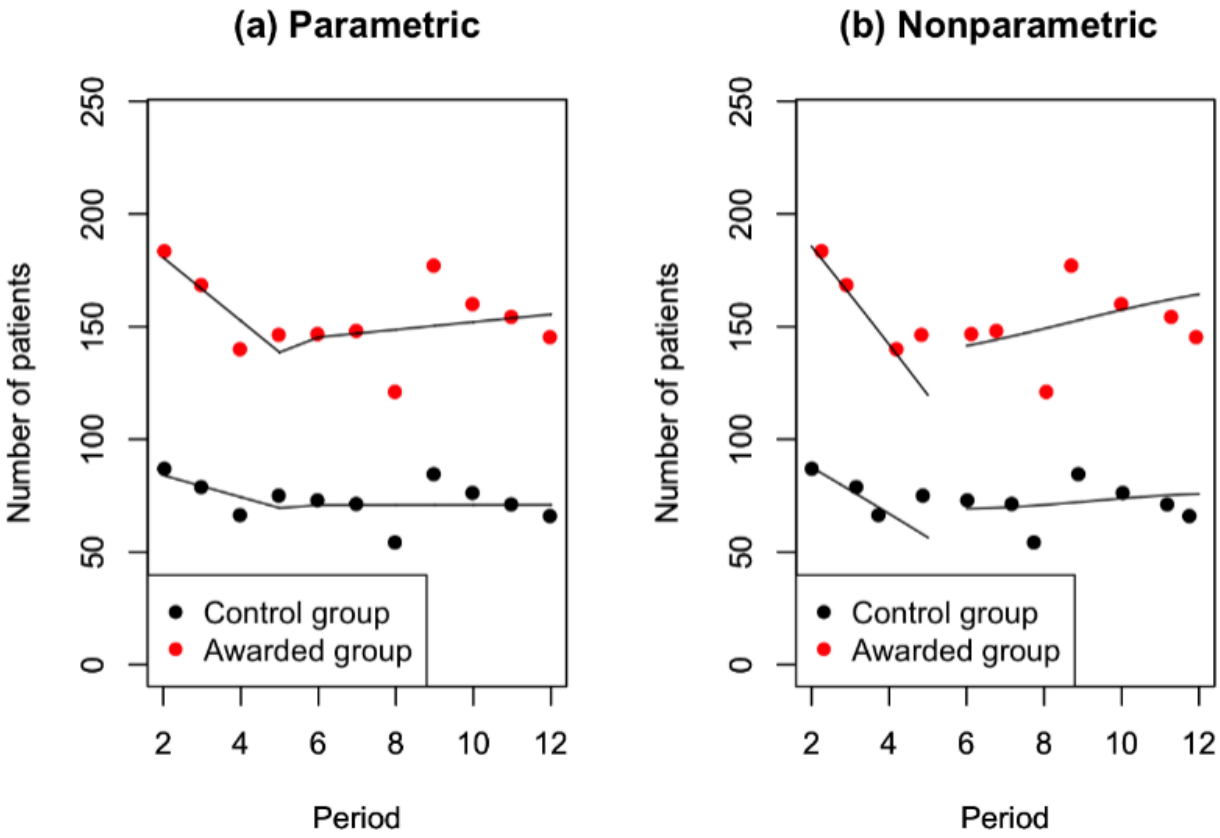
Table 3. Parametric and nonparametric estimates of the coefficient (the treatment effect).

Group and methods	Estimate	SE	N[a]	*P* value
Monthly consultations				
Control				
Parametric	6.113	5.970	3696	.31
Nonparametric	–9.286	8.087	1680	.25
Treatment				
Parametric	20.699	9.095	3806	.02
Nonparametric	5.133	10.277	2422	.62
Monthly home page views				
Control				
Parametric	80,666	10,704	3696	<.001
Nonparametric	66,814	13,535	2352	<.001
Treatment				
Parametric	84,340	14,535	3806	<.001
Nonparametric	73,351	18,386	2422	<.001

[a]N: number of physicians in the experimental data.

XSL•FO
RenderX

Figure 6. Estimation of the number of monthly consultations before and after physicians receive honorary titles.



In Table 3, the results show that the parametric estimates of the coefficient (the treatment effect of the honorary title) are significantly positive (P<.001) on monthly home page views. The estimates were 80,666 for the control group and 84,340 for the treatment group. Similar results were obtained for the nonparametric estimates of the coefficients, which were 66,814 for the control group and 73,351 for the treatment. These results indicate that the physicians with honorary titles had more monthly home page views than their counterparts. Meanwhile, honorary titles highlighted the physicians, accelerating the increase of the monthly home page views more greatly than the others. The number of monthly consultations of the physicians with the honorary titles was larger than those without the titles.

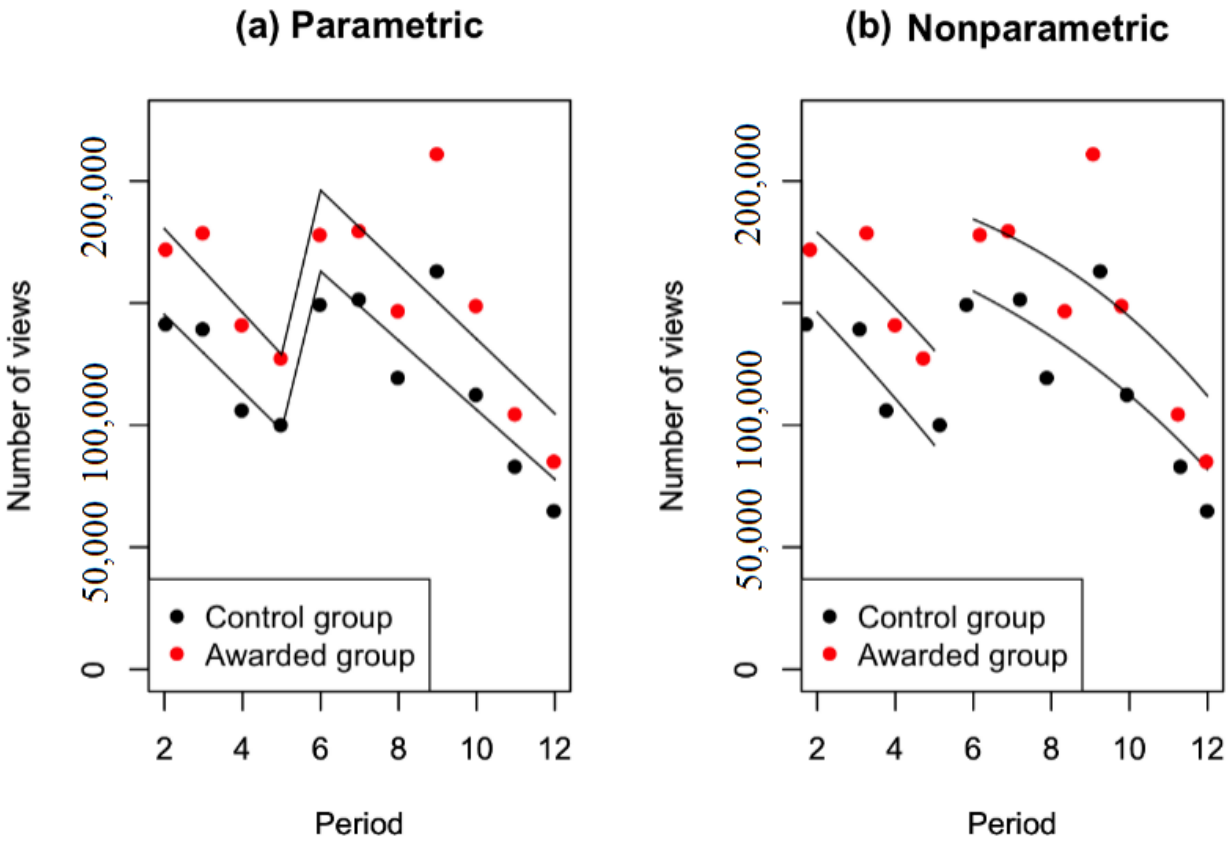
However, the parametric estimates of the effect on monthly consultations are positive (6.113 for the control and 20.699 for the treatment group) but not significant. The nonparametric estimates of the effect on monthly consultations were negative (–9.286) for the control and positive (5.133) for the treatment group (also insignificant). These results support our argument that the effect of honorary titles for physicians can significantly multiply the increases in the monthly home page views, yet they cannot significantly impact the monthly consultations.

In Figure 6, despite the decreasing treads, there was also a jump for monthly consultations during the period of receiving honorary titles. The results also indicated that 2 months of lag existed in the RDD. However, the jump in the volume of monthly consultations was insignificant or did not occur instantly. The honorary titles of physicians were awarded in January 2018, whereas a sharp regression discontinuity occurred in March 2018.

In Figure 7, the results show that honorary titles caused a jump in the monthly volume of home page views, that is, a sharp regression discontinuity. Moreover, the trends of the period before the jump were decreasing, which illustrated that the counterfactual observations of monthly volumes would be much less than those before this jump. Therefore, these results provide empirical evidence that regression discontinuity existed at the cutoff of the period.

XSL•FO
RenderX

Figure 7. Estimation of the number of monthly home page views before and after physicians receive honorary titles.



Discussion

Principal Findings

In this paper, we investigated the causal effect of honorary titles for physicians in terms of the changes in their service volumes in online health care communities. Monthly home page views and consultations were chosen as the 2 proxy variables for the outcomes. To identify the causal effect, multiple covariates, including physicians' rating scores, thank you letters, virtual gifts, online contribution scores, and posted articles, were considered for PS estimation. With pure randomization, the bias of effect estimation was reduced with the assigned samples. Through PSM, the results showed an adequate overlap of the PSs, with a good control match for each treated case. The results showed that honorary titles caused a jump in the monthly volumes of monthly consultations and home page views, specifically a sharp regression discontinuity.

Compared with the discontinuity regressions, the jump in monthly consultations is not as sharp as that of home page views. There may be many reasons for these sharp regression discontinuities. For example, if honorary titles are implemented for ranking and deploying physicians to users, then entitled physicians get more clicks because they appear first. However, consultation capacity is limited by the physician's schedule. This leads to the limit of the increase in monthly consultations for physicians. In contrast, there is no limit to the increase in home page views, which causes the jump in the monthly volumes of home page views to be much sharper than that of the monthly consultations. In total, changes in the monthly volumes of monthly consultations and home page views reflect the differential treatment effect of honorary titles on physicians' service volumes. The effect of the incentive policy with honorary titles is objectively estimated from the perspective of both online and offline medical services.

Although the causal study design was performed rigorously with the PSM method, this study still has some limitations. The number of articles posted by physicians was collected monthly for this study, with potential seasonable noise data. Meanwhile, the historical monthly data of the home page views and the online consultation frequency may be a cause of the honorary title for physicians in the current period. In this study, we introduced the time factor to distinguish the factors of the cause and effect among these variables. In future studies, their historical data can also be implemented as covariates to improve the balance of the comparing groups in the causal inference. To further investigate the proxy of physicians' service volumes, more characteristics could be abstracted from physicians' articles. In addition, falsification checks could also be implemented in future studies, including investigating whether covariates jump during the period of honorary titles (when jumps occur at cutoff thresholds) and what are the most reasonable time lags.

Conclusions

In OHCs, the platform management provider offers incentive policies (eg, honorary title) to encourage physicians' web-based involvement. However, the impact of the incentive policy on patients' online choice of physicians is still unclear. In this study, we investigated the causal effect of honorary titles for

XSL•FO
RenderX

physicians on changes in their service volumes, including monthly home page views and consultations. By stratifying the samples with the treatment (honorary title) and the period (pre- vs postaward), the RDD method was applied to identify the impact of the incentive policy on the service volumes of physicians. A sharp discontinuity was found at the time of the physician receiving the honorary title. The results showed that both parametric and nonparametric estimated coefficients were significantly positive (P<.001) for monthly home page views. The effect of honorary titles for physicians can significantly multiply the increases in the monthly home page views, yet its impact on monthly consultations was insignificant. Therefore, these results provide empirical evidence for our claim that regression discontinuity existed at the cutoff of the period. In the future, more investigations can be conducted to identify the time lag of the RDD effect.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 71601026) and the Key Research and Development Program of Sichuan Province (No. 2018SZ0114 and No. 2019YFS0271). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank the platform for the availability of the data.

Authors' Contributions

All coauthors are justifiably credited with authorship, according to the authorship criteria. The final approval is given by each coauthor. HY led the research and designed and performed all data analysis and interpretation of results. YW, JW, HQ, and YQ participated in the conception, design, and implementation of the study. JW and YQ made substantial contributions to data acquisition. HY and YW drafted the manuscript. YW, YQ, and MG revised the manuscript. MG confirmed the approach and provided essential editorial support for the manuscript.

Conflicts of Interest

None declared.

References

1. Wang J, Chiu Y, Yu H, Hsu Y. Understanding a nonlinear causal relationship between rewards and physicians' contributions in online health care communities: longitudinal study. J Med Internet Res 2017 Dec 21;19(12):e427 [FREE Full text] [doi: 10.2196/jmir.9082] [Medline: 29269344]
2. Yu H, Chen J, Wang J, Chiu Y, Qiu H, Wang L. Identification of the differential effect of city-level on the Gini coefficient of health service delivery in online health community. Int J Environ Res Public Health 2019 Jun 29;16(13):2314 [FREE Full text] [doi: 10.3390/ijerph16132314] [Medline: 31261952]
3. Gibson TB, Maclean JR, Carls GS, Moore BJ, Ehrlich ED, Fener V, et al. Engagement in health and wellness: an online incentive-based program. Prev Med Rep 2017 Sep;7:86-90 [FREE Full text] [doi: 10.1016/j.pmedr.2017.05.013] [Medline: 28626625]
4. Wang J, Du J, Chiu Y. Can online user reviews be more helpful? Evaluating and improving ranking approaches. Inf Manag 2020 Feb:103281 (forthcoming). [doi: 10.1016/j.im.2020.103281]
5. Guo X, Guo S, Vogel D, Li Y. Online healthcare community interaction dynamics. J Manag Sci Eng 2016 Dec;1(1):58-74. [doi: 10.3724/sp.j.1383.101004]
6. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of china's good doctor website. J Med Internet Res 2019 Mar 8;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
7. Heckman JJ. Causal parameters and policy analysis in economics: a twentieth century retrospective. Q J Econ 2000 Feb;115(1):45-97. [doi: 10.1162/003355300554674]
8. Contandriopoulos D, Perroux M. Fee increases and target income hypothesis: data from Quebec on physicians' compensation and service volumes. Healthcare Policy 2013 Nov 27;9(2):30-35. [doi: 10.12927/hcpol.2013.23613]
9. Athey S, Imbens GW. The state of applied econometrics: causality and policy evaluation. J Econ Perspect 2017 May 1;31(2):3-32. [doi: 10.1257/jep.31.2.3]
10. Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. J Am Stat Assoc 2000 Jun;95(450):573-585. [doi: 10.1080/01621459.2000.10474233]
11. Hill J, Reiter JP. Interval estimation for treatment effects using propensity score matching. Stat Med 2006 Jul 15;25(13):2230-2256. [doi: 10.1002/sim.2277] [Medline: 16220488]
12. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]

XSL•FO
RenderX

13. Austin PC, Stuart EA. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. Stat Methods Med Res 2017 Aug;26(4):1654-1670 [FREE Full text] [doi: 10.1177/0962280215584401] [Medline: 25934643]
14. Yu H, Shen J, Xu M. Temporal case matching with information value maximization for predicting physiological states. Inf Sci 2016 Nov;367-368:766-782. [doi: 10.1016/j.ins.2016.05.042]
15. Yu H, Shen J, Xu M. Resilient parallel similarity-based reasoning for classifying heterogeneous medical cases in MapReduce. Digit Commun Netw 2016 Aug;2(3):145-150. [doi: 10.1016/j.dcan.2016.07.003]
16. Qiu H, Yu H, Wang L, Yao Q, Wu S, Yin C, et al. Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. Sci Rep 2017 Nov 27;7(1):16417 [FREE Full text] [doi: 10.1038/s41598-017-16665-y] [Medline: 29180800]
17. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-differences models. Econometrica 2006 Mar;74(2):431-497. [doi: 10.1111/j.1468-0262.2006.00668.x]
18. Al-Zainal MH, Anvery S, Al-Jewair T. Clear aligner therapy may not prevent but may decrease the incidence of external root resorption compared to full fixed appliances. J Evid Based Dent Pract 2020 Jun;20(2):101438. [doi: 10.1016/j.jebdp.2020.101438] [Medline: 32473807]
19. Lee DS, Lemieux T. Regression discontinuity designs in economics. J Econ Lit 2010 Jun;48(2):281-355. [doi: 10.1257/jel.48.2.281]
20. Lagunes-García G, Rodríguez-González A, Prieto-Santamaría L, del Valle EP, Zanin M, Menasalvas-Ruiz E. How wikipedia disease information evolve over time? An analysis of disease-based articles changes. Inf Process Manag 2020 May;57(3):102225. [doi: 10.1016/j.ipm.2020.102225]
21. Imbens GW. Nonparametric estimation of average treatment effects under exogeneity: a review. Rev Econ Stat 2004 Feb;86(1):4-29. [doi: 10.1162/003465304323023651]
22. Imbens G, Kalyanaraman K. Optimal bandwidth choice for the regression discontinuity estimator. Rev Econ Stud 2011 Nov 24;79(3):933-959. [doi: 10.1093/restud/rdr043]
23. Dehejia RH, Wahba S. Propensity score-matching methods for nonexperimental causal studies. Rev Econ Stat 2002 Feb;84(1):151-161. [doi: 10.1162/003465302317331982]
24. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci 2010 Feb 1;25(1):1-21 [FREE Full text] [doi: 10.1214/09-STS313] [Medline: 20871802]
25. Keller B, Tipton E. Propensity score analysis in R: a software review. J Educ Behav Stat 2016 Jun;41(3):326-348. [doi: 10.3102/1076998616631744]
26. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]

Abbreviations

ATE: average treatment effect
DID: difference-in-differences
OHC: online health community
PPI: physician-patient interaction
PRS: physician rating score
PSM: propensity score matching
RDD: regression discontinuity design

Edited by G Eysenbach; submitted 03.03.20; peer-reviewed by D Carvalho; comments to author 24.03.20; revised version received 19.05.20; accepted 03.06.20; published 09.07.20.

Please cite as:
Yu H, Wang Y, Wang JN, Chiu YL, Qiu H, Gao M
Causal Effect of Honorary Titles on Physicians' Service Volumes in Online Health Communities: Retrospective Study
J Med Internet Res 2020;22(7):e18527
URL: https://www.jmir.org/2020/7/e18527
doi:10.2196/18527
PMID:32673232

©Haiyan Yu, Yali Wang, Jying-Nan Wang, Ya-Ling Chiu, Hang Qiu, Mingyue Gao. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 09.07.2020. This is an open-access article distributed under the terms of the

XSL•FO
RenderX

Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Gender, Soft Skills, and Patient Experience in Online Physician Reviews: A Large-Scale Text Analysis

Zackary Dunivin[1], BA; Lindsay Zadunayski[2], BA; Ujjwal Baskota[3], BS; Katie Siek[1], PhD; Jennifer Mankoff[4], PhD

[1]Indiana University, Bloomington, IN, United States

[2]Rensselaer Polytechnic Institute, Troy, NY, United States

[3]Jackson State University, Jackson, MS, United States

[4]University of Washington, Seattle, WA, United States

Corresponding Author:
Jennifer Mankoff, PhD
University of Washington
Bill & Melinda Gates Center for Computer Science and Engineering
3800 E Stevens Way NE
Seattle, WA, 98112
United States
Phone: 1 4125677720
Email: jmankoff@acm.org

Abstract

Background: Online physician reviews are an important source of information for prospective patients. In addition, they represent an untapped resource for studying the effects of gender on the doctor-patient relationship. Understanding gender differences in online reviews is important because it may impact the value of those reviews to patients. Documenting gender differences in patient experience may also help to improve the doctor-patient relationship. This is the first large-scale study of physician reviews to extensively investigate gender bias in online reviews or offer recommendations for improvements to online review systems to correct for gender bias and aid patients in selecting a physician.

Objective: This study examines 154,305 reviews from across the United States for all medical specialties. Our analysis includes a qualitative and quantitative examination of review content and physician rating with regard to doctor and reviewer gender.

Methods: A total of 154,305 reviews were sampled from Google Place reviews. Reviewer and doctor gender were inferred from names. Reviews were coded for overall patient experience (negative or positive) by collapsing a 5-star scale and coded for general categories (process, positive/negative soft skills), which were further subdivided into themes. Computational text processing methods were employed to apply this codebook to the entire data set, rendering it tractable to quantitative methods. Specifically, we estimated binary regression models to examine relationships between physician rating, patient experience themes, physician gender, and reviewer gender).

Results: Female reviewers wrote 60% more reviews than men. Male reviewers were more likely to give negative reviews (odds ratio [OR] 1.15, 95% CI 1.10-1.19; *P*<.001). Reviews of female physicians were considerably more negative than those of male physicians (OR 1.99, 95% CI 1.94-2.14; *P*<.001). Soft skills were more likely to be mentioned in the reviews written by female reviewers and about female physicians. Negative reviews of female doctors were more likely to mention candor (OR 1.61, 95% CI 1.42-1.82; *P*<.001) and amicability (OR 1.63, 95% CI 1.47-1.90; *P*<.001). Disrespect was associated with both female physicians (OR 1.42, 95% CI 1.35-1.51; *P*<.001) and female reviewers (OR 1.27, 95% CI 1.19-1.35; *P*<.001). Female patients were less likely to report disrespect from female doctors than expected from the base ORs (OR 1.19, 95% CI 1.04-1.32; *P*=.008), but this effect overrode only the effect for female reviewers.

Conclusions: This work reinforces findings in the extensive literature on gender differences and gender bias in patient-physician interaction. Its novel contribution lies in highlighting gender differences in online reviews. These reviews inform patients' choice of doctor and thus affect both patients and physicians. The evidence of gender bias documented here suggests review sites may be improved by providing information about gender differences, controlling for gender when presenting composite ratings for physicians, and helping users write less biased reviews.

(J Med Internet Res 2020;22(7):e14455) doi:10.2196/14455

XSL•FO
RenderX

KEYWORDS

reviews; physician-patient relationship; gender; soft-skills

Introduction

Background

Physician review sites are relatively new and were initially greeted with concern by some in the medical community. In particular, some physicians were critical of the lack of transparency in composite statistics [1] and were concerned that online reviews could harm their careers [2]—perhaps unfairly [3,4]. Although ratings are generally high [2,3,5,6], negative ratings undoubtedly influence patient behavior [7] and impact doctors [1,8]. Some doctors have attempted to *gag* patients by contractually prohibiting them from writing online reviews [9].

Most studies of online physician reviews have focused on portals such as HealthGrades [10], RateMDs [11], Vitals [12], and Yelp [13]. Studies tend to have a small sample size, analyzing approximately 5400 reviews [6]. Many studies aim to understand the factors that influence quantitative physician ratings. The qualitative analysis of 712 reviews by López et al [5] established thematic categories that tended to appear in reviews. Paul et al [14] replicated and expanded this work with a natural language processing (NLP) approach, which they applied to the text of 50,000 online reviews downloaded from RateMDs [11]. Their novel joint topic–sentiment modeling approach found that certain textual accounts of interpersonal skills such as *rude*, *arrogant*, and *condescending* are strongly associated with negative reviews, and drew attention to the role of patient experience of bureaucratic process in reviews, noting that these experiences were often reflected in reviews. Wallace et al [15] expanded on the work by López et al [5] and Paul et al [14] by analyzing 60,000 reviews to identify relationships between overall rating, health outcomes, and cost of care. To date, the only study to investigate the relationship between physician reviews and gender is that by Nwachukwu et al [16] on surgeon quality in sports medicine. They found that communication style influenced the valence of ratings for top- and bottom-tier surgeons and that female surgeons typically had higher ratings [16]. These and other studies of online reviews endeavor to understand how clinical experiences influence patient satisfaction and health outcomes. However, they tend to overlook or minimize questions about whether online review data reflect real experiences of medical care. Reviews may not be representative of the public or reflect demographic variation in health care utilization; indeed, doctor reviews are typically written by educated, younger, affluent, and healthier people [5]. However, a study comparing ratings of over 3000 physicians with licensing data showed a clear relationship between doctor quality and ratings [3].

Little research has studied the impact of gender or other demographic factors on the content and ratings of online physician reviews. Although qualitative studies of doctor-patient relationships have considered both negative and positive experiences [5], including the impact of demographics [17-19], the nature of these studies makes it difficult to estimate the size or scope of gender and other demographic variation in online physician reviews. Furthermore, gender differences, in particular biased interpretations of clinical experiences based on gender stereotypes, may impact online review content, which in turn may negatively impact both patients and physicians and perpetuate false gender stereotypes. The large-scale systematic study we present here documents gender differences in patient reviews with respect to both patients and doctors and proposes improvements for online review systems that could help reduce these disparities, thus improving information quality.

Gender and Health Care

Although we know little about gender in the context of online reviews, gender has been studied extensively in the social sciences for over half a century. Much of this work investigates the role of gender in medical care and health systems more generally.

Gender is a cultural construct that affects people's expectations and actions [20]. In social contexts and practices [21], gender is *assessed* independently of one's identity [22]. Thus, any name appearing in online text is likely to be interpreted in terms of the man/woman binary, which is reflected in the use of gender in the current doctor review literature.

The expectations of one's behavior differ depending on one's assessed gender. Indeed, leadership traits praised in men are penalized in women, while traditional *feminine* behavior is seen as ineffective [23]. When writing references, men are described with more standout and ability-based words and fewer *grindstone* words (eg, *hardworking*, *conscientious*) [24], and women are described with more communal words [25]. Even when all factors are controlled, people rated teachers differently on hard skills (eg, *promptness*, *fairness*) and soft skills depending on the gender portrayed by the instructor [26]. Thus, bias may influence review content even when performance is identical.

In health care, gender differences influence doctors' communication with patients [27-29], doctor and patient trust [30], and even diagnosis error rates [27]. Female *doctors* are seen as partners and more involved in the patient-doctor relationship, whereas female *patients* are treated with more condescension [27,29,31], have their concerns dismissed [29] and credibility doubted [27]. Conversely, patient satisfaction is dependent on more caring communication styles for women than for men [32]. However, many studies documenting these trends are small in scale or have weak evidence [33]. One exception is a study of over 10,000 people experiencing long-term illness in Sweden, where women reported being blamed, interrupted, disbelieved, doubted, and regarded as stupid [19]. These gender differences are likely to impact review scores, as lower patient satisfaction is correlated with high physician dominance, which can manifest itself in gendered actions (eg, poor information sharing and use of medical jargon) [34]. In this study, we investigate how these gender biases are represented in online reviews, which affect patients, physicians, and people using the reviews.

XSL•FO
RenderX

Study Design and Motivation

The goal of this study was to broaden and deepen our understanding of the impact of gender bias and other gender differences on online physician reviews. We leveraged reports of patient sentiments about their doctors through a large-scale analysis of online reviews. The Google Place review data analyzed here allowed us to identify patient and reviewer gender and characterize patient sentiment or experience in terms of both overall quality (a reviewer-entered Likert-type scale) and thematic content. Specifically, we formulated the following hypotheses (H):

- H1a—Physician ratings and physician gender: female physicians are more likely to receive negative reviews than male physicians.
- H1b—Physician ratings and reviewer gender: Female reviewers are more likely to report negative experiences with doctors.
- H2a—Soft skills and physician gender: Female physicians are more likely to receive criticism mentioning soft skills than male physicians.
- H2b—Soft skills and reviewer gender: Female reviewers are more likely to mention soft (interpersonal) skills in negative reviews.
- H3—Reviewer gender and physician gender: Female reviewers are more likely to report negative experiences with male doctors.

Hypotheses H1a and H1b relate to physician gender and reflect the findings of prior work on gender inequalities in reviewing in other fields [24-26]. Hypotheses H2a, H2b, and H3 are based on prior work documenting gender differences in clinical encounters [27,28].

Our approach comprised both qualitative and quantitative perspectives on review content that mutually informed one another throughout the research process. We collected and analyzed a corpus of 154,305 reviews of doctors that constitutes a large nationally representative sample of physician reviews across all medical subfields and clinical contexts. Our focus was specifically on characterizing the differences in experience quality and content as they relate to reviewer and doctor gender.

This study contributes to the larger body of work on the impact of gender on clinical interactions and provides insight into what patients value in their doctors. Furthermore, we add to the small but growing body of literature that seeks to develop a general understanding of online reviews and the systems that collect and display them. Our results must be interpreted with caution due to the unstructured and often short nature of the patient narratives in reviews, the relative crudeness of the NLP techniques employed (as compared with human interpretation), and selection biases introduced by nonrepresentative variation in demographic characteristics of reviews and the types of experiences that motivate patients to write reviews. Such selection biases are almost unknown although we make a novel contribution here.

Online reviews provide an opportunity to learn, at scale, about patients' perceptions of their doctors. Our findings have direct implications for the design of review sites and the presentation of search results. We argued that there are a number of useful ways in which gender differences can be reflected and potentially corrected in the presentation of information on the internet.

Methods

This study is a large-scale ($N>10^5$) text analysis of reviews of US physicians in the form of social media trace data. Social media trace data have the advantage of feature richness, (often) wide availability, and relation to genuine human social behavior outside of an experimental or survey context. The analytic approach we pursued in this study followed a similar process to the hybrid ethnographic and NLP approach advocated by Nelson [35]. The methodological framework is a recursive process whereby qualitative text analysis (sometimes called *deep reading* or *content analysis*) informs computational feature extraction, which is then evaluated through further qualitative analysis. This refinement process continues until the patterns in the computationally derived features match the intuitions and examples accumulated through qualitative analysis fairly well. Ultimately, a quantitative analysis, in this case regression modeling, is applied to validate large-scale patterns in the data. The rest of this section details the specifics of this approach, the review sampling process, inference of reviewer and physician gender, and modeling of associations and interactions among the variables of interest.

Physician Review Collection

To examine how gender influences patient experience at scale, we sought a representative sample of reviews of US physicians. As gender and other demographic variables are rare in social media, we sought data that contained physician and reviewer names, which we used as a proxy for gender.

Review Collection Application Programming Interface

After exploring possible sources of physician reviews considering various application programming interface (API) features and use in prior work, we selected the Google *Places* API [36]. The API provides access to patient and physician names, which we leveraged to infer gender, as well as a broad range of areas and specialties. The Google My Business API has a 5-review limit for any particular doctor. Unfortunately, the API documentation does not provide information on how these reviews are selected. We can be fairly confident, however, that reviewer gender is not a factor. Thus, it is unlikely that this introduces bias into the sample with regard to the variables of interest. Furthermore, we took measures to ensure that our personal search histories did not influence review collection.

Geographical Sampling

Reviews, physicians, and practices are likely to vary by location. Differences in locale, such as ruralness or urbanity, can influence health outcomes and care options, as can regional differences. For instance, in the United States, mortality rates of particular conditions have been shown to differ greatly from state to state [37]. Samples were taken across states from multiple regions of the United States using the Google Places API to control for the effect of locale. We steadily increased latitude and longitude

XSL•FO
RenderX

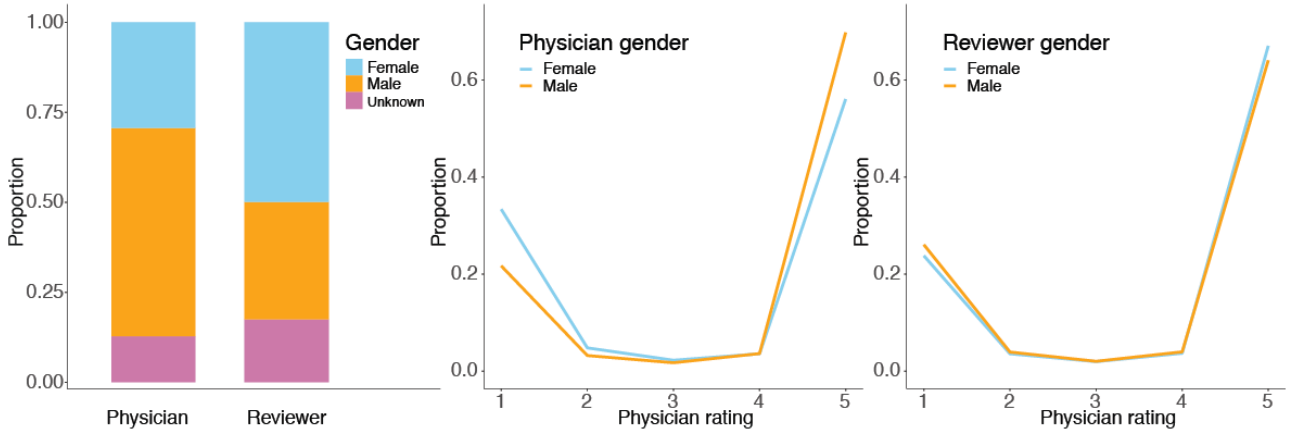
intervals throughout each state with a 10,000-m radius to capture both urban and rural regions.

Data Summary

Reviews returned by the Google My Business API were either for a *place* (such as a practice with multiple doctors) or a *physician*. The API provided additional *review-specific* data for each review: a 5-point Likert-type rating assigned by the reviewer to the doctor or practice, doctor name, reviewer name, the location of the practice, and the text content of the review. The reviewer and physician gender were determined automatically using third-party software described in the following section.

The collection strategy described in this section yielded 154,305 reviews of physicians across the United States. These reviews spanned 2007 to 2017. Doctor ratings were highly polarized, exhibiting a U-shaped distribution (Figure 1). Of the reviews collected, 46,605 were rated 1 star or 2 stars (*negative* reviews) and 107,700 were rated 4 stars or 5 stars (*positive* reviews). Another 3,208 reviews were rated 3 stars and were omitted from our analyses. We did not screen for particular specialties.

Figure 1. Left: Gender distribution in the complete data set (N=154,305). The unknown category represents clinics and names that were androgynous or unknown to the gender classifier. Middle: Distribution of physician ratings by physician gender (N=137,329). Right: Distribution of physician ratings by reviewer gender (N=129,985).



The mean length of positive reviews was 50 words, while the mean length of negative reviews was 100 word (both follow heavy-tailed distributions). The distribution of inferred gender for both doctors and reviewers is shown in Figure 1. Note that we expected to see a fairly high rate of unknown gender in these data because some reviews are for medical practices that include multiple physicians. In addition, some physicians' gender could not be identified. Given the presence of nonperson entities among *physicians*, it is perhaps surprising that reviewers have a greater rate of gender ambiguity. This likely reflects typos and pseudonyms among reviewers, who have a weaker incentive to use the correct name. However, the high rate of gender detection suggests that this should not be a great concern, as almost all reviewers use *real* names, not screen names. We were less certain about the rate of pseudonymous users but proceeded on the assumption that even when pseudonyms are used, they accurately reflect the reviewer's gender.

Qualitative Coding

Sampling Strategy

To guide our quantitative analysis and support the validation of our approach, we additionally selected a small sample of reviews for hand coding. A total of 200 reviews were selected for hand coding using stratified sampling for a distribution of 60% negative (3/5 from each state) and 40% positive (2/5 from each state) reviews because our initial read-throughs indicated that negative verbiage was less prevalent than positive verbiage in our sample.

As this sample was relatively small in comparison with the number of total reviews collected, we used search terms intended to select for specialties that would help us focus on specific patient genders. We selected 50 reviews by mostly female reviewers using *maternal*, *fetal*, *fertility*, *natal* as search terms for clinic name and *maternity*, *fetal*, *miscarriage*, *trimester*, *fertility*, *natal*, *birth*, *pregnancy*, *delivery*, *baby*, midwife, *Ob/Gyn* in the review. We also selected 50 reviews by mostly male reviewers using *cancer*, *prostate* in the review itself, along with 100 reviews of both male and female urology reviewers. All of these reviews were manually assessed to ensure that they represented the assigned group. Although pregnancy and prostate cancer are not comparable medically, we chose them because they selected for patients by biological sex. They are the only common conditions that affect only one biological sex. This reduces uncertainty in our interpretation of the qualitative data.

We analyzed these reviews to construct a codebook (Multimedia Appendix 1) and develop an intuition for the patterns of thematic content and gendered interaction in physician reviews. Throughout this paper, we reflect on these patterns or illustrate particular situations by quoting largely from the reviews in this sample. The intuitions we developed through our qualitative and quantitative analyses have led us to conclude that situations reported in exemplar quotations generalize beyond the women's health or urological contexts. Accordingly, our quantitative analysis and some follow-up quantitative investigations span the entire data set.

When we quote reviews, they will be cited with the following descriptors: physician gender (Male; Female; Unknown), star rating (1-star, 2-star,..., 5-star), doctor type (O=OBGYN; U=UROLOGY). For example, (Male; 5-star; O5654) is an

XSL•FO
RenderX

OBGYN review of a male physician with a 5-star rating. Quotes are exemplary of many similar statements found in the reviews, with some synthesis and paraphrasing to support anonymization.

Codebook Construction

We developed a codebook to regulate our analysis of the reviews. These codes were divided into 2 parts: regular codes and context codes. Context codes relate to demographic information, doctor gender, or specialty, while regular codes reflect the content of the review, for example, *professional* or *rude*.

We began with codes identified by López et al [5]. We also used iterative open coding to identify common categories of statements in the data. After coding each of the qualitative samples in Text Analysis Markup System [38], we took each coded section and created an affinity diagram, grouping similar segments independent of the initial codebook to represent the content of the reviews most accurately. These were then grouped into overarching categories that could be used for the analysis of the full data set. The process resulted in 7 main thematic areas: *process*, *candor*, *trust*, *investment*, *amicability*, *indifference*, and *disrespect*. The themes represented 2 general categories: those pertaining to nursing and administrative *process* (process) and those pertaining to soft skills. The latter group was further subdivided into *positive* (candor, trust, investment, amicability) and *negative* (indifference, disrespect) soft skills. After developing the codebook, 3 authors coded 20 reviews to assess interrater reliability using Cohen kappa (Table 1) and then refined the categories to improve agreement.

Table 1. Themes that emerged in affinity diagramming and examples of the associated terms in the dictionaries used in the quantitative analysis.

Theme	Sample terms	Kappa[a]	Count	Accuracy[b]	Precision	Recall
Positive soft skills		0.61	113	0.77	0.74	0.90
Candor	Honest, explain, answer, direct	0.95	41	0.84	0.57	0.76
Trust	Support, safe, reassure, comfort	0.98	27	0.42	0.41	0.63
Investment	Respect, care, compassion, listen	0.93	65	0.42	0.60	0.74
Amicability	Warm, friendly, personable, funny	0.94	31	0.82	0.46	0.84
Negative soft skills		0.53	25	0.87	0.44	0.16
Indifference	Cold, dismiss, ignored, abandoned	0.74	16	0.87	0.50	0.16
Disrespect	Rude, harass, condescending, arrogant	0.40	25	0.92	0.50	0.44
Process	Cost, nurse, staff, wait	0.83	115	0.87	0.84	0.96

[a]Kappa represents interrater agreement (on 20 reviews).

[b]Accuracy, precision, and recall, respectively, on a random sample (N=100) of 200 total reviews. A review is labeled as pertaining to a theme if at least one of the words in the theme in dictionary is presented in the review. Note the infrequency of negative soft skills (16 and 25 for indifference and disrespect, respectively), contributing to low precision and recall.

Computational Feature Extraction

Gender Detection

A third-party Python library [39] was used to infer the probable gender of physicians and reviewers based on their name. Although not everyone identifies within the female-male gender binary [22], currently, gender is typically assessed and reacted to with respect to this binary [21], so a binarized gender of reviewers and physicians was extracted using the names provided. Although a binary definition of gender does not capture the spectrum of gender and gender relations, capturing a more complex understanding of gender is infeasible given the scope of our data and the lack of identifying information for reviewers and doctors beyond name.

To verify the accuracy of the gender inference procedure, we took a random sample of 200 reviews and compared automatically inferred gender with our human-coded gender determinations informed by close reading of the reviews informed by name and gender pronouns. Automated physician gender inference was 98% accurate. The accuracy of reviewer gender was not examined because the only available measure was the reviewer's name.

The distribution of genders is shown in Figure 1 (left). Gender could not be inferred for 12.8% of physicians and 17.5% of patients. A logistic regression model (not shown) estimated with high confidence that female physicians are 1.41 times more likely to be reviewed by female patients (P<.001). A second model (also not shown) indicates that physicians are 4.36 times more likely to be reviewed by patients of the same gender (P<.001). These figures do not necessarily represent the actual gender distribution of patients *seen* by doctors, as there may be selection bias for or against intragender reviews.

Thematic Content of Reviews

Informed by the qualitative analysis, dictionaries were developed relating to the 7 themes (*process*, *candor*, *trust*, *investment*, *amicability*, *indifference*, and *disrespect*) identified in the qualitative coding. Review text was stemmed using Porter stemmer and tagged with a binary label for each theme if the review mentioned a word in the theme's dictionary. Working separately, we coded 200 reviews to assess the ability of the codebook to identify each of the themes. Table 1 presents summaries of the themes, their kappa statistics, and the performance of the binary variables when applied as a single-feature classifier against a random test set of 100 hand-coded reviews that were not used to inform the codebook.

XSL•FO
RenderX

Tables 2 and 3 display the frequencies with which the terms appear in the corpus and the proportion of reviews for each combination of gender/physician rating for physician gender and patient gender separately.

Table 2. Prevalence of themes by physician gender.

Theme	Doctors (female, n=36,847; male, n=74,189)[a,b]			
	Female (negative; n=13,874), n (%)[c]	Female (positive; n=22,973), n (%)[c]	Male (negative; n=17,906), n (%)[c]	Male (positive; n=56,283), n (%)[c]
Positive soft skills	7403 (53.36)	16,984 (73.93)	7530 (42.05)	37,993, (67.50)
Candor	2756 (19.86)	3531 (15.37)	2227 (12.44)	8431 (14.98)
Trust	754 (5.43)	2725 (11.86)	745 (4.16)	5566 (9.89)
Investment	4724 (34.05)	12,266 (53.39)	4830 (26.97)	26,032 (46.25)
Amicability	1928 (13.90)	7481 (32.56)	1621 (9.05)	16,047 (28.51)
Negative soft skills	4629 (33.36)	488 (2.12)	4505 (25.16)	721 (1.28)
Indifference	868 (6.26)	163 (0.71)	836 (4.67)	298 (0.53)
Disrespect	4112 (29.64)	343 (1.49)	3942 (22.01)	442 (0.79)
Process	9330 (67.25)	10,099 (43.96)	9981 (55.74)	23,837 (42.35)

[a]Many reviews contain multiple themes, so the overall rows (bold) have smaller numbers than the sum of themes would indicate. This table includes only those reviews for which a gender was assigned (n=111,036).

[b]The physician rating is denoted as negative/positive.

[c]Percentages represent the proportion of reviews containing the theme for that particular gender/rating combination.

Table 3. Prevalence of themes by reviewer gender.

Theme	Reviewers (female, n=67,857; male, n=43,179)[a,b]			
	Female (negative; n=18,780), n (%)[c]	Female (positive; n=49,077), n (%)[c]	Male (negative; n=13,000), n (%)[c]	Male (positive; n=30,179), n (%)[c]
Positive soft skills	9102 (48.47)	35,334 (72.00)	5831 (44.85)	19,643 (65.09)
Candor	3104 (16.52)	7815 (15.92)	1879 (14.45)	4147 (13.74)
Trust	906 (4.82)	5636 (11.48)	593 (4.56)	2655 (8.80)
Investment	5834 (31.06)	24,994 (50.93)	3720 (28.62)	13,304 (44.08)
Amicability	2231 (11.88)	15,630 (31.85)	1318 (10.14)	7898 (26.17)
Negative soft skills	5868 (31.25)	790 (1.61)	3266 (25.12)	419 (1.39)
Indifference	1101 (5.86)	305 (0.62)	603 (4.64)	156 (0.52)
Disrespect	5178 (27.57)	513 (1.05)	2876 (22.12)	272 (0.90)
Process	11,469 (61.07)	21,728 (44.27)	7842 (60.32)	12,208 (40.45)

[a]Many reviews contain multiple themes, so the overall rows (bold) have smaller numbers than the sum of themes would indicate. This table includes only those reviews for which a gender was assigned (N=111,036).

[b]The physician rating is denoted as negative/positive.

[c]Percentages represent the proportion of reviews containing the theme for that particular gender/rating combination.

Dictionary-based text analysis is crude in that it cannot determine valence, that is, the dictionary approach cannot distinguish between the phrase “Dr. X listens,” for example, “Great bedside manner. She was kind and listened to everything I had to say” (Female; 5-star; 15180), and the many variants of its negation, for example, “I never felt like she truly listened” (Female; 1-star; 25068). Positive soft skills are more likely to be negated than negative ones, largely because double negatives are far less common than single negatives in English. This is borne out by the associations between each theme and *negative reviews* in the set of models summarized in Table 4. Determining valence is further complicated by constructs that contradict a negative, such as “If you want a doctor who knows what’s best after not listening to you for 5 minutes, don’t see him. This quality of care is almost impossible to find” (Male; 5-star; 40558). As reviewers have assigned a general valence to their experience, we leveraged physician ratings to distinguish positive from negative sentiment. However, this applies only in analyses where positive reviews are considered in isolation from negative reviews and vice versa. In the quantitative analyses below, we typically controlled for interactions between gender and physician rating. Therefore, when considering a soft

XSL•FO
RenderX

skill in the context of reviews with ratings of the opposite valence, the soft skill should be interpreted as the negation of that interpersonal trait. For instance, when *amicability* appears in a negative review, the reader should interpret this as the *absence* of amicability, whereas *disrespect* in the same review should be regarded as the *presence* of disrespect.

Table 4. Logistic regression on the presence of a theme in review (n=106,325).

Model[a]	Intercept	$Doctor_F$[b]	$Reviewer_F$[b]	$Rating_{Neg}$	$Doctor_F \times Rating_{Neg}$	$Reviewer_F \times Rating_{Neg}$	$Reviewer_F \times Reviewer_F$
No interactions							
Candor	−1.92[c]	0.15[c]	0.16[c]	−0.10[c]	—[d]	—	—
Trust	−2.40[c]	0.19[c]	0.27[c]	−1.01[c]	—	—	—
Investment	−0.40[c]	0.27[c]	0.21[c]	−0.88[c]	—	—	—
Amicability	−1.57[c]	0.25[c]	0.23[c]	−1.20[c]	—	—	—
Indifference	−5.26[c]	0.29[c]	0.18[c]	2.19[c]	—	—	—
Disrespect	−5.30[c]	0.35[c]	0.24[c]	3.39[c]	—	—	—
$Doctor_F \times Rating_{Neg}$							
Candor	−1.88[c]	0.01	0.16[c]	−0.31[c]	0.47[c]	—	—
Trust	−2.39[c]	0.18[c]	0.27[c]	−1.03[c]	0.04	—	—
Investment	−0.40[c]	0.26[c]	0.21[c]	−0.89[c]	0.01	—	—
Amicability	−1.56[c]	0.20[c]	0.23[c]	−1.34[c]	0.29[c]	—	—
$Reviewer_F \times Rating_{Neg}$							
Candor	−1.92[c]	0.15[c]	0.16[c]	−0.10[c]	—	−0.00	—
Trust	−2.42[c]	0.19[c]	0.31[c]	−0.86[c]	—	−0.23[c]	—
Investment	−0.42[c]	0.27[c]	0.25[c]	−0.80[c]	—	−0.14[c]	—
Amicability	−1.58[c]	0.25[c]	0.25[c]	−1.12[c]	—	−0.13[e]	—
$Doctor_F \times Reviewer_F$							
Disrespect	−5.34[c]	0.46[c]	0.31[c]	3.39[c]	—	—	−0.16[e]

[a]Rows represent distinct logit models for each of the 7 themes. Each cell reports the log-likelihood that a variable is associated with the given theme. Sentences containing terms related to the process have been removed from the reviews.

[b]Female=1, male=0.

[c]$P<.001$.

[d]Missing value indicates that no coefficient was estimated for the given endogenous variable.

[e]$P<.01$.

Quantitative Analysis

A total of 3 sets of logistic regression models were fitted to these data. Reviews for which either reviewer or physician gender could not be identified were removed from the analysis in all of the models presented in the Results section. This reduced the data set by 28% from 154,305 reviews to 111,036.

The set of models we present investigated the association between negative reviews and reviewer/doctor gender. The second set comprised models examining the likelihood that a review mentions a soft skill. As our primary variables of interest are binary, and we are interested in interactions among those binary variables, the interpretation of the logistic regression variables is complicated. All interaction terms disrupt the interpretation of their component variables, but this interpretation is even more difficult when estimating all pairwise interactions of a set of variables (in this case, 3). Effectively, this decomposes each main effect coefficient into different components, which must be carefully interpreted and summed to construct odds ratios (ORs) for various conditions. We present one set of regressions on each review theme for each interaction rather than estimating all 3 interactions in a single model for readability. Through these sets, we tested *physician gender*, *reviewer gender*, and *physician rating* for pairwise interactions. The intercept and noninteracting effects are only marginally altered between these models, if at all. An additional set of models estimates the main effects of each dependent variable. When reporting these results, models are grouped by these model classes rather than by the dependent variable, as the relation between the independent variables takes priority over the particulars of the review themes. Finally, a pair of models

XSL•FO
RenderX

estimates how administrative process correlates with gender, physician ratings, and soft skills.

Results

This section discusses the statistical models fit to the review data. The report and discussion of these results is supplemented with excerpts of real reviews examined in the qualitative component of this study. We present these reviews to illustrate and contextualize the quantitative findings and the computational method of feature extraction.

Physician Ratings and Gender

Owing to the possibility that reviews could be influenced by clinical processes outside of physician control, we first fit a pair of models with and without mentions of bureaucratic *process*. The logit models summarized in Table 5 indicate that physician ratings are extensively influenced by gender, irrespective of mentions of *process*. Four models examine the correlation between sets of independent variables and the probability of a *negative review*. The *A* models were fit on all reviews for which we were able to infer both doctor and reviewer gender (*N*=111,036). In the data set on which we regressed the *B* models, we stringently controlled for mentions of *process*, which refer not to the physician but instead to the clinical aspects beyond the patient-physician relationship. The data set for the *B* models is the result of filtering sentences that mention terms associated with *process* from the reviews and then removing any reviews that were left without text. This process yielded a slightly smaller data set (*N*=106,325). Models 1A and 1B present a base model that includes only physician and reviewer gender and an interaction term. Models 2A and 2B control for soft skills, with Model 2A controlling for *process*. The overlapping coefficient estimates do not differ substantially between the 4 models, suggesting that mentions of *process* do not substantially alter the correlations captured by the variables of interest in these data. However, we conservatively controlled for them in the remainder of this section, except in models that consider *process* explicitly.

Table 5. Logistic regression on rating negative (A: n=111,036; B: n=106,325). Models 1 and 2 differ in the inclusion of review content themes. The B variants show the effects of filtering sentences mentioning process from each review.

Variable	Model 1A	Model 1B (no process)	Model 2A	Model 2B (no process)	*P* value
Intercept	−1.05	−1.08	−1.12	−0.80	<.001
$Doctor_F$[a]	0.67	0.69	0.63	0.71	<.001
$Reviewer_F$[a]	−0.16	−0.14	−0.14	−0.11	<.001
$Doctor_F \times Reviewer_F$	−0.03	−0.05	−0.02	−0.03	NS[b]
Candor	N/A[c]	N/A	0.06	0.05	<.05
Trust	N/A	N/A	−0.95	−0.92	<.001
Investment	N/A	N/A	−0.89	−0.90	<.001
Amicability	N/A	N/A	−1.49	−1.18	<.001
Indifference	N/A	N/A	2.35	2.28	<.001
Disrespect	N/A	N/A	3.42	3.45	<.001
Process	N/A	N/A	0.84	N/A	<.001

[a]Female=1, male=0.

[b]NS: not significant.

[c]Not applicable.

Although the dictionaries that contain terms related to review themes were developed in conjunction with our qualitative analysis and were thus thoroughly vetted, there remained concerns that these term lists do not adequately capture the themes they purportedly represent. As discussed in the Methods section, dictionary-based or *bag of words* (presence or absence of terms) approaches to natural language understanding often struggle to overcome or capture nuance in word use, notably suffering an inability to distinguish positive use from negation. We fit 2 models (2A and 2B) to verify that the soft skill dictionaries are correlated with negative reviews, as expected. The coefficients for the themes indicate that our dictionaries capture the basic tendency we anticipated: 3 of 4 positive soft skills, *trust*, *investment*, and *amicability*, are correlated with positive reviews (negative coefficients; *P*<.001), and 2 negative soft skills, *indifference* and *disrespect*, are correlated with negative reviews (*P*<.001). Notably, *candor* is not associated with either positive or negative reviews. This suggests the possibility that *candor* was mischaracterized by its dictionary. However, the intercoder agreement and classifier performance in Table 1 imply that the dictionary for *candor* captures the theme equally well as the other positive soft skills. Rather, it seems that *candor* was misclassified as a *positive* soft skill and, as defined by its dictionary, is perhaps better understood as a *neutral* soft skill, appearing equally in positive and negative reviews. The models estimate large absolute effects for the three other positive soft skills. A negative soft skill, *indifference*, fits a considerably greater effect than any positive soft skill, and *disrespect* fits an even greater effect. This is consistent with the descriptive statistics in Tables 2 and 3, which suggest that

XSL•FO
RenderX

although positive soft skills are more strongly associated with positive reviews, they also appear often in negative reviews. Negative soft skills by contrast occur overwhelmingly in negative reviews. Generally, these findings further suggest that our dictionaries accurately represent the themes they attempt to capture.

H1a: Female Physicians Are More Likely to Receive Negative Reviews Than Male Physicians

> *She was harsh and short. I always felt rushed and uncomfortable... it was like she was just making sure she did what was required. No sympathy at all.* [Female; 1-star review; O6198]

The logit models on physician ratings (Table 5) indicate that female doctors are considerably more likely to receive negative reviews. Model 1B, which includes only physician and reviewer gender and an interaction term, estimates that female physicians' reviews are 2.00 (95% CI 1.90-2.10) times as likely to be negative than the reviews of male physicians (log OR 0.69, 95% CI 0.65-0.74; $P<.001$). Model 2B, which controls for mentions of soft skills, estimates a slightly larger coefficient.

H1b: Female Reviewers Are More Likely to Report Negative Experiences With Doctors

Contrary to our hypothesis, Model 1B estimates that men write negative reviews at 1.15 (95% CI 1.08-1.16) times the rate of women ($P<.001$). Controlling for review content themes (Model 2B) fit a slightly smaller estimate. There was no interaction between physician gender and reviewer gender, indicating that female patients are no more likely to give a doctor of a particular gender a negative review than men are.

Patient Experience, Physician Ratings, and Gender

Several batteries of logistic regression models were fit to investigate how specific aspects of the patient experience (review themes) interact with gender and overall patient experience (physician rating). The coefficients estimated by these models are listed in Table 4. As described in the Methods section, we fit separate models for each of the interaction terms, as the interpretation of multiple interaction terms is complicated, and separating them into distinct models does not significantly alter the results. We fit models that interact for gender and physician ratings for positive soft skills only, as negative soft skills are almost exclusively found in negative reviews. Furthermore, we report the gender×gender interaction model for disrespect only, as no other model estimated a significant interaction.

As comments about *process* may be wrongly ascribed to a doctor's soft skills, the model for each theme controls for mentions of *process*. We found that *process* was significantly associated with all soft skills ($P<.001$), including a strong correlation with *amicability* and *disrespect* (see Table 6 and *Process and Gender* section for a more detailed treatment of this model). The model estimates that *process* is 2.73 (95% CI 2.65-2.86) times as likely to co-occur with *amicability* and 2.02 (95% CI 1.92-2.13) times as likely to co-occur with *disrespect*. This is unsurprising given our qualitative investigation, which found that the reviewers commonly commented on the friendliness or rudeness of the staff. For instance, when positive reviews mentioned both *process* and *disrespect*, it almost always contrasted a positive experience with a physician with a negative process experience. Reviewers seemed to be fairly capable of separating feelings about bureaucratic process from their experience with a physician, setting their dissatisfaction with, for example, staff, insurance, or booking aside when assigning a rating to a doctor who otherwise provided a good clinical experience.

Physician Ratings and Soft Skills

In our qualitative analysis, we observed that reviews mentioning positive soft skills were primarily associated with high scores for male and female doctors. Reviewers wrote positively about physicians who were candid and direct, "Ladies this doctor listens and responds with respect, she does not talk down to you either" (Female; 5-star; U940); who were trustworthy and supportive, "The delivery would've been terrifying without him" (Male; 4-star; O7390); invested, "She asks questions and listens. She makes me feel like I am important" (Female; 5-star; U933); and amicable, for example, "She is an amazing doctor. Kind, caring, empathetic, warm, knowledgeable, quick thinking, funny, and honest" (Female; 5-star; O679). As stated in the discussion of the logit models in Table 5, positive soft skills are more likely to appear in positive reviews ($P<.001$). These models also estimate large effects for positive soft skills except *candor*, being at least 2.4 times as likely to appear in positive reviews than negative ones.

We also coded for 2 negative soft skills, *indifference* and *disrespect*. *Indifference* was relatively rare, appearing in only 2% of the reviews. *Disrespect* was more common, occurring in 8% of all reviews. Unlike positive soft skills, which appeared in both positive and negative reviews, negative soft skills were almost exclusively found in negative reviews. When they did appear in positive reviews, it almost always referred to bureaucratic process, not the physician. Typical reviewer comments coded for negative soft skills relate experiences with doctors who lack courtesy, patience, and warmth toward their patients, for example, "I could not believe how condescending and snippy she was!" (Female; 1-star; O8376); "He was very rude, condescending, arrogant, and appeared angry" (Male; 1-star; U1122). Reviewers also described feeling ignored, "I felt passed around and ignored" (Female; 1-star; O10061), or that their concerns were dismissed, "...brushed it off" (Female; 1-star; O10100) and "I was in tears because he was too stubborn to listen" (Male; 1-star; U1047). Other complaints included ignoring patients' understanding of their own medical condition and lack of inclusion in decision-making. As reviewers mentioned, "The doctor does not listen to you and forces his opinion down your throat without considering your view. Do not visit here" (Male; 1-star; U398) and "He did not want to listen to anything I had to say and he definitely didn't want me getting a second opinion. He got defensive and standoffish at the mention of any other opinion which says SHADY all over it" (Male; 1-star; U198).

The logit models for negative soft skills indicate that negative soft skills are far more likely to appear in negative reviews than positive ones. The log-likelihood coefficients are considerably stronger for negative than for positive soft skills. This is because

XSL•FO
RenderX

positive soft skills can be negated to note the absence of a positive quality, whereas negative soft skills are rarely negated to indicate a positive quality.

H2a: Female Physicians Are More Likely to Receive Criticism Mentioning Soft Skills Than Male Physicians

> *I feel 100 percent comfortable telling her anything because I know she holds no judgment and treats everybody with fairness and kindness.* [Female; 5-star review; O8982]

The coefficients in the $Doctor_F \times Rating_{Neg}$ indicate that positive soft skills are more likely to occur in reviews of female physicians. In positive reviews, *trust*, *investment*, and *amicability* were more strongly associated with the reviews of female doctors than those of male doctors ($P<.001$). Trust and investment show no significant interaction between physician gender and physician rating, indicating that all reviews of female physicians are more likely to mention *trust* and *investment* than those of men. The models estimate that *trust* occurs 1.20 (95% CI 1.15-1.26) times and *investment* 1.31 (95% CI 1.27-1.34) times as often in reviews of female physicians than in those of male physicians. *Amicability* is estimated to have a significant effect on positive reviews and an additional amplifying effect in negative reviews. Positive reviews of female physicians reported *amicability* more often than those of male physicians (log OR 0.20, 95% CI 0.17-0.24; $P<.001$). To calculate the probability of mentions of *amicability* in female doctors' negative reviews, we summed the *base* (that of positive reviews) log OR (0.20) with the log OR (0.29, 95% CI 0.19-0.38; $P<.001$) from the interaction term. The model estimates that *amicability* is much more likely to be mentioned in negative reviews of female physicians than in reviews of male physicians (log OR 0.49, 95% CI 0.36-0.62). Similarly, *candor* is much more likely to appear in negative reviews of female physicians (log OR 0.48, 95% CI 0.35-0.60; $P<.001$), although it is equally likely to appear in male and female physicians' positive reviews.

Negative soft skills are more easily interpreted than positive ones, as they are less likely to be negated and thus occur predominately with negative valence. As discussed earlier, this is supported by the stronger associations of negative soft skills with negative reviews than those of positive soft skills and positive reviews. Qualitative analysis of reviews indicated that *disrespect*, when it occurs in positive reviews, usually refers to *process*. However, occasionally, reviewers will contradict or justify negative soft skills when referring to positive experiences with physicians, for example, "Some may misinterpret her candor as rudeness, but I appreciate that about her - she always gets right to the point" (Male; 5-star; 951) and "He can seem somewhat arrogant, but I've been seeing him for a while now, and he really knows his stuff and takes his patients very seriously. He has a right to think highly of himself!" (Male; 5-star; 7181).

Given their overwhelmingly negative valence, it is sufficient to model negative soft skills without interactions for gender and overall review quality (the *No Interactions* section of Table 4). However, it is important to control for physician ratings given the much higher rate of negative reviews in women's reviews. Both *indifference* (log OR 0.29, 95% CI 0.20-0.37; $P<.001$) and *disrespect* (log OR 0.35, 95% CI 0.30-0.41; $P<.001$) are more likely to be mentioned in reviews of female physicians than in reviews of male physicians.

H2a: Female Reviewers Are More Likely to Mention Soft Skills in Negative Reviews

Our qualitative analysis did not reveal consistent patterns of association between mentions of soft skills and reviewer gender. As this investigation was limited to a small sample of reviews, we expected the quantitative results to yield patterns consistent with observations of gender differences in patients' clinical experience reported in the literature. The logit models of soft skills offer 2 advantages in detecting gender bias in clinical settings. First, the high volume of observations may detect a pattern that was too rare to emerge from the qualitative analysis. Second, it may be that gender bias is not explicitly identified in most reviews but rather emerges when looking at the reviews in aggregate. The logistic models of soft skills demonstrate several associations between reviewer gender and soft skills, including interesting interaction effects.

The $Doctor_F \times Rating_{Neg}$ models in Table 4 find that all positive soft skills are more likely to occur in reviews written by women than those written by men ($P<.001$). *Candor* is roughly 1.17 (95% CI 1.13-1.21) times more likely to appear in all reviews written by women. The other 3 positive soft skills demonstrate a higher rate among women in positive reviews and a compensatory effect in women's negative reviews. However, this effect merely dampens the greater probability of occurring in reviews written by women, not equalizing it. In the models that did not fit coefficients for gender and rating interactions (*No Interactions*), both *indifference* (log OR 0.18, 95% CI 0.09-0.27; $P<.001$) and *disrespect* (log OR 0.24 95% CI 0.18-0.30; $P<.001$) were more likely to appear in reviews written by women.

H3: Female Reviewers Are More Likely to Report Negative Experiences With Male Doctors

> *I have hunted for a female Urologist for several years. I was dealing with a male doctor who kept blowing off my concerns as a woman and telling me what women think they feel or know.* [Female; 5-star review; U940]

When women mentioned soft skills, they occasionally related difficulties with their doctor to physician gender. However, female reviewers rarely attributed poor treatment to their womanhood or to male physicians treating women poorly. It was also rare that women commented on the absence of bias in settings where they might have expected it, for example, "While he treats women and men, I think his sensitivity makes him especially good with women" (Male; 5-star; U1081).

To examine whether women or men report differential treatment depending on the gender of their doctor, each model tested an interaction effect between physician gender and patient gender. Only *disrespect* produced a highly significant gender×gender interaction. The model for *disrespect* estimates log OR of −0.16 (95% CI −0.28-−0.04; $P=.008$) for women who review female doctors. The reader may be inclined to interpret the negative coefficient for gender×gender interaction as evidence that

XSL•FO
RenderX

women are less likely to report *disrespect* when seeing a female physician. This is true, but it must be qualified when we ask the question, *less likely relative to what*?

The odds ratio for the gender×g interaction indicates that the discrepancy in reports of disrespect between women and men is not so great when seeing a female doctor as we would have expected given the difference between women and men when seeing a male doctor. Female reviewers would seem to benefit from seeing female doctors, as we cannot reject the null that men and women report *disrespect* from a female doctor at equal probability. The $Doctor_F \times Reviewer_F$ model of *disrespect* estimates women to be 1.37 (95% CI 1.26-1.49) times as likely (log OR 0.31, 95% CI 0.23-0.40; $P<.001$) as men to report *disrespect* when seeing a male doctor (the $Reviewer_F$ column). Given the interaction term, this probability represents the OR that a woman (compared with a man) reports *disrespect* from a *male doctor*. Summing this *base* probability with the interaction coefficient (log OR −0.16) estimates that female reviewers are 1.16 (95% CI −0.05-0.36) times (log OR 0.15) more likely to report *disrespect* when seen by a female doctor than a *man* seeing a female doctor. As the 95% confidence interval overlaps 0; we cannot reject the null hypothesis, that men and women report disrespect with the same probability when seeing women doctors. Similarly, when a man reviews a female physician, he is 1.58 (95% CI 1.43-1.74; log OR 0.46, 95% CI 0.36-0.55) times more likely to associate her with *disrespect* than he would a male doctor. The compensatory effect of the gender×gender interaction coefficient diminishes, but does not dissolve, the probability that a female doctor is reported to be disrespectful. When reviewed by female patients, female doctors are 1.35 (95% CI 1.08-1.67) times more likely to be associated with *disrespect* than a male doctor.

Process and Gender

Although we made no predictions about administrative *process*, it is worth noting several patterns that emerged from the logistic models on mentions of *process*. We fit 2 models of *process* reported in Table 6 in the reviews for which gender could be inferred and which did not filter out mentions of *process* (N=111,036). Model 1 parallels the models of themes in Table 4 and accordingly reports interaction effects as separate models. Model 2 estimates correlations between each theme and *process*, controlling for rating and gender.

Table 6. Logistic regression on the presence of process in review (N=111,036).

Variable	Model 1			Model 2
	$Doctor_F$[a]×$Rating_{Neg}$	$Reviewer_F$[a]×$Rating_{Neg}$	$Doctor_F$×$Reviewer_F$	Themes
Intercept	−0.37[b]	−0.41[b]	−0.43[b]	−0.84[b]
$Doctor_F$	0.06[b]	0.20[b]	0.19[b]	0.12[b]
$Reviewer_F$	0.11[b]	0.11[b]	0.14[b]	0.04[c]
$Rating_{Neg}$	0.54[b]	0.70[b]	0.78[b]	0.83[b]
$Doctor_F \times Rating_{Neg}$	0.42[b]	—[d]	—	—
$Reviewer_F \times Rating_{Neg}$	—	−0.02	—	—
$Doctor_F \times Reviewer_F$	—	—	−0.13[b]	—
Candor	—	—	—	0.26[b]
Trust	—	—	—	0.34[b]
Investment	—	—	—	0.22[b]
Amicability	—	—	—	1.01[b]
Indifference	—	—	—	0.00
Disrespect	—	—	—	0.70[b]

[a]Female=1, male=0.

[b]$P<.001$.

[c]$P<.01$.

[d]Missing value indicate that no coefficient was estimated for the given endogenous variable.

Process is much more likely to be mentioned in negative reviews ($P<.001$). This is consistent across all models. This may be because when process is smooth, it is more likely to go unnoticed, whereas poor experiences with process are more likely to color the overall experience. Parallel to the trend observed in soft skills, *process* more often occurs in reviews written by and about women ($P<.001$). The $Doctor_F \times Rating_{Neg}$ model estimates that negative reviews of female physicians are 2.61 (95% CI 2.41-2.88) times as likely to mention process than negative reviews of male physicians (log OR 0.96, 95% CI 0.88-1.06; $P<.001$). By contrast, female and male reviewers mention process in negative reviews at equal rates. Women who see female doctors are less likely to mention *process*, which

XSL•FO
RenderX

produces an equalizing effect that offsets the greater rate of reports of *process* for both female reviewers and physicians.

We also examined the association of *process* with soft skills. Model 2 demonstrates a positive correlation between mentions of *process* and soft skills. The correlations with *amicability* and *disrespect* are sizable (log OR 1.01, 95% CI 0.98-1.04 and 0.70, 95% CI 0.65-0.76, respectively; $P<.001$), indicating that patients value the ease of interpersonal interactions with staff, that is, whether they are *friendly* or *rude*, and likely interpret bureaucratic competence through the framework of how *nice* the staff are. In our qualitative analysis, we found that positive reviews expressing *disrespect* overwhelmingly do so with regard to *process*, indicating that reviewers are able to separate relationships with their doctor and the overall clinical experience as reflected in the final review. This compartmentalization is well illustrated by a patient who reported a positive experience with a doctor but faced problems with poor administration such that the patient ultimately severed their relationship with the clinic:

> *Terrific bedside manner!! Really dedicates time to patients and will even follow up by phone. The staff are rude and incompetent though. They repeatedly failed to file paperwork with my insurance. I got fed up with it and had to find a new doctor.* [Unknown; 5-star review; 3686]

Discussion

Interpreting These Results

Our results provide compelling evidence for a number of effects of gender on patient experience, as reported in physician reviews. These findings may be interpreted through 2 distinct frames. First, a *patient experience* frame attempts to interpret gender dynamics in the context of the patient-physician relationship. This frame should be familiar to readers versed in the literature on gender and clinical experience. The second frame, the *online review system* frame, seeks not to understand or improve the clinical aspects of health care but rather considers how gender differences may subvert or be leveraged to improve reviews as a valuable public resource that informs decisions about care-seeking.

The following discussion of the results in context of the hypotheses of this study assumes the *patient experience* perspective. The Summary and Recommendations section, however, largely reflects the *online review system* perspective, which is less concerned with controlled statistical inference than it is how the descriptive statistical patterns in Tables 2 and 3 might affect public perception of physician quality and how we might design online review systems to improve physician-patient matching and offset bias.

H1a: Physician Ratings and Physician Gender

We hypothesized that female physicians would be more likely to receive lower ratings. This hypothesis was supported by our study. These data indicate that there is considerable reviewer bias against female physicians. This is consistent with well-documented patterns of bias against women in other fields, notably when reviewing instructor performance in a controlled online classroom [26].

H1b: Reviewer Gender and Rating

We hypothesized that female patients would report overall worse clinical experiences than men. The results here support the opposite hypothesis. We found that men are slightly more likely to report negative experiences than women.

There are several valid interpretations of these findings. Men may receive worse care than women, as captured by patient experiences (rather than health outcomes). Alternatively, it could be that men have higher expectations for care than women or are less competent at navigating the clinical setting. Both are plausible given that men less frequently utilize health services [40,41]. Finally, we might attribute the discrepancy to women's greater propensity for agreeableness [42,43], and for forgiveness [44] and compromise [45] in interpersonal conflict. These interpretations are not mutually exclusive, and further research is warranted to account for this trend.

H2a: Soft Skills and Physician Gender

We hypothesized that reviews of female physicians would be more likely to critique their soft skills. Our results indicate that this is true of all soft skills. All soft skills were more often mentioned in the reviews of female physicians. In negative reviews of female physicians, reviewers were considerably more likely to mention *candor* and *amicability*.

The results supporting this hypothesis indicate that female physicians' soft skills are more likely to be critiqued and that female physicians are much more likely to be associated with *disrespect*. We also present evidence that women may be penalized for lacking *candor* and *amicability* to a much greater degree than men. This may be attributed to failure to live up to a positive stereotype, as women are generally expected to be more open and personable, and female physicians in particular are expected to be more caring [46]. Furthermore, the physician role is one of authority, and research has extensively documented that women are punished for leadership styles that men are rewarded for [47,48].

H2b: Soft Skills and Reviewer Gender

We hypothesized that female reviewers are more likely to reflect on their doctor's soft skills. This hypothesis was supported by our study.

In positive reviews, women mentioned all positive soft skills with greater probability. However, the magnitude of these effects was diminished in negative reviews. These patterns suggest that women may be more inclined to value a physician's soft skills. However, it also indicates that men may be more sensitive to a lack of positive soft skills than their presence.

We found that women were more likely to mention the negative soft skills, *indifference* and *disrespect*. This likely reflects the wealth of literature documenting the tendency for physicians to take women's concerns less seriously and treat them with condescension [27,29,31].

XSL•FO
RenderX

H3: Reviewer Gender and Physician Gender

We expected that female patients would report more negative experiences with male doctors. Our results support this hypothesis. Although there was no significant interaction effect of reviewer and physician gender on the probability of a negative experience, we found a significant interaction between reviewer and physician gender on the likelihood of reporting *disrespect*, which overwhelmingly occurs in negative reviews. Female reviewers apparently benefit from seeing female doctors, as they are less likely to mention *disrespect* when reviewing female physicians than when they review men. This compensating effect neutralized the overall greater association of *disrespect* with female reviewers, but not female doctors.

Given the literature and our previous finding that female physicians are subject to biased reviews, these results suggest that even women harbor bias against female physicians. However, this bias is considerably smaller among women than it is among men and is complicated by physician-gender/patient-gender preferences for different communication styles [32].

Process and Physician Gender

We made no predictions regarding the relationship between *process* and soft skills or reviewer/physician gender. However, we found that both *amicability* and *disrespect* were highly correlated with *process*, suggesting that the ease of social interaction with staff is important to reviewers. Importantly, negative reviews of female physicians are considerably more likely to mention aspects of the clinical experience beyond experiences with the doctor.

The strong association between *process* and negative reviews of female doctors may reflect a tendency for patients to assess male doctors "on their own merits," whereas women are more likely to be held accountable for poor process. This reflects a *bias against women* interpretation. A *realist* account might hypothesize that female doctors are more likely to work in clinics with less competent or accommodating staff.

Limitations

This work is complementary to previous qualitative studies on the influence of gender on the doctor-patient relationship. We acknowledge that it is unclear whether gender differences reflect patient perception or the reality of physician behavior. For example, given the ample evidence in other contexts on bias against women, the high rate of negative reviews for female physicians likely reflects reviewer bias against physicians rather than genuine differences in treatment. However, our approach does not allow certainty in this regard. The Discussion section provides a more detailed explanation of the interpretation of gender differences in the context of the findings of previous studies. We also recognize that the data do not contain information related to patients' health outcomes. Although the health outcome of each patient is not represented in our data, other studies have shown that reviews can reflect real health consequences [14]. Similarly, treatment noncompliance, unwarranted recalcitrance, and other patient characteristics beyond the reviewer's narrative are not captured in the reviews.

In this paper and similar research, gender representation is reported as binary, which does not capture the full spectrum of gender or gendered interaction. Even though gender is likely to be interpreted in a binary fashion by most review readers [21], the doctor-patient relationship is more complex, and other data could offer more nuanced and richer perspectives. Furthermore, our work does not consider the intersection of gender and other identities, such as race [49,50].

As noted in the Methods section, the Google Places API limits data collection to 5 reviews per physician or practice. Google provides no documentation on how these reviews are chosen from all the reviews written. We acknowledge that the small sample may not be fully representative of a doctor or practice; however, our contribution is more focused on the biases within the reviews and not on the doctors themselves. We also note that the Google Places reviews are subject to selection biases. Demographics undoubtedly play a role in determining who writes physician reviews (eg, consider the high proportion of female reviewers in our data set). However, the data likely suffer selection biases, irrespective of demographic differences. For instance, it is probable that the U-shaped distribution of physician ratings is both a product of overall polarized attitudes and strong experiences providing greater motivation to write a review. It seems likely that other such selection biases were present in these data but were unknown to us as we performed our analysis.

Summary and Recommendations

The increasing prevalence of online reviews of physicians affects both medical practices and patient choices. However, little is known about biases that may be present in these reviews or whether they reflect the real biases documented in doctor-patient interactions [27,28]. Conversely, most studies of gender bias in doctor-patient interactions to date have been limited to qualitative analyses, smaller-scale data sets, or specific medical conditions.

This study is the first to provide evidence that gender biases and other gender differences are observable at scale in physician reviews. We provide extensive evidence of differences in physician ratings and review content with respect to both physician and reviewer gender. Our statistical inference indicates that these differences are robust when controlling for possible confounding relationships and therefore are likely to reflect gender differences and biases in the patient-physician relationship.

It is difficult to disentangle which aspects of these gender differences may be attributed to review selection bias, gender bias, or gender behavioral variation. However, these patterns undeniably affect prospective patients as they peruse online reviews to select a doctor. Thus, it is important to consider how we might educate the public about the effects of gender bias on physician ratings and how online review systems could be improved to control for bias. We propose several concrete steps that could be taken to better support patients.

Review systems can draw attention to gender differences in reviews to aid prospective patients in building their own understanding of a physician's potential gender biases. One

XSL•FO
RenderX

way to do this is to organize reviews by gender. This could either be the default presentation or a special gender-separated view. Alternatively, reviewer gender might be indicated explicitly only in automatically generated summaries of physician reviews. Additionally, prospective patients might benefit from a panel that provides a sense of how a particular physician compares with other physicians on gender. For example, if there is a discrepancy between men's and women's ratings for a physician that differs greatly from the gender discrepancy of other similar physicians, a prospective patient might benefit from reading their reviews with this information in mind.

An online review system could also help to correct for gender differences that generally affect reviews. As female physicians receive many more negative reviews on average, a prospective patient might find it easier to select among physicians if ratings are adjusted to control for the physician's gender or if ratings are reported relative to physicians of the same gender. On the other hand, online review systems could implement measures to reduce bias in the reviews. For instance, when writing a review and using a word that is commonly used to critique female physicians, the system could prompt the reviewer about gender stereotyping in word choice or alternative terms that are gender neutral. This might encourage a more balanced approach to review writing and help reviewers recognize their own biases. Alternatively, information can be solicited from the reviewer in such a way that greatly reduces the effects of gender stereotypes on performance evaluation [51].

These approaches would ideally lead to reviews that more accurately reflect the quality of care provided by physicians. Finally, this study draws attention to several important areas for future work. We advocate researchers adopt mixed methods approaches similar to the one presented here when pursuing quantitative analyses of text. Furthermore, this study raises questions specifically related to online review systems as objects of study in their own right. Little is known about how readers interpret online reviews, notably in the context of health care and gender. It also highlights the need to study how review systems can be designed to improve review accuracy and inform review readers and writers on gender bias in online reviews. We propose that experimental studies in review cognition and system design will be most fruitful to these ends.

Authors' Contributions

JM and KS designed the study. LZ and UB collected the data. LZ performed the initial analysis. ZD performed the statistical analysis. All authors assisted with writing the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook.
[PDF File (Adobe PDF File), 41 KB - jmir_v22i7e14455_app1.pdf]

References

1. Rosenbaum L. Scoring no goal-further adventures in transparency. N Engl J Med 2015 Oct 8;373(15):1385-1388. [doi: 10.1056/NEJMp1510094] [Medline: 26332360]
2. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
3. Gao G, McCullough JS, Agarwal R, Jha AK. Are Doctors Created Equal? An Investigation of Online Ratings by Patients. In: Proceedings of the Work- shop on Information Systems and Economics. 2010 Presented at: WISE'10; June 10-11, 2010; St Louis, Missouri URL: https://www.semanticscholar.org/paper/Are-doctors-created-equal-An-investigation-of-by-Gordon-McCullough/59225be0be8ea64b249f701b6f6bc4749a6e0f93
4. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
5. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
6. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What do patients say about doctors online? A systematic review of studies on patient online reviews. J Med Internet Res 2019 Apr 8;21(4):e12521 [FREE Full text] [doi: 10.2196/12521] [Medline: 30958276]
7. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
8. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]

XSL•FO
RenderX

9. Lee S. 'I hate my doctor': reputation, defamation, and physician-review websites. Health Matrix Clevel 2013;23(2):573-606. [Medline: 24341082]
10. Healthgrades. URL: http://www.healthgrades.com/ [accessed 2020-05-14]
11. RateMDs. URL: http://www.ratemds.com/ [accessed 2020-05-14]
12. Vitals. URL: http://www.vitals.com/ [accessed 2020-05-14]
13. Yelp. URL: http://www.yelp.com/ [accessed 2020-05-14]
14. Paul MJ, Wallace BC, Dredze M. What Affects Patient (Dis)satisfaction? Analyzing Online Doctor Ratings with a Joint Topic-Sentiment Model. In: AAAI Press Technical Reports. 2014 Presented at: AAAI'13; July 14-18, 2013; Bellevue, Washington URL: https://www.semanticscholar.org/paper/What-Affects-Patient-(Dis)satisfaction-Analyzing-a-Paul-Wallace/a23628e9d88cacafc35454ba77047f3de2e69f86
15. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
16. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a sports medicine surgeon's 'quality' in the modern era: an analysis of popular physician online rating websites. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
17. Johnson RL, Roter D, Powe NR, Cooper LA. Patient race/ethnicity and quality of patient-physician communication during medical visits. Am J Public Health 2004 Dec;94(12):2084-2090. [doi: 10.2105/ajph.94.12.2084] [Medline: 15569958]
18. Mike M. Doctor Sues Patients Over Bad Yelp Reviews. Techdirt. 2010. URL: https://www.techdirt.com/articles/20101110/19053611809/doctor-sues-patients-over-bad-yelp-reviews.shtml [accessed 2020-06-01]
19. Upmark M, Borg K, Alexanderson K. Gender differences in experiencing negative encounters with healthcare: a study of long-term sickness absentees. Scand J Public Health 2007;35(6):577-584. [doi: 10.1080/14034940701362194] [Medline: 17852974]
20. Cohoon JM, Aspray W. Women and Information Technology: Research on Underrepresentation. Cambridge, Massachusetts: The MIT Press; 2008.
21. West C, Zimmerman DH. Accounting for doing gender. Gend Soc 2009 Feb;23(1):112-122. [doi: 10.1177/0891243208326529]
22. Richards C, Bouman WP, Seal L, Barker MJ, Nieder TO, T'Sjoen G. Non-binary or genderqueer genders. Int Rev Psychiatry 2016;28(1):95-102. [doi: 10.3109/09540261.2015.1106446] [Medline: 26753630]
23. Rojo D, Esteban CG. The gender of power: the female style in labour organizations. In: Lazar M, editor. Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse. Houndmills, United Kingdom: Palgrave Macmillan; 2005:61-89.
24. Schmader T, Whitehead J, Wysocki VH. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. Sex Roles 2007;57(7-8):509-514 [FREE Full text] [doi: 10.1007/s11199-007-9291-4] [Medline: 18953419]
25. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. J Appl Psychol 2009 Nov;94(6):1591-1599. [doi: 10.1037/a0016539] [Medline: 19916666]
26. MacNell L, Driscoll A, Hunt AN. What's in a name: exposing gender bias in student ratings of teaching. Innov High Educ 2014 Dec 5;40(4):291-303. [doi: 10.1007/s10755-014-9313-4]
27. Elderkin-Thompson V, Waitzkin H. Differences in clinical communication by gender. J Gen Intern Med 1999 Feb;14(2):112-121 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00296.x] [Medline: 10051782]
28. Schmittdiel J, Grumbach K, Selby JV, Quesenberry CP. Effect of physician and patient gender concordance on patient satisfaction and preventive care practices. J Gen Intern Med 2000 Nov;15(11):761-769 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.91156.x] [Medline: 11119167]
29. Street Jr RL. Gender differences in health care provider-patient communication: are they due to style, stereotypes, or accommodation? Patient Educ Couns 2002 Dec;48(3):201-206. [doi: 10.1016/s0738-3991(02)00171-4] [Medline: 12477604]
30. Halbert CH, Armstrong K, Gandy OH, Shaker L. Racial differences in trust in health care providers. Arch Intern Med 2006 Apr 24;166(8):896-901. [doi: 10.1001/archinte.166.8.896] [Medline: 16636216]
31. National Survey of Women's Health. The Commonwealth Fund. 1993. URL: https://www.commonwealthfund.org/publications/publication/1993/jul/national-survey-womens-health [accessed 2020-05-14]
32. Mast MS, Hall JA, Roter DL. Disentangling physician sex and physician communication style: their effects on patient satisfaction in a virtual medical visit. Patient Educ Couns 2007 Sep;68(1):16-22. [doi: 10.1016/j.pec.2007.03.020] [Medline: 17482418]
33. Sandhu H, Adams A, Singleton L, Clark-Carter D, Kidd J. The impact of gender dyads on doctor-patient communication: a systematic review. Patient Educ Couns 2009 Sep;76(3):348-355. [doi: 10.1016/j.pec.2009.07.010] [Medline: 19647969]
34. Mast MS. Dominance and gender in the physician-patient interaction. J Mens Health Gend 2004 Dec;1(4):354-358. [doi: 10.1016/j.jmhg.2004.10.013]
35. Nelson LK. Computational grounded theory: a methodological framework. Sociol Methods Res 2017 Nov 21;49(1):3-42. [doi: 10.1177/0049124117729703]

XSL•FO
RenderX

36. Overview: Introducing the API. Google Developers. URL: https://developers.google.com/places/web-service/intro [accessed 2020-05-14]
37. Prostate Cancer Rates by State. Centers for Disease Control and Prevention. 2014. URL: https://www.cdc.gov/cancer/prostate/statistics/state.htm [accessed 2020-05-14]
38. Weinstein M. TAMS Analyzer for Macintosh OS X. TAMS Analyzer - SourceForge. URL: http://tamsys.sourceforge.net/ [accessed 2020-05-14]
39. Elmas F. SexMachine 0.1.1. The Python Package Index. URL: https://pypi.org/project/SexMachine/ [accessed 2020-05-14]
40. Ladwig K, Marten-Mittag B, Formanek B, Dammann G. Gender differences of symptom reporting and medical health care utilization in the German population. Eur J Epidemiol 2000 Jun;16(6):511-518. [doi: 10.1023/a:1007629920752] [Medline: 11049093]
41. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
42. Feingold A. Gender differences in personality: a meta-analysis. Psychol Bull 1994 Nov;116(3):429-456. [doi: 10.1037/0033-2909.116.3.429] [Medline: 7809307]
43. Costa Jr PT, Terracciano A, McCrae RR. Gender differences in personality traits across cultures: robust and surprising findings. J Pers Soc Psychol 2001 Aug;81(2):322-331. [doi: 10.1037/0022-3514.81.2.322] [Medline: 11519935]
44. Miller AJ, Worthington EL, McDaniel MA. Gender and forgiveness: a meta–analytic review and research agenda. J Soc Clin Psychol 2008 Oct;27(8):843-876. [doi: 10.1521/jscp.2008.27.8.843]
45. Gayle BM, Preiss RW, Allen M. A meta-analytic interpretation of intimate and nonintimate interpersonal conflict. In: Interpersonal Communication Research: Advances Through Meta-analysis. Abingdon, United Kingdom: Routledge; 2002:345-368.
46. Fennema K, Meyer DL, Owen N. Sex of physician: patients' preferences and stereotypes. J Fam Pract 1990 Apr;30(4):441-446. [Medline: 2324696]
47. Eagly AH, Mladinic A, Otto S. Are women evaluated more favorably than men?: an analysis of attitudes, beliefs, and emotions. Psychol Women Q 2016 Nov 25;15(2):203-216. [doi: 10.1111/j.1471-6402.1991.tb00792.x]
48. Eagly AH, Karau SJ. Role congruity theory of prejudice toward female leaders. Psychol Rev 2002 Jul;109(3):573-598. [doi: 10.1037/0033-295x.109.3.573] [Medline: 12088246]
49. Crenshaw K. Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chic Leg Forum 1989;1989(1):139-167. [doi: 10.4324/9780429500480-5]
50. Young IM, Allen DS. Justice and the Politics of Difference. Princeton, New Jersey: Princeton University Press; 2011.
51. Bauer CC, Baltes BB. Reducing the effects of gender stereotypes on performance evaluations. Sex Roles 2002;47(9-10):465-476. [doi: 10.1023/A:1021652527696]

Abbreviations

NLP: natural language processing
OR: odds ratio

Edited by G Eysenbach; submitted 21.04.19; peer-reviewed by M Reblin, M Reynolds, Y Motoki; comments to author 01.10.19; revised version received 15.03.20; accepted 28.04.20; published 30.07.20.

Please cite as:
Dunivin Z, Zadunayski L, Baskota U, Siek K, Mankoff J
Gender, Soft Skills, and Patient Experience in Online Physician Reviews: A Large-Scale Text Analysis
J Med Internet Res 2020;22(7):e14455
URL: https://www.jmir.org/2020/7/e14455
doi:10.2196/14455
PMID:32729844

©Zackary Dunivin, Lindsay Zadunayski, Ujjwal Baskota, Katie Siek, Jennifer Mankoff. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Changing the Health Behavior of Patients With Cardiovascular Disease Through an Electronic Health Intervention in Three Different Countries: Cost-Effectiveness Study in the Do Cardiac Health: Advanced New Generation Ecosystem (Do CHANGE) 2 Randomized Controlled Trial

Jordi Piera-Jiménez[1,2], MSc; Marjolein Winters[3], MSc; Eva Broers[4,5], PhD; Damià Valero-Bover[2], MSc; Mirela Habibovic[4,5], PhD; Jos W M G Widdershoven[4,5], MD, PhD; Frans Folkvord[1,6], PhD; Francisco Lupiáñez-Villanueva[1,7], PhD

[1]Open Evidence Research Group, Universitat Oberta de Catalunya, Barcelona, Spain

[2]Department of R&D, Badalona Serveis Assistencials, Badalona, Spain

[3]Smart Homes, Eindhoven, Netherlands

[4]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

[5]Department of Cardiology, Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

[6]Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

[7]Department of Information and Communication Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Jordi Piera-Jiménez, MSc
Open Evidence Research Group
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona, 08018
Spain
Phone: 34 651041515
Email: jpieraj@uoc.edu

Abstract

Background: During the last few decades, preventing the development of cardiovascular disease has become a mainstay for reducing cardiovascular morbidity and mortality. It has been suggested that interventions should focus more on committed approaches of self-care, such as electronic health techniques.

Objective: This study aimed to provide evidence to understand the financial consequences of implementing the "Do Cardiac Health: Advanced New Generation Ecosystem" (Do CHANGE 2) intervention, which was evaluated in a multisite randomized controlled trial to change the health behavior of patients with cardiovascular disease.

Methods: The cost-effectiveness analysis of the Do CHANGE 2 intervention was performed with the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing tool, based on a Markov model of five health states. The following two types of costs were considered for both study groups: (1) health care costs (ie, costs associated with the time spent by health care professionals on service provision, including consultations, and associated unplanned hospitalizations, etc) and (2) societal costs (ie, costs attributed to the time spent by patients and informal caregivers on care activities).

Results: The Do CHANGE 2 intervention was less costly in Spain (incremental cost was −€2514.90) and more costly in the Netherlands and Taiwan (incremental costs were €1373.59 and €1062.54, respectively). Compared with treatment as usual, the effectiveness of the Do CHANGE 2 program in terms of an increase in quality-adjusted life-year gains was slightly higher in the Netherlands and lower in Spain and Taiwan.

Conclusions: In general, we found that the incremental cost-effectiveness ratio strongly varied depending on the country where the intervention was applied. The Do CHANGE 2 intervention showed a positive cost-effectiveness ratio only when implemented in Spain, indicating that it saved financial costs in relation to the effect of the intervention.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT03178305; https://clinicaltrials.gov/ct2/show/NCT03178305

(J Med Internet Res 2020;22(7):e17351) doi:10.2196/17351

KEYWORDS

cost-effectiveness; randomized controlled trial; RCT; eHealth; cardiovascular disease; engagement; behavior change; digital health

Introduction

Background

In the last few decades, prevention at both population and individual levels in patients with established cardiovascular disease (CVD) has become a mainstay for reducing cardiovascular morbidity and mortality [1]. However, CVD remains the leading cause of death globally [2].

One of the cornerstones of CVD prevention is the promotion of lifestyle changes, including physical activity, a healthy diet, and avoidance of unhealthy behaviors such as smoking and drinking alcohol [1]. However, providing patients with relevant information regarding the importance of lifestyle habits seems to be insufficient to prompt these changes and maintain them over time [3]. Instead, it has been suggested that the preventive paradigm should shift from passive to more committed approaches of self-care based on the following three core elements: self-care maintenance, self-care monitoring, and self-care management [4-6].

The emergence of solutions based on information and communication technologies (ICTs), such as telemedicine, has greatly contributed to filling some of the gaps of effective self-care. One of the most classical ICT solutions has been the use of self-monitoring devices in patients with high cardiovascular risk to facilitate successful blood pressure (BP) control [7]. The expansion of mobile apps and their peripheral devices has raised the number of ICT-based interventions aimed at improving not only self-monitoring but also behavior changes in various patient profiles, including older people with high cardiovascular risk [8-10]. The emergence of lifestyle data-driven apps illustrates the increasing interest in this approach in various health care areas [11].

To date, evidence regarding the efficacy of these interventions is still evolving. Clinical guidelines for the prevention of CVD highlight that cost-effectiveness data from randomized controlled trials (RCTs) are scarce, and most data regarding the cost-effectiveness of cardiovascular prevention strategies combine clinical evidence with simulation approaches [1,12,13]. Simulation modeling is currently used to address important issues in clinical practice and health policy that have been very difficult to study within high-quality clinical trials but provide necessary insights for making health care decisions. Nonetheless, assumptions and personal choices are required to conduct simulation modeling, leading to potentially biased outcomes. Transparency in decision-making is therefore critical to adequately understand the observed outcome [14]. In this regard, there is a need for providing the various stakeholders, particularly policy makers, with evidence from nonsimulated research trials to understand the financial consequences of scaling up ICT solutions for health care systems [15]. In this study, we aimed to determine whether the Do Cardiac Health: Advanced New Generation Ecosystem (Do CHANGE) 2 preventive intervention is a cost-effective alternative for patients with CVD in Spain, the Netherlands, and Taiwan.

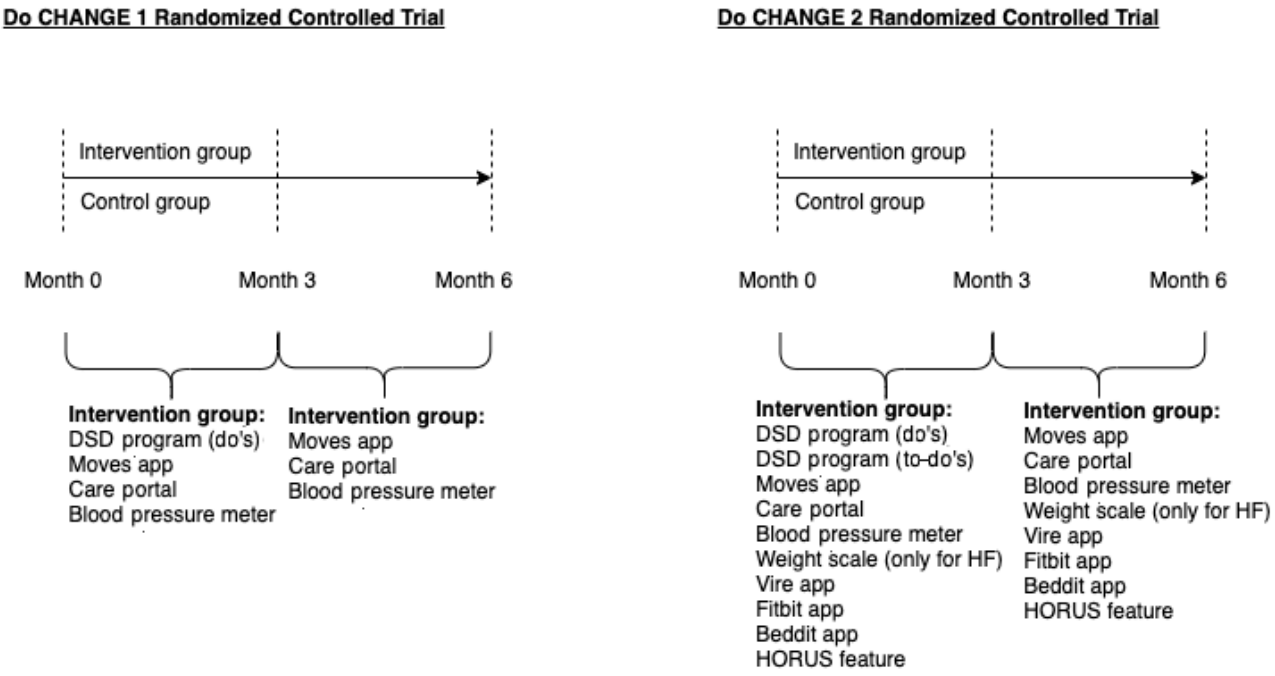
The Do CHANGE Project

The Do CHANGE program was developed as an ICT-based alternative for providing health education, which leads to behavioral changes in care recipients [16,17]. The Do CHANGE program consists of a 6-month intervention with a set of devices that include self-monitoring tools and the Do Something Different (DSD) behavior change program (only available during the first 3 months of the intervention), which has been shown to be effective in changing health behaviors in previous studies targeting different populations [18]. The Do CHANGE program included the following two phases: Do CHANGE 1 and Do CHANGE 2, which were assessed in two consecutive RCTs (Figure 1).

Patients included in the Do CHANGE 1 study received the DSD behavior change program, which was provided via text messages on patients' mobile phones. Behavioral flexibility is associated with a broad range of behavioral repertoires, making people more open to experience and adopt new behaviors [19]. This is achieved by disrupting patients' daily behavioral routines for a short period (eg, a few seconds) with behavioral prompts (referred to as "do's") delivered through patients' mobile phones. These messages challenge patients to do something different and get out of their comfort zone. Do's have been developed by a multidisciplinary team including cardiologists and psychologists, ensuring that they apply to the target population and are thus related to daily behaviors or needs. Patients received a total of 32 do's during the 3-month intervention period (2-3 do's per week). The program was tailored to the cardiac population with slight differences in the program depending on patients' primary diagnosis (coronary artery disease, heart failure, and hypertension), as the preferred health behaviors may vary depending on the diagnosis [20]. In order to obtain objective measures on patients' physical functioning, all participants received a BP monitor, the Moves app (Facebook Inc; to register GPS location), and the Careportal (Docobo Ltd, home monitoring device measuring daily symptoms and an electrocardiogram).

XSL•FO
RenderX

Figure 1. Do CHANGE 1 and 2 randomized controlled trial design including intervention details. Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem; DSD: Do Something Different; HF: heart failure.



The main additional features of the Do CHANGE 2 compared with the Do CHANGE 1 trial were the greater number of devices for self-monitoring and collecting behavioral information, and the capacity of the DSD program to tailor the behavioral prompts to the actual behavior of the care recipient, thus allowing for a personalized approach. Do CHANGE 2 integrates the principle of theory-driven behavioral change techniques, which can guide behavior change, within the offered interventions [21]. As a natural evolution from the Do CHANGE 1 trial, the second phase aimed to increase the ability of a person to express behavior in a more context-dependent way [22], thus being more open to experience and increasing the likelihood of adopting new behaviors [19].

Care recipients perceived the Do CHANGE 1 program as helpful and easy to use; however, it failed to prompt relevant lifestyle changes (measured with the Health Promotion Lifestyle Profile-II questionnaire) compared with treatment as usual (TAU) [20]. The Do CHANGE 2, based on a more personalized approach, resulted in a relevant change in lifestyle behavior over time in the intervention group. In addition, the intervention was perceived as useful and feasible by patients and health care professionals [23]. In order to provide a broader perspective of the effects of this program, we present herein the results of the cost-effectiveness analysis of the Do CHANGE 2 compared with TAU.

Methods

Trial Design and Patients

This was a multisite RCT to assess the cost-effectiveness of an ICT-based program to change behavior in patients with CVD compared with TAU. Local clinical specialists and research assistants recruited adult patients treated in the following three hospitals in three different countries: *Badalona Serveis Assistencials* (Spain), *Elisabeth TweeSteden Ziekenhuis* (the Netherlands), and *Buddhist Tzu-Chi Dalin General Hospital* (Taiwan). The planned sample size based on the available project resources was 75 patients for Spain, 75 for the Netherlands, and 100 for Taiwan. Once accepted to participate in the study, patients at each study site were randomized to receive either the TAU or the Do CHANGE 2 intervention. The primary outcomes were lifestyle change and quality of life. Additionally, behavioral flexibility was considered a mediator variable in this relationship. As the project aimed to provide proof of concept and examine the feasibility of the intervention, no sample size calculation was performed a priori. Recruiting a comparable number of patients across the countries was considered relevant to provide proof of concept. The details regarding the study patients and trial design are described in the report by Habivovic et al [16].

The most remarkable changes from the original study protocol (Do CHANGE 1) were the changes in the DSD program (moving from predefined messages according to the patient psychological profile to nudges tailored according to their behavior as gathered by the measurement devices), the addition of two new wearable devices (Beddit [Apple] and Fitbit [Fitbit Inc]), and the Vire app (Do CHANGE app). Considering the importance of weight in heart failure (HF), patients with this diagnosis also received a weight scale.

Inclusion Criteria

Participants were screened from adult patients (aged 18-75 years) who had been primarily diagnosed with either hypertension (ie, systolic BP [SBP]/diastolic BP [DBP] ≥140/90 mmHg in two consecutive measurements), coronary artery disease (ie, occurrence of myocardial infarction or angina pectoris, or previous percutaneous coronary intervention and/or coronary artery bypass graft surgery), or symptomatic HF (ie, New York Heart Association class I-IV). Patients also had to have two or more of the following risk factors: increased

XSL•FO
RenderX

cholesterol, smoking, diabetes, sedentary lifestyle, and psychosocial risk factors. The presence or absence of each of the risk factors was assessed following the local guidelines in each participant country. For HF patients, additional inclusion criteria were a diagnosis of systolic or diastolic HF and the presence of HF symptoms (eg, exhaustion, shortness of breath, and chest pain). Other general inclusion criteria were an adequate level of the native language, access to the internet at home, having a smartphone compatible with the apps used in the study, and having the skills necessary to use a personal computer and a smartphone.

Exclusion Criteria

Patients with life expectancy less than 1 year, life-threatening comorbidities, a history of psychiatric diseases other than anxiety and depression, and relevant cognitive impairments and those on the waiting list for heart transplantation were excluded from the study.

The reasoning for establishing the exclusion criteria was to prevent the inclusion of patients whose disease severity may critically increase during the intervention. These patients may perceive participation as an extra burden, are more likely to drop out due to illness-related complaints or early mortality, and may be less likely to benefit from a lifestyle intervention owing to severe comorbidities. Patients with mental illness were also excluded because the intervention might become stressful and trigger symptoms in these patients. The selected exclusion criteria are in line with the DO CHANGE 1 trial, safeguarding that the study is not perceived as a burden and meets patients' needs as much as possible.

Intervention

The Do CHANGE 2 program implemented in this trial was similar to that described by Broers et al for Do CHANGE 1 [20]. Patients randomized to the intervention group received devices for measuring key clinical parameters needed for monitoring their CVD, such as a BP monitor, weight scale (in HF patients only), and the Careportal, which allowed monitoring of daily symptoms and an electrocardiogram. The patient's location was monitored by the Moves app. In addition to the aforementioned ICT solutions (also used in the Do CHANGE 1 program), the Do CHANGE 2 program included the Vire app, a purpose-designed app to integrate the input from all the monitoring devices, so that the patient could interact with a unique easy-to-use source of information. The app integrates the information coming from the following apps: the Beddit app (provided with the device under the mattress cover sheet) aimed at monitoring sleep efficiency, the Fitbit app (with the wristband) aimed at measuring physical activity through step count, and the HORUS feature embedded in and aimed at collecting pictures of the different meals of the patients in order to provide diet recommendations. Study participants in the Do CHANGE 2 intervention group were also provided with leaflets (Multimedia Appendix 1) and multimedia resources explaining the use of the Do CHANGE environment.

Like in the Do CHANGE 1 program, patients in the intervention group received a 3-month behavior change program. The program was based on providing care recipients with short messages aimed at disturbing daily routines. Messages were delivered through their mobile phones and suggested them to "do something different." However, unlike the Do CHANGE 1 program, in the Do CHANGE 2 program, behavioral nudges were not only predefined according to the patient's personality profile but also tailored to the patient's behavior, as recorded by the monitoring devices. These behavior-driven messages called to-do's were delivered to the patients based on their current functioning. Patients receive their do's and to-do's through the Careportal or the Vire app, or via SMS, depending on patients' preferences [16].

The GPS data from the Moves app and activity data from the Fitbit device were used to calculate higher abstraction scores called *activity*, *variety*, and *social opportunities*. The to-do's were tailored based on the trends of these scores over time (eg, a patient with declining activity would receive a message that focuses on increasing activity). The granularity of this system is not restricted to one score; alternatively, new scores are calculated for each update of data, and the system determines whether the score is a "target" for a message. Multiple combinations of targets are possible; therefore, a to-do can tackle both *variety* and *social opportunities* if needed based on the scores. A detailed description of the construction process of to-do's can be found in previously published work [23,24].

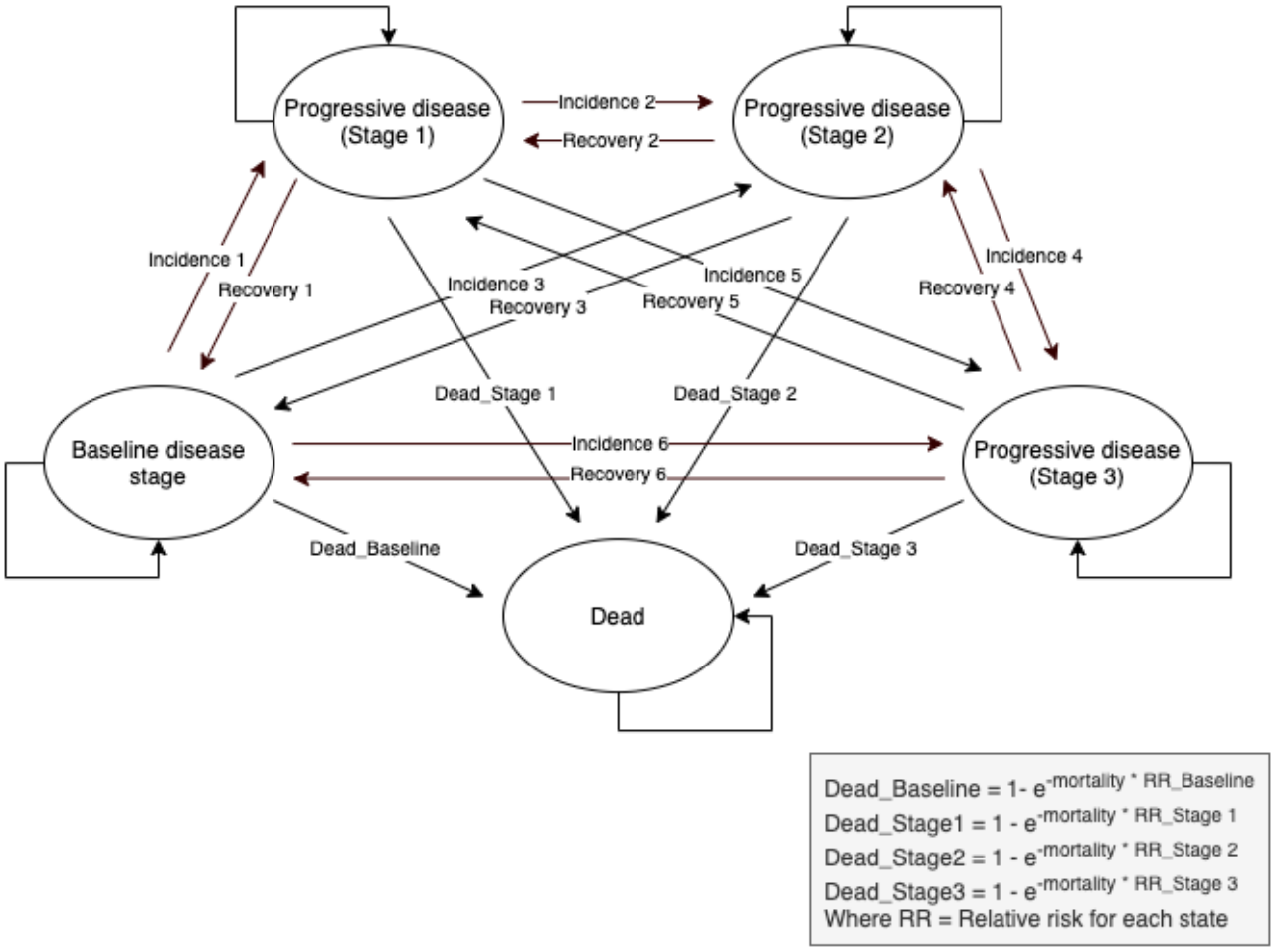
Besides receiving personalized prompts (eg, based on activity levels), patients were contacted each week by the research assistant to check how everything was going and to provide dietary coaching. This might have greatly contributed to the high adherence rate during the first 3 months, as they received personalized feedback. After this period, patients were not contacted in person anymore; however, they were allowed to keep all the devices (eg, Fitbit, Beddit, etc) in order to monitor their behavior for the remaining 3 months.

Cost-Effectiveness Analysis

The cost-effectiveness analysis of the Do CHANGE 2 RCT was conducted using the Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing (MAFEIP) tool [25]. The MAFEIP tool performs a cost-utility analysis through a web app that analyzes incremental costs and effects. The cost-effectiveness estimates are based on the principles of decision analytic modeling and Markov models that assess the impacts that health-related innovations have in terms of health outcomes and resource usage. For Do CHANGE 2, we parametrized the tool on a Markov model of five health states from the perspective of the three service providers (Figure 2) [26].

XSL•FO
RenderX

Figure 2. Markov model of five health states applied for the Do CHANGE cost-effectiveness analysis. Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.



The MAFEIP tool requires the user to provide the following three main types of inputs: (1) the health states and the corresponding transition probabilities between them, (2) the costs, and (3) the utility, for which the EuroQol five-dimension three-level (EQ-5D-3L) questionnaire was used as recommended by the National Institute for Health and Care Excellence [27].

In order to estimate the incremental health gain from a particular intervention delivered, the defined model needs to be run twice. Once with parameter estimates for the respective intervention under assessment (ie, Do CHANGE 2), and once with parameters corresponding to the standard care scenario (ie, TAU). In the model, these two scenarios may differ in terms of transition probabilities (disease incidence, recovery, and mortality), as well as the utility weight and health care and societal costs related to the health states. When the model simulates a hypothetical cohort of patients moving between these health states over time, the differences in survival, utility, and cost accumulate until reaching an estimate of the incremental costs (ΔC) and health effects (ΔE) that can be expected from the intervention under evaluation. Therefore, the tool can be used to estimate the incremental cost-effectiveness ratio (ICER=ΔC/ΔE) or the incremental net monetary benefit (ΔE×λ−ΔC) of one intervention compared with another, where λ is defined as the willingness-to-pay (WTP) threshold for an additional unit of health gain.

Besides the transition probabilities among health states, the utility, and the cost, the tool allows the user to include the relative risks for mortality, the discount rates, and the time horizon for the analysis (cycle length). The parameters included in this analysis and their assessments during the study are explained in detail below.

Definition of Health States

The first stage in the construction of a Markov model is defining the different states of the disease in relation to the important clinical and economical effects of the disease. Evidence suggests that high BP is the predominant risk factor for CVD [28]. Following the scientific evidence and for the purpose of the Do CHANGE assessment, the health states were established based on SBP and DBP, according to the classification of the American Heart Association [29] as follows: baseline disease stage (SBP <120 mmHg and DBP <80 mmHg), progressive disease stage 1 (SBP 120-129 mmHg and DBP <80 mmHg), progressive disease stage 2 (SBP 130-139 mmHg or DBP 80-89 mmHg), progressive disease stage 3 (SBP ≥140 mmHg or DBP ≥90 mmHg), and death. On a side note, the latest stage was not included, as it is considered to be a hypertensive crisis (SBP >180). The transition probabilities were calculated based on the changes between the initial health states (at baseline) and those at 3 months. These transitions can be of incidence (ie, the annual probability for an individual to move from baseline to each

XSL•FO
RenderX

progressive stage of the disease) and recovery (ie, the annual probability of improving).

Cost Estimate

The following two types of costs were considered for both study groups: (1) health care costs (ie, costs associated with the time spent by health care professionals on service provision, including consultations, unplanned hospitalizations, etc) and (2) societal costs (ie, costs associated with the time spent by patients and informal caregivers on care activities).

The data collected by the research team in each country were provided in local currency units (Euro for both Spain and the Netherlands and New Taiwan Dollars for Taiwan). Taiwan prices were converted into a common basis of 2018 Euros using simple exchange rate conversion factors, reflecting the average market exchange rate between New Taiwan Dollars and Euros during the year in question (NT $1=€0.02862). A currency exchange rate of €1=US $1.12 is applicable (average exchange rate for 2018).

For computing the time spent by health care professionals, we considered an average duration of 15 minutes and 25 minutes for a visit to a general practitioner and specialist, respectively. The personnel cost was established based on the average cost for one full - time employee, including employer contributions to social security. The average hourly costs were €29.23 (Spain), €59 (the Netherlands), and €16.88 (Taiwan) for a general practitioner; €20.79 (Spain), €40 (the Netherlands), and €6.33 (Taiwan) for a nurse; and €34.81 (Spain), €113.50 (the Netherlands), and €84.91 (Taiwan) for a specialist. The estimations of cost per bed - day for hospitalizations were €733.56 (Spain) and €1853.57 (the Netherlands), which were obtained by dividing the expenditure for inpatient curative care in hospitals by hospital bed - days for services of curative care (both publicly available) [30]. The corresponding costs for Taiwan were calculated by dividing the average expenses of hospitalization by the number of hospital days, which was €342.50 [31].

Societal costs differed according to the study group. For patients allocated to the control group, we considered the extra travel time spent by patients and caregivers in usual care compared with Do CHANGE 2, whereas for those allocated to the intervention group, we considered the time spent by patients using the service.

Additionally, for patients in the Do CHANGE group, the following costs were added: time spent by professionals in service development and training (4 hours per professional, divided by the number of randomized subjects), time spent by nurses in training patients (30 minutes per participant) and installing the Do CHANGE service ecosystem (45 minutes per patient), and the cost of the devices (including taxes). Data are provided in Euro (2018).

Utility Calculation

Utility was estimated using the EQ-5D-3L tool [32]. The EQ-5D is a standardized questionnaire-based measure of self-rated health-related quality of life developed by the EuroQol Group to provide a simple and generic measure widely used for both clinical and economic appraisals. In the case of the Do CHANGE 2 project, we used the EQ-5D-3L version, which was administered at baseline and at 3 and 6 months.

The resulting scores of the questionnaire were weighted using the trade-off method previously described for Spain [33], the Netherlands [34], and Taiwan [35]. The EQ-5D health states, defined by the EQ-5D descriptive system, were subsequently converted into a single summary index by applying specific weights to each of the levels in each dimension of quality of life. The index was calculated by deducting the appropriate weights from 1, which was the value assigned to full health. In the case of the cost-effectiveness analysis of the Do CHANGE intervention, our interest was to measure the change over time, rather than the absolute values. Therefore, we calculated the changes in utility for each of the five health states and for each of the study conditions and added a common initial measure for the whole sample to each of them. The MAFEIP requires EQ-5D utility scores combined with time indicators to compute quality-adjusted life-years (QALYs) automatically.

Relative Risks of Mortality, Discount Rates, and Time Horizon

The MAFEIP tool allows mortality rates to be internally calculated by using the all-cause mortality rates (age- and sex-dependent) extracted from the Human Mortality Database. The relative risk of mortality is a measure that estimates the mortality in a specific population (eg, people who participated in the Do CHANGE 2 study) compared with (ie, divided by) the mortality in a reference population or condition (in this case, from the Human Mortality Database). The reference condition considers CVD mortality for the population of the specified country (ie, Spain, the Netherlands, and Taiwan).

The discount factors for costs and effects are used to estimate outcomes while taking into account the future costs and health effects, that is, adjusting for differences in the timing of costs (expenditure) compared with health benefits (outcomes). Therefore, adequately applied discount factors express future costs or benefits at today's equivalent value. In Do CHANGE 2, we followed the recommendations from the Health Technology Assessment authorities in each country [36-40]. The discount factors for costs and health outcomes applied in Do CHANGE 2 were 3% for both costs and health outcomes in Taiwan and Spain, and 4% for costs and 1.5% for health factors in the Netherlands.

Finally, the MAFEIP framework allows specifying the number of cycles that the model will run, which represents the timeframe in which the impact of the intervention will be evaluated. Markov models are used to simulate both short-term and long-term processes (ie, CVDs) [41]. In the case of Do CHANGE 2, we wanted to see estimates of the incremental costs (and effects) of the intervention in a time horizon of 5 years. The cycle length we selected is not in line with CVD's etiology (ie, a long disease development process) [41] but considers the nature of the intervention and the maximum time frame it can be sustained in light of the deprecation of the wearable and medical device technology that was used.

XSL•FO
RenderX

WTP Threshold

The balance between the economic benefit and clinical effectiveness varies and is entirely dependent on the relationship between the ICER and the threshold value the society is willing to pay at a specific point of time, which is known as the WTP threshold. The fact that WTP thresholds can be specified after the ICER is calculated raises concerns about researchers selecting WTP thresholds that suit their hypothesis, hence compensating for technology of relatively lower value [42].

While there is an agreement about CVDs being preventable to a certain extent, there has also been a discussion as to whether prevention interventions offer good value for money. Previous research has shown a positive relation between lower lifetime risk for CVD mortality and increased survival and quality of life [43]. Prevention strategies can bring relevant benefits at lower costs relative to most treatment options provided that their cost-effectiveness value is almost always below the accepted societal WTP [44].

For the Do CHANGE 2, we selected a WTP threshold of €15,000/QALY for the three countries, not corresponding to the value recommended by local Health Technology Assessment guidelines. The WTP threshold is lower in all cases. We set a lower WTP threshold in order to avoid the concerns mentioned above and to fit the results of the technology, and considering comparisons with other preventive interventions.

Data Collection and Analysis

The questionnaires, as defined in the study protocol, were loaded into the web tool LimeSurvey [45] and collected by local research assistants. Data from the medical devices (built-in electrocardiogram monitor, blood pressure meter, and weight scale) were collected through the Careportal. The data generated by the wearable devices (Fitbit and Beddit) were continuously monitored and integrated through the Vire app. Information regarding resource consumption (eg, hospitalization costs) was collected by local research assistants from the local electronic medical records.

IBM SPSS Statistics for Windows Version 21.0 (IBM Corp) was used to perform the statistical analyses of the effectiveness study, and R (R Core Team 2018) and RStudio (RStudio Team 2016) were used to calculate the transition probabilities and utilities. We used the MAFEIP tool to perform the cost-effectiveness analysis.

Results

Study Participants

Figure 3 shows the overall flow chart of participant recruitment in the Do CHANGE 2 project. Of the 4540 patients assessed for eligibility at all three sites, 238 were enrolled in the study (120 in the intervention group [Do CHANGE] and 118 in the control group).

Owing to relevant differences in patient recruitment strategies between sites, most patients screened in Spain and the Netherlands agreed to participate, whereas many patients in Taiwan refused to participate. Based on retrospective investigation, it appeared that all inclusion criteria in Taiwan were checked after the patients were approached for participation. Hence, the number of patients that were approached appeared to be much higher. In Spain and the Netherlands, patients who fulfilled the basic inclusion criteria were offered to participate. If the same strategy was applied in Taiwan, the refusal rate would have been much lower. Both Spain and the Netherlands met the target number of participants as defined in the project plan, whereas Taiwan did not reach the planned number of participants.

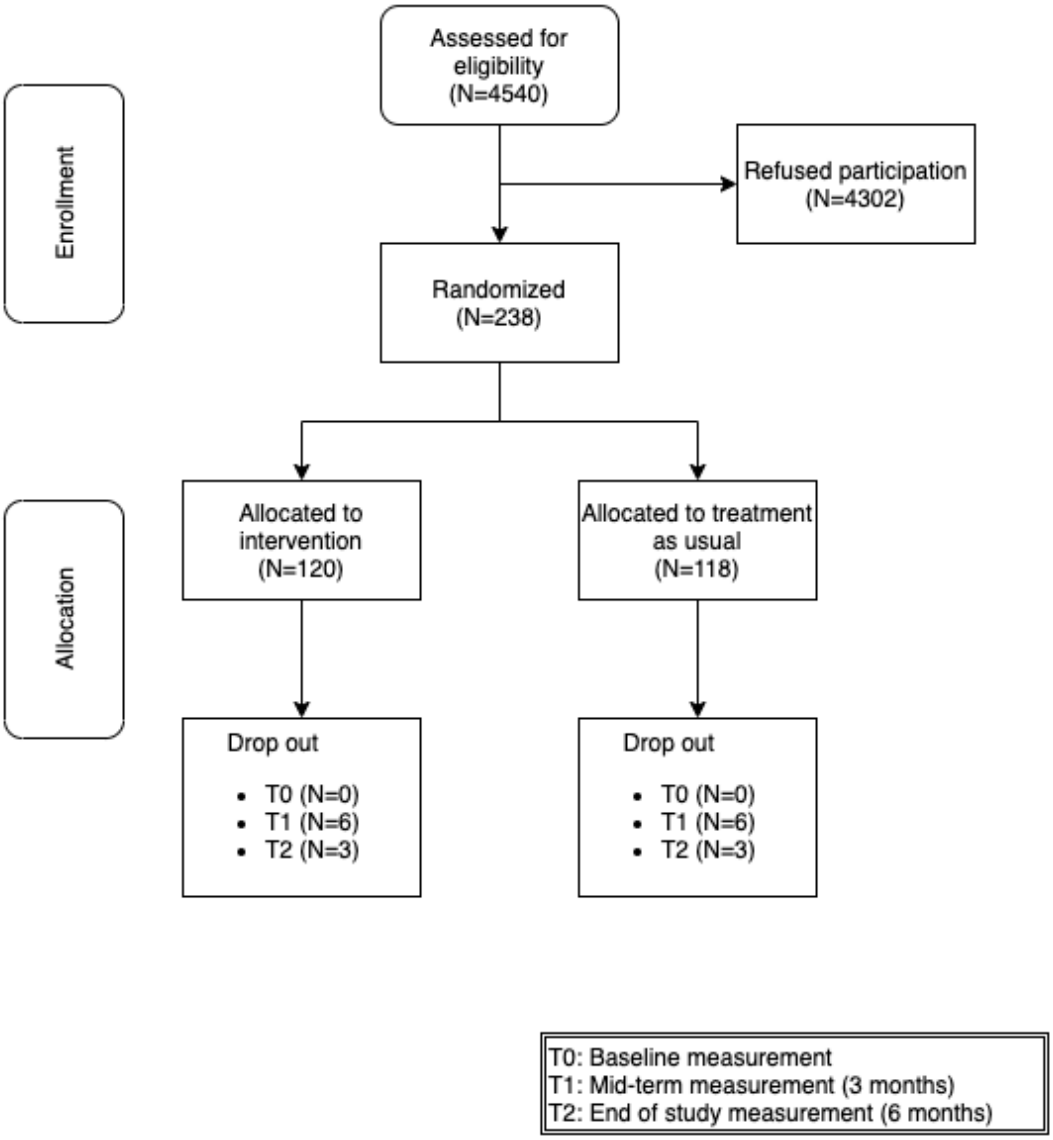
Eighteen patients dropped out before the end of the 6-month follow-up, demonstrating a very high adherence rate. Owing to the personalized nature of the intervention (eg, relevant behavioral prompts and personalized feedback), we expected the adherence to be high. We believe that the nature of the 3-month intervention, where blended and personalized care was provided, combined with the monitoring devices after that period contributed to the high adherence. Patients were engaged in their health management, and therefore, they might be more willing to proceed with monitoring.

Of the 114 participants who were in the program for at least 3 months, 72 (60%) claimed to have carried out all the nudges provided by the DSD and 86 (72%) reckoned the program was useful. Reasons for not adhering to the program were mainly having no time (8/84, 6.7%), not feeling like it (2/84, 1.7%), and falling ill (2/84, 1.7%). One of the participants who quit the program disclosed that being confronted with the illness on a daily basis became too stressful. Moreover, in some cases, the confrontation for some partners to deal with the illness of their husband or wife caused anxiety.

Table 1 summarizes the main demographic characteristics of the three participating countries. None of the variables collected showed relevant systematic differences between study conditions at baseline, except for Spain, with participants in the Do CHANGE 2 group being younger (Do CHANGE 2 group vs TAU group: mean 53.8 years, SD 15.8 years vs mean 67.4 years, SD 7.5 years), having a higher education (mean 14.5 years, SD 6.3 years vs mean 9.1 years, SD 5.5 years), and showing a higher employed proportion (17/37, 45.9% vs 4/37, 10.8%). Multimedia Appendix 2 presents the clinical characteristics of the study sample, medication, and psychological symptoms. The only significant difference at baseline was observed in psychotropic medication for participants in the TAU group in Spain (P=.03), which is consistent with the population in the control group being older.

XSL•FO
RenderX

Figure 3. Flow chart of participant recruitment (aggregated numbers for Spain, the Netherlands, and Taiwan).



XSL•FO
RenderX

Table 1. Demographic baseline characteristics of the total sample (N=238).

Characteristic	Spain (N=75)	The Netherlands (N=75)	Taiwan (N=88)	Total (N=238)
Sample size, n (%)				
Do CHANGE[a] 2	38 (50.7)	38 (50.7)	44 (50.0)	120 (50.4)
TAU[b]	37 (49.3)	37 (49.3)	44 (50.0)	118 (49.6)
Total	75 (100.0)	75 (100.0)	88 (100.0)	238 (100.0)
Age (years), mean (SD)				
Do CHANGE 2	53.8 (15.8)	63.0 (9.2)	58.2 (9.9)	58.3 (12.3)
TAU	67.4 (7.5)	63.9 (7.4)	56.7 (9.1)	62.3 (9.2)
Total	60.5 (14.1)	63.4 (8.3)	57.5 (9.5)	60.3 (11.1)
Gender (male), n (%)				
Do CHANGE 2	27 (71.1)	32 (84.2)	30 (68.2)	89 (74.2)
TAU	19 (51.4)	29 (78.4)	38 (86.4)	86 (72.9)
Total	46 (61.3)	61 (81.3)	68 (77.3)	175 (73.5)
Education (years), mean (SD)				
Do CHANGE 2	14.5 (6.3)	12.9 (5.1)	14.9 (5.5)	14.1 (5.7)
TAU	9.1 (5.5)	13.16 (7.9)	16.4 (5.0)	13.1 (6.9)
Total	11.8 (6.5)	13.0 (6.6)	15.7 (5.3)	13.6 (6.3)
Marital status (partner), n (%)				
Do CHANGE 2	27 (71.1)	34 (89.5)	39 (88.6)	100 (83.3)
TAU	27 (73.0)	33 (89.2)	42 (95.5)	102 (86.4)
Total	54 (72.0)	67 (89.3)	81 (92.0)	202 (84.9)
Working status (paid job), n (%)				
Do CHANGE 2	17 (45.9)	13 (34.2)	26 (59.1)	56 (46.7)
TAU	4 (10.8)	16 (43.2)	28 (63.6)	48 (40.7)
Total	21 (28.0)	29 (38.7)	54 (61.4)	104 (43.7)
Smoking (yes), n (%)				
Do CHANGE 2	7 (18.4)	3 (7.9)	2 (4.5)	12 (10.0)
TAU	5 (13.5)	7 (18.9)	4 (9.1)	16 (13.6)
Total	12 (16.0)	10 (13.3)	6 (6.8)	28 (11.8)

[a]Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.

[b]TAU: treatment as usual.

Model Input

The model input for the MAFEIP tool included data regarding the health states (and transition probabilities) of study participants, the costs associated with each study group, and the utility estimate. Table 2 summarizes the distribution of study participants across the MAFEIP health states at study start, as well as the transition probabilities between these states, computed by considering data recorded at month 3. Table 3 summarizes the total health care and societal costs for each group and each state. The detailed amounts for each type of health care and societal cost are provided in Multimedia Appendix 3. The specific costs associated with the implementation of the Do CHANGE environment are presented in Table 4. The utility values calculated from the EQ-5D-3L scores and the estimated utility computed by adding the initial common measure are described in Table 5. No systematic differences were observed between study conditions at baseline. The utility values for the whole study sample in Spain, the Netherlands, and Taiwan were 0.897, 0.842, and 0.854, respectively.

XSL•FO
RenderX

Table 2. Frequency and percentage of patients across the various health states (N=207).

Variable	Do CHANGE 2[a,b] (N=92)			TAU[b,c] (N=115)		
	Spain (n=27)	The Netherlands (n=29)	Taiwan (n=36)	Spain (n=36)	The Netherlands (n=37)	Taiwan (n=42)
Health states at study start, n (%)[d]						
Baseline disease stage	5 (19.4%)	4 (14.3%)	9 (25.0%)	5 (13.5%)	5 (13.3%)	8 (18.6%)
Progressive disease stage 1	3 (11.1%)	1 (2.9%)	0	1 (2.7%)	2 (6.7%)	3 (7.0%)
Progressive disease stage 2	10 (36.1%)	9 (31.4%)	15 (40.9%)	17 (46.0%)	10 (26.7%)	19 (44.2%)
Progressive disease stage 3	9 (33.3%)	15 (51.4%)	12 (34.1%)	14 (37.8%)	20 (53.3%)	13 (30.2%)
Transition probabilities, %						
Incidence 1 (baseline disease stage to progressive disease stage 1)	14.3%	0.0%	18.2%	80.0%	25.0%	12.5%
Recovery 1 (progressive disease stage 1 to baseline)	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
Incidence 2 (baseline disease stage to progressive disease stage 2)	14.3%	20.0%	0.0%	20.0%	0.0%	37.5%
Recovery 2 (progressive disease stage 2 to baseline)	7.7%	0.0%	5.6%	17.7%	0.0%	10.5%
Incidence 3 (progressive disease stage 1 to stage 2)	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Recovery 3 (progressive disease stage 2 to stage 1)	38.5%	0.0%	11.1%	29.4%	12.5%	10.5%
Incidence 4 (baseline disease stage to progressive disease stage 3)	14.3%	0.0%	0.0%	0.0%	50.0%	12.5%
Recovery 4 (progressive disease stage 3 to baseline)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Incidence 5 (progressive disease stage 1 to stage 3)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Recovery 5 (progressive disease stage 3 to stage 1)	0.0%	0.0%	0.0%	0.0%	0.0%	7.7%
Incidence 6 (progressive disease stage 2 to stage 3)	38.5%	27.3%	16.7%	5.9%	0.0%	47.4%
Recovery 6 (progressive disease stage 3 to stage 2)	50.0%	5.6%	33.3%	42.9%	0.0%	30.8%

[a]Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.

[b]Distribution of study participants at study start and the corresponding transition probabilities (in percentage).

[c]TAU: treatment as usual.

[d]Baseline disease stage: systolic blood pressure (SBP) <120 mmHg and diastolic blood pressure (DBP) <80 mmHg; Progressive disease stage 1: SBP 120-129 mmHg and DBP <80 mmHg; Progressive disease stage 2: SBP 130-139 mmHg or DBP 80-89 mmHg; Progressive disease stage 3: SBP ≥140 mmHg or DBP ≥90 mmHg.

XSL•FO
RenderX

Table 3. Total health care and societal costs for each of the study groups (N=207).

Variable	Do CHANGE 2[a,b] (N=92)			TAU[b,c] (N=115)		
	Spain (n=27)	The Netherlands (n=29)	Taiwan (n=36)	Spain (n=36)	The Netherlands (n=37)	Taiwan (n=42)
Health care costs[d]						
Baseline disease stage	299.90	489.82	156.94	646.08	343.97	114.25
Progressive disease stage 1	729.25	166.43	244.96	1284.41	88.76	71.45
Progressive disease stage 2	942.39	240.36	161.58	2381.76	313.00	93.44
Progressive disease stage 3	2176.32	240.36	138.67	3484.41	88.76	114.94
Societal costs[d]						
Baseline disease stage	309.61	512.90	158.94	648.48	367.43	113.35
Progressive disease stage 1	737.38	198.51	247.90	1289.75	76.53	70.71
Progressive disease stage 2	953.98	277.55	163.44	2386.12	323.81	93.06
Progressive disease stage 3	2198.96	287.20	140.62	3485.76	73.47	115.09

[a]Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.

[b]Data are presented in € (2018; €1=US $1.12). The detailed costs of each category are provided in Multimedia Appendix 3.

[c]TAU: treatment as usual.

[d]Baseline disease stage: systolic blood pressure (SBP) <120 mmHg and diastolic blood pressure (DBP) <80 mmHg; Progressive disease stage 1: SBP 120-129 mmHg and DBP <80 mmHg; Progressive disease stage 2: SBP 130-139 mmHg or DBP 80-89 mmHg; Progressive disease stage 3: SBP ≥140 mmHg or DBP ≥90 mmHg.

Table 4. Costs associated with the implementation of the Do CHANGE 2 intervention (N=92).

Variable	Spain[a] (n=27)	The Netherlands[a] (n=29)	Taiwan[a] (n=36)
Time spent by professionals[b] (overhead of 18%)	50.02	99.79	19.19
Time spent by specialists (service development, receiving training, and adaptation)	1.86	7.12	4.53
Time spent by nurses (service development, receiving training, and adaptation)	1.39	2.67	0.42
Time spent by nurses on training provision to patients	25.99	50	7.91
Time spent by nurses on installation of the Do CHANGE[c] ecosystem	20.79	40	6.33
Cost of the set of devices included within the Do CHANGE ecosystem	748.99	748.99	748.99
Total	799.01	848.78	768.18

[a]Data are presented in € (2018; €1=US $1.12).

[b]For the personnel cost, we use the average cost for one full - time employee including employer contributions to social security. The average hourly costs are as follows: €29.23 (Spain), €59 (the Netherlands), and €16.88 (Taiwan) for a physician; €20.79 (Spain), €40 (the Netherlands), and €6.33 (Taiwan) for a nurse; and €34.81 (Spain), €113.50 (the Netherlands), and €84.91 (Taiwan) for a specialist.

[c]Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.

XSL•FO
RenderX

Table 5. Calculation of utility (N=207).

Disease stage[a] and assessment	Spain (N=63)			The Netherlands (N=66)			Taiwan (N=78)		
	Do CHANGE[b] 2 (N=27)	TAU[c] (N=36)	*P* value[d]	Do CHANGE 2 (N=29)	TAU (N=37)	*P* value[d]	Do CHANGE 2 (N=36)	TAU (N=42)	*P* value[d]
Baseline disease stage									
M0[e]	0.896	0.869	.74	0.854	0.936	.40	0.875	0.638	.01
M3[f]	0.900	0.950	.49	0.931	0.904	.83	0.911	0.847	.59
Δ[g]	0.004	0.081		0.077	−0.032		0.036	0.209	
Progressive disease stage 1									
M0	0.871	0.719	—[h]	0.861	0.807	—	0.726	1	—
M3	0.853	0.898	.57	0.861	0.904	—	0.726	1	.19
Δ	−0.018	0.179		0	0.097		0	0	
Progressive disease stage 2									
M0	0.912	0.875	.49	0.896	0.821	.23	0.877	0.895	.75
M3	0.938	0.853	.07	0.886	0.825	.42	0.841	0.883	.56
Δ	0.026	−0.022		−0.010	0.004		−0.036	−0.012	
Progressive disease stage 3									
M0	0.964	0.889	.09	0.805	0.843	.42	0.832	0.870	.62
M3	0.944	0.866	.21	0.852	0.872	.64	0.766	0.838	.32
Δ	−0.020	−0.023		0.047	0.029		−0.066	−0.032	

[a]Baseline disease stage: systolic blood pressure (SBP) <120 mmHg and diastolic blood pressure (DBP) <80 mmHg; Progressive disease stage 1: SBP 120-129 mmHg and DBP <80 mmHg; Progressive disease stage 2: SBP 130-139 mmHg or DBP 80-89 mmHg; Progressive disease stage 3: SBP ≥140 mmHg or DBP ≥90 mmHg.

[b]Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem.

[c]TAU: treatment as usual.

[d]A *P* value <.05 is considered significant.

[e]M0: baseline assessment.

[f]M3: assessment at 3 months.

[g]Δ: M3 – M0.

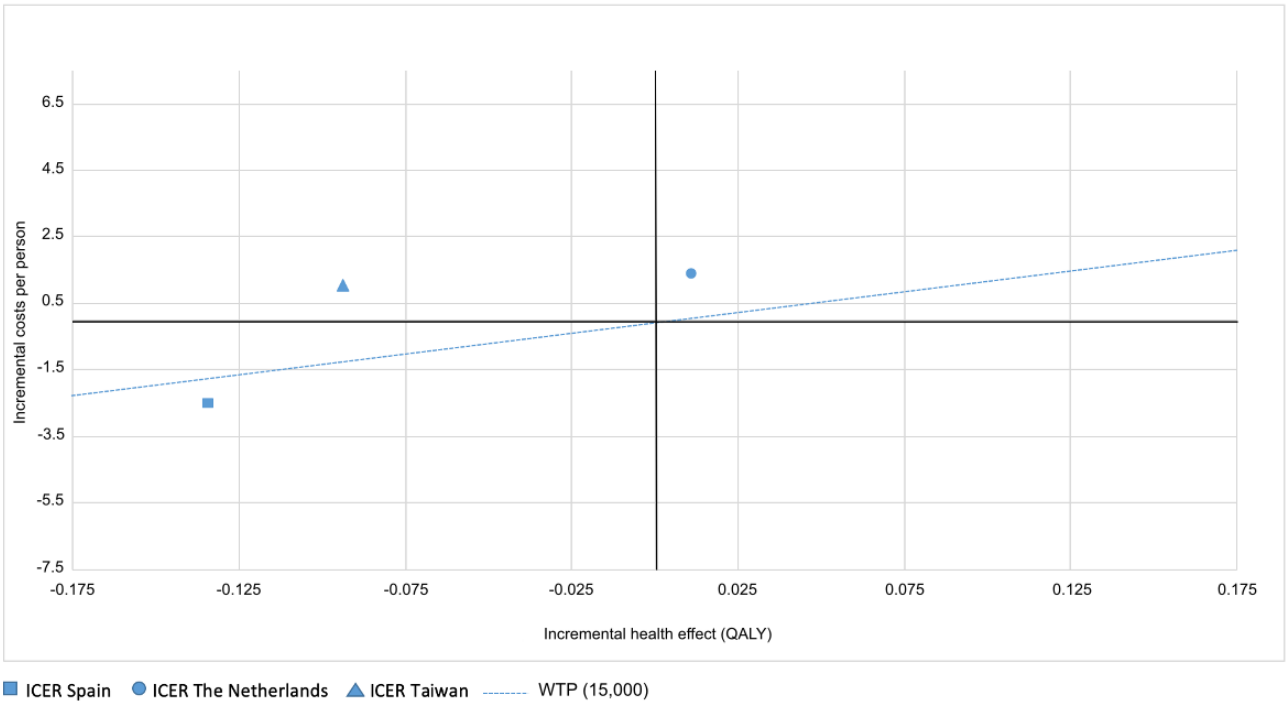
[h]Not enough data to calculate a *P* value.

Cost-Effectiveness

After cleaning the data, 207 participants were included in the cost-effectiveness analysis (92 in the intervention group and 115 in the control group). The Do CHANGE 2 intervention was less costly in Spain (incremental cost was −€2514.90) and more costly in the Netherlands and Taiwan (incremental costs were €1373.59 and €1062.54, respectively). Figure 4 shows the cost-effectiveness plane for the three countries. The cost-effectiveness plane plots the incremental cost of the intervention on the y-axis and the incremental health outcome (measured in QALYs) on the x-axis. The diagonal line represents the WTP per additional QALY gained, which is the maximum amount that the society is willing to give in exchange for a better quality of life. Different thresholds may also be selected. Depending on the location of the ICER point in this plane, one would be able to interpret whether an intervention is cost-effective. When the ICER point is within the lower-right quadrant, it means the intervention is accepted (it is more effective and cheaper), and when it is within the upper-left quadrant, it means that the intervention is not accepted (it is less effective and more expensive). If the ICER point lies in the other two quadrants, then the intervention may or may not be accepted depending on the ICER and WTP threshold values.

XSL•FO
RenderX

Figure 4. Cost-effectiveness plane for the Do CHANGE intervention in Spain, the Netherlands, and Taiwan. The dotted line shows the willingness-to-pay threshold of €15,000 per QALY. Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life-year; WTP: willingness to pay.



Compared with usual care, the effectiveness of the Do CHANGE 2 program in terms of QALY gains was slightly higher in the Netherlands (incremental effect of 0.011) and lower in Spain and Taiwan (incremental effects of −0.134 and −0.094, respectively). Even though the Do CHANGE program was more effective than usual care in the Netherlands, the relative costs for gained utility (€124,489.27 per QALY) were too high to accept this intervention. Taken together, the Do CHANGE intervention would only be accepted in Spain, where it would help save €18,769.05 per QALY.

We also calculated the incremental cost and health-related quality of life for every age-gender combination in the specified target population. The data are presented in Multimedia Appendix 4.

Discussion

Principal Findings

In this cost-effectiveness analysis of an ICT-based intervention to change the health behavior of patients with CVD (the Do CHANGE program) assessed in a multicenter RCT, we found that the ICER strongly varied depending on the country where the intervention was applied. The Do CHANGE 2 program was slightly more effective than usual care in the Netherlands only, albeit at an incremental cost too high to accept the intervention at the selected WTP threshold (€15,000 per QALY). The same intervention was less effective but less costly than usual care in Spain. In Taiwan, the intervention resulted in the dominated option (less effective and more expensive). Therefore, implementation of the Do CHANGE 2 intervention is only recommended in Spain, where it could allow saving financial costs taking into account the costs and effects of the intervention.

We further tested the results with higher WTP thresholds (ie, €30,000 per QALY), with results remaining in the same line.

Contextualization With Previous Work

There is a large body of evidence showing that ICT solutions, including mobile-based telemonitoring, improve the quality and outcomes of care in patients with CVD [46]. Unfortunately, the cost-effectiveness assessment is often disregarded, and many studies reporting cost information do not meet a quality standard high enough to determine the cost-effectiveness or cost-utility of the intervention [46,47].

Regardless of the quality in reporting of individual studies, evidence on the cost-effectiveness of ICT-based lifestyle interventions is rather controversial, and many authors have acknowledged difficulties in drawing strong conclusions in this regard [48-50]. Overall, cost-effectiveness evaluations of secondary and tertiary prevention strategies for patients with CVD are challenged by the multiple factors influencing the outcomes and costs, such as baseline cardiovascular risk, the cost of drugs or other interventions, reimbursement procedures, and implementation of preventive strategies [1]. In the case of telemedicine approaches, it has been recognized that cost-effectiveness depends largely on local aspects of the individual service (and care as usual) being evaluated, and a service may be highly cost-effective in one context but highly ineffective when transferred to another context [47]. This was the case in our analysis, which yielded controversial results regarding the cost-effectiveness of the intervention in the different countries involved. Importantly, the success of an ICT-based lifestyle intervention strongly depends on the willingness of individuals to adopt the intervention, which is likely to be associated with cultural constraints and, therefore, to be country specific.

XSL•FO
RenderX

These differences were particularly pervasive between the Netherlands and Taiwan, where the cost-effectiveness planes showed an almost opposite profile, although the same intervention was implemented. We associate this situation with the majority of patients recruited in Taiwan having hypertension as the primary diagnosis and medical consultations involving health care professionals (ie, physicians) in the Netherlands being too expensive for them to devote time to a prevention intervention that could perhaps be conducted by nurses. This finding supports the need to evaluate the cost-effectiveness of these types of interventions within each context in order to provide the various stakeholders with evidence to understand the financial consequences of scaling up ICT solutions for health care.

Limitations

The main limitation of our trial was the sample size, which was constrained by budget restrictions. The discrepancy between the target and the actual sample size was mainly due to technical difficulties in recruiting participants, who had to enroll in the trial for a minimum duration of 6 months. The low sample size might have constrained the representativeness of the results and made them more sensitive to biases associated with patients with extreme behaviors (ie, outliers). Actually, the extremely low number of participants within, for instance, the states "baseline" and "progressive disease stage 1," might explain the unrealistic utilities of these patients before the intervention (eg, 0.962 and 1, respectively, for the control group), which were considerably higher than the average reported in larger RCTs involving HF (utility 0.84) [51]. Another example associated with this limitation is the relevant differences for patients allocated to the intervention group in Spain, who were younger and had higher education.

Second, an acknowledgment must be made regarding the limitation associated with the heterogeneous characteristics of the study sample for the primary diagnosis and cultural setting (ie, Spain, the Netherlands, and Taiwan), which may have contributed to the heterogeneous results across countries.

Unlike other cost-effectiveness analyses, we did not consider the contribution of medication to the health care costs. Although this might have increased the accuracy of the absolute costs, from a clinical point of view, it is unrealistic that the prescribed medicines would change throughout a 3-month time lapse. Since our main interest was assessing the change in costs rather than describing the actual values, we considered that it was more appropriate to exclude this concept from the cost-effectiveness analysis.

Finally, although RCTs are considered the gold standard for assessing cost-effectiveness, some authors criticize that they might miss information regarding how the intervention fits into routine practice [15].

Conclusions

Our results suggest that the Do CHANGE 2 environment may help reduce health care costs associated with the management of patients with CVD in certain settings. However, changing health behavior and assessing the impact of this change on health care and societal costs remain big challenges. In line with previous research in this field, our assessment does not allow drawing strong conclusions in this regard. Irrespective of the specific cost-effectiveness of the Do CHANGE 2 program, our results highlight the high heterogeneity that ICT-based interventions might show depending on the country where they are implemented and stress the need for assessing each intervention in all areas before scaling up implementation.

Acknowledgments

The continuous development of the MAFEIP tool is funded by the European Commission's Horizon 2020 program under the projects WE4AHA (grant number: 769705) and DigitalHealthEurope (grant number: 826353). The Do CHANGE study was funded by the European Commission's Horizon 2020 program (grant number: 463735). The funder had no role in data collection or analysis, the decision to publish, or the preparation of the manuscript. The Do CHANGE team received funding for Research and Innovation from the European Union for the project. Within the consortium, two small- and medium-sized enterprises (Docobo and Do Something Different) and one start-up (Onmi) were financially supported to develop their products. The authors would like to acknowledge all the patients who were willing to participate in this study within the three countries and were willing to share their data, thus making this work possible. Moreover, the authors would like to thank all students and health care professionals for their support in recruitment and data management. The authors also want to acknowledge the inestimable work conducted by the Do CHANGE consortium members (www.do-change.eu). Lastly, the authors would like to thank Gerard Carot-Sans (PhD) for providing medical writing assistance during the preparation of the manuscript and Leah Bührmann (MSc) for her help with SPSS analysis.

Authors' Contributions

JPJ, JWMGW, MH, EB, and DVB contributed to the study design and data collection. JPJ, MW, and DVB conducted the statistical analyses. JPJ drafted the manuscript. All authors critically revised and approved the final version of the manuscript. FLV and FF supervised the study.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Do CHANGE leaflet.
[PDF File (Adobe PDF File), 796 KB - jmir_v22i7e17351_app1.pdf]

Multimedia Appendix 2
Clinical characteristics of the study sample, medication, and psychological symptoms.
[PDF File (Adobe PDF File), 179 KB - jmir_v22i7e17351_app2.pdf]

Multimedia Appendix 3
Health care and societal costs.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v22i7e17351_app3.xlsx]

Multimedia Appendix 4
Incremental costs and effects (age and gender specific) per pilot site.
[PDF File (Adobe PDF File), 674 KB - jmir_v22i7e17351_app4.pdf]

References

1. Piepoli M, Hoes A, Agewall S, Albus C, Brotons C, Catapano A, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
2. Epidemiology of cardiovascular disease. Nature Reviews Cardiology. 2019 Feb 28. URL: https://www.nature.com/collections/bedbejdbij [accessed 2019-07-09]
3. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
4. Broers ER, Gavidia G, Wetzels M, Ribas V, Ayoola I, Piera-Jimenez J, et al. Usefulness of a Lifestyle Intervention in Patients With Cardiovascular Disease. Am J Cardiol 2020 Feb 01;125(3):370-375 [FREE Full text] [doi: 10.1016/j.amjcard.2019.10.041] [Medline: 31761149]
5. Charlson ME, Wells MT, Peterson JC, Boutin-Foster C, Ogedegbe GO, Mancuso CA, et al. Mediators and moderators of behavior change in patients with chronic cardiopulmonary disease: the impact of positive affect and self-affirmation. Transl Behav Med 2014 Mar;4(1):7-17 [FREE Full text] [doi: 10.1007/s13142-013-0241-0] [Medline: 24653772]
6. Riegel B, Moser D, Buck H, Dickson V, Dunbar S, Lee C, American Heart Association Council on CardiovascularStroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of CareOutcomes Research. Self-Care for the Prevention and Management of Cardiovascular Disease and Stroke: A Scientific Statement for Healthcare Professionals From the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9) [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
7. Omboni S. Connected Health in Hypertension Management. Front Cardiovasc Med 2019;6:76 [FREE Full text] [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
8. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017;3:51 [FREE Full text] [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
9. Brunetti ND, Scalvini S, Acquistapace F, Parati G, Volterrani M, Fedele F, et al. Telemedicine for cardiovascular disease continuum: A position paper from the Italian Society of Cardiology Working Group on Telecardiology and Informatics. Int J Cardiol 2015 Apr 01;184:452-458. [doi: 10.1016/j.ijcard.2015.02.023] [Medline: 25755064]
10. Franklin NC, Lavie CJ, Arena RA. Personal health technology: A new era in cardiovascular disease prevention. Postgrad Med 2015 Mar;127(2):150-158 [FREE Full text] [doi: 10.1080/00325481.2015.1015396] [Medline: 25690685]
11. Fernandez-Luque L, Aupetit M, Palotti J, Singh M, Fadlelbari A, Baggag A. Health Lifestyle Data-Driven Applications Using Pervasive Computing. In: Big Data, Big Challenges: A Healthcare Perspective. Lecture Notes in Bioengineering. Switzerland: Springer; 2019:115-126.
12. McConnachie A, Walker A, Robertson M, Marchbank L, Peacock J, Packard CJ, et al. Long-term impact on healthcare resource utilization of statin treatment, and its cost effectiveness in the primary prevention of cardiovascular disease: a record linkage study. Eur Heart J 2014 Feb;35(5):290-298 [FREE Full text] [doi: 10.1093/eurheartj/eht232] [Medline: 23839541]

XSL•FO
RenderX

13. Mistry H, Morris S, Dyer M, Kotseva K, Wood D, Buxton M, EUROACTION study group. Cost-effectiveness of a European preventive cardiology programme in primary care: a Markov modelling approach. BMJ Open 2012;2(5):e001029 [FREE Full text] [doi: 10.1136/bmjopen-2012-001029] [Medline: 23065443]
14. Ramsey SD, McIntosh M, Etzioni R, Urban N. SIMULATION MODELING OF OUTCOMES AND COST EFFECTIVENESS. Hematology/Oncology Clinics of North America 2000 Aug;14(4):925-938. [doi: 10.1016/s0889-8588(05)70319-1]
15. Bergmo T. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]
16. Habibović M, Broers E, Piera-Jimenez J, Wetzels M, Ayoola I, Denollet J, et al. Enhancing Lifestyle Change in Cardiac Patients Through the Do CHANGE System ("Do Cardiac Health: Advanced New Generation Ecosystem"): Randomized Controlled Trial Protocol. JMIR Res Protoc 2018 Feb 08;7(2):e40 [FREE Full text] [doi: 10.2196/resprot.8406] [Medline: 29422454]
17. Do CHANGE. URL: https://www.do-change.eu/ [accessed 2019-12-01]
18. Fletcher B, Hanson J, Page N, Pine K. FIT – Do Something Different. Swiss Journal of Psychology 2011 Jan;70(1):25-34. [doi: 10.1024/1421-0185/a000035]
19. Pine K, Fletcher B. Time to shift brain channels to bring about effective changes in health behaviour. Perspect Public Health 2014 Jan;134(1):16-17. [doi: 10.1177/1757913913514705] [Medline: 24395840]
20. Broers ER, Kop WJ, Denollet J, Widdershoven J, Wetzels M, Ayoola I, et al. A Personalized eHealth Intervention for Lifestyle Changes in Patients With Cardiovascular Disease: Randomized Controlled Trial. J Med Internet Res 2020 May 22;22(5):e14570 [FREE Full text] [doi: 10.2196/14570] [Medline: 32441658]
21. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]
22. Duckworth RA. Evolution of Personality: Developmental Constraints on Behavioral Flexibility. The Auk 2010 Oct;127(4):752-758. [doi: 10.1525/auk.2010.127.4.752]
23. Broers ER, Widdershoven J, Denollet J, Lodder P, Kop WJ, Wetzels M, et al. Personalized eHealth Program for Life-style Change. Psychosomatic Medicine 2020;82(4):409-419. [doi: 10.1097/psy.0000000000000802]
24. Ayoola I. Advanced cardiac-health transformation ecosystem ACTivatE: contributions to the development of a health ecosystem for integrated cardiovascular disease management and an approach to changing behaviour. The Netherlands: Technische Universiteit Eindhoven; Sep 26, 2018:1-263.
25. Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing. European Comission. URL: https://www.mafeip.eu/the-tool [accessed 2019-12-01]
26. Alhambra-Borrás T, Durá-Ferrandis E, Ferrando-García M. Effectiveness and Estimation of Cost-Effectiveness of a Group-Based Multicomponent Physical Exercise Programme on Risk of Falling and Frailty in Community-Dwelling Older Adults. Int J Environ Res Public Health 2019 Jun 13;16(12) [FREE Full text] [doi: 10.3390/ijerph16122086] [Medline: 31200434]
27. Position statement on use of the EQ-5D-5L valuation set for England. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l [accessed 2019-12-01]
28. Fuchs FD, Whelton PK. High Blood Pressure and Cardiovascular Disease. Hypertension 2020 Feb;75(2):285-292. [doi: 10.1161/hypertensionaha.119.14240]
29. Whelton P, Carey R, Aronow W, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
30. Eurostat - Microdata. European Commission. 2017. URL: https://ec.europa.eu/eurostat/web/microdata [accessed 2019-12-01]
31. Cheng K, Lu C, Chung Y, Huang M, Shen H, Chen H, et al. ICU service in Taiwan. J Intensive Care 2014;2(1):8 [FREE Full text] [doi: 10.1186/2052-0492-2-8] [Medline: 25908981]
32. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
33. Badia X, Roset M, Herdman M, Kind P. A Comparison of United Kingdom and Spanish General Population Time Trade-off Values for EQ-5D Health States. Med Decis Making 2016 Jul 02;21(1):7-16. [doi: 10.1177/0272989x0102100102]
34. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006 Oct;15(10):1121-1132. [doi: 10.1002/hec.1124] [Medline: 16786549]
35. Lee H, Hung M, Hu F, Chang Y, Hsieh C, Wang J. Estimating quality weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. J Formos Med Assoc 2013 Nov;112(11):699-706 [FREE Full text] [doi: 10.1016/j.jfma.2012.12.015] [Medline: 24183199]

XSL•FO
RenderX

36. Mora-Ripoll R, Gilabert-Perramon A, Oliva-Moreno J, Puig-Junoy J. Guidance For Economic Evaluation and Budget Impact Analysis For Pharmaceuticals in Catalonia (Spain). Value Health 2014 Nov;17(7):A447 [FREE Full text] [doi: 10.1016/j.jval.2014.08.1193] [Medline: 27201217]
37. Rovira J, Antoñanzas F. Economic analysis of health technologies and programmes. A Spanish proposal for methodological standardisation. Pharmacoeconomics 1995 Sep;8(3):245-252. [doi: 10.2165/00019053-199508030-00007] [Medline: 10155620]
38. López-Bastida J, Oliva J, Antoñanzas F, García-Altés A, Gisbert R, Mar J, et al. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ 2010 Oct;11(5):513-520. [doi: 10.1007/s10198-010-0244-4] [Medline: 20405159]
39. PE Guidelines Around The World: The Netherlands. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). 2017. URL: https://tools.ispor.org/PEguidelines/countrydet.asp?c=22&t=1 [accessed 2019-12-01]
40. Demarteau N, Tang C, Chen H, Chen C, Van Kriekinge G. Cost-effectiveness analysis of the bivalent compared with the quadrivalent human papillomavirus vaccines in Taiwan. Value Health 2012;15(5):622-631 [FREE Full text] [doi: 10.1016/j.jval.2012.02.012] [Medline: 22867770]
41. Saidi O, O'Flaherty M, Zoghlami N, Malouche D, Capewell S, Critchley J, et al. Comparing Strategies to Prevent Stroke and Ischemic Heart Disease in the Tunisian Population: Markov Modeling Approach Using a Comprehensive Sensitivity Analysis Algorithm. Comput Math Methods Med 2019;2019:2123079 [FREE Full text] [doi: 10.1155/2019/2123079] [Medline: 30838048]
42. Padula W, Chen H, Phelps C. PRM39 - Is the Choice of Willingness-to-Pay Threshold in Cost-Utility Analysis Endogenous to the Resulting Value of the Technology? Value in Health 2018 Oct;21:S362. [doi: 10.1016/j.jval.2018.09.2162]
43. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PW, et al. Prediction of Lifetime Risk for Cardiovascular Disease by Risk Factor Burden at 50 Years of Age. Circulation 2006 Feb 14;113(6):791-798. [doi: 10.1161/circulationaha.105.548206]
44. Eze-Nliam CM, Zhang Z, Weiss SA, Weintraub WS. Cost-effectiveness Assessment of Cardiac Interventions: Determining a Socially Acceptable Cost Threshold. Interv Cardiol (Lond) 2014;6(1):45-55 [FREE Full text] [doi: 10.2217/ica.13.81] [Medline: 26136831]
45. Lime Survey. URL: https://www.limesurvey.org/es/ [accessed 2019-12-01]
46. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile Health Devices as Tools for Worldwide Cardiovascular Risk Reduction and Disease Management. Circulation 2015 Nov 24;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
47. Whitten P, Mair F, Haycox A, May C, Williams T, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [doi: 10.1136/bmj.324.7351.1434] [Medline: 12065269]
48. Young LA, Buse JB, Weaver MA, Vu MB, Mitchell CM, Blakeney T, Monitor Trial Group. Glucose Self-monitoring in Non-Insulin-Treated Patients With Type 2 Diabetes in Primary Care Settings: A Randomized Trial. JAMA Intern Med 2017 Jul 01;177(7):920-929 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1233] [Medline: 28600913]
49. Greving JP, Kaasjager HA, Vernooij JW, Hovens MM, Wierdsma J, Grandjean HM, et al. Cost-effectiveness of a nurse-led internet-based vascular risk factor management programme: economic evaluation alongside a randomised controlled clinical trial. BMJ Open 2015 May 20;5(5):e007128 [FREE Full text] [doi: 10.1136/bmjopen-2014-007128] [Medline: 25995238]
50. Monahan M, Jowett S, Nickless A, Franssen M, Grant S, Greenfield S, et al. Cost-Effectiveness of Telemonitoring and Self-Monitoring of Blood Pressure for Antihypertensive Titration in Primary Care (TASMINH4). Hypertension 2019 Jun;73(6):1231-1239. [doi: 10.1161/hypertensionaha.118.12415]
51. Lee D, Wilson K, Akehurst R, Cowie MR, Zannad F, Krum H, Eplerenone in Mild Patients Hospitalization And Survival Study in Heart Failure (EMPHASIS-HF) Study. Cost-effectiveness of eplerenone in patients with systolic heart failure and mild symptoms. Heart 2014 Nov;100(21):1681-1687 [FREE Full text] [doi: 10.1136/heartjnl-2014-305673] [Medline: 24993605]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
DBP: diastolic blood pressure
DSD: Do Something Different
Do CHANGE: Do Cardiac Health: Advanced New Generation Ecosystem
EQ-5D: EuroQol five dimensions
HF: heart failure
ICER: incremental cost-effectiveness ratio
ICT: information and communication technology

XSL•FO
RenderX

MAFEIP: Monitoring and Assessment Framework for the European Innovation Partnership on Active and Healthy Ageing
QALY: quality-adjusted life-year
RCT: randomized controlled trial
SBP: systolic blood pressure
TAU: treatment as usual
WTP: willingness to pay

Edited by G Eysenbach; submitted 09.12.19; peer-reviewed by C Vis, T Bergmo, M Alrige; comments to author 24.01.20; revised version received 20.03.20; accepted 03.06.20; published 28.07.20.

Please cite as:
Piera-Jiménez J, Winters M, Broers E, Valero-Bover D, Habibovic M, Widdershoven JWMG, Folkvord F, Lupiáñez-Villanueva F Changing the Health Behavior of Patients With Cardiovascular Disease Through an Electronic Health Intervention in Three Different Countries: Cost-Effectiveness Study in the Do Cardiac Health: Advanced New Generation Ecosystem (Do CHANGE) 2 Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17351
URL: http://www.jmir.org/2020/7/e17351/
doi:10.2196/17351
PMID:32720908

©Jordi Piera-Jiménez, Marjolein Winters, Eva Broers, Damià Valero-Bover, Mirela Habibovic, Jos W M G Widdershoven, Frans Folkvord, Francisco Lupiáñez-Villanueva. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Cyclic Value-Context Reinforcement Model of Problematic Internet Use: Empirical Validation Using a Thematic Analysis of Children's Counseling Data

Young Yim Doh[1], PhD; Bugeun Kim[2], MS; Seul Lee[3], MA; Gahgene Gweon[2], PhD

[1]Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[2]Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

[3]Advanced Institute of Convergence Technology, Suwon, Republic of Korea

Corresponding Author:
Gahgene Gweon, PhD
Graduate School of Convergence Science and Technology
Seoul National University
1 Gwanakro
Seoul
Republic of Korea
Phone: 82 318889172
Email: ggweon@snu.ac.kr

Abstract

Background: Research on problematic internet use has focused on devising diagnostic criteria or describing the factors that influence internet overuse. However, a paradigm shift is necessary in studying the phenomenon of increased internet use not just from a pathological point of view but also from a developmental point of view that considers children's behavior of adapting to a technology-oriented society.

Objective: In this paper, we propose the Cyclic Value-Context Reinforcement Model (CVCRM) to understand problematic internet use behavior. The purpose of our study was to construct a developmental process model that provides a holistic understanding of problematic internet use behavior of children and to empirically validate the proposed model by conducting a thematic analysis on actual counseling data.

Methods: To validate the CVCRM, we conducted thematic analysis using the counseling data from 312 Korean children aged 7-18 years. For the coding process, 7 master's and doctoral student researchers participated as coders, and 2 professors supervised the coding process and results.

Results: This project was funded from October 2015 to September 2019 to analyze counseling data from 312 children who participated in counseling sessions during January 2012 to May 2014. Based on the data analysis, we present the CVCRM, which integrates existing theoretical approaches and encompasses the 3 interacting aspects that induce and reinforce problematic internet use in children: psychosocial value, environmental context, and internet utility. Specifically, using counseling data, we empirically ascertained that problematic internet use behavior feeds into children's psychosocial values and environmental contexts, which in turn facilitates problematic internet use in a cyclical manner.

Conclusions: Through this empirical validation, the CVCRM can provide a theoretical framework and an integrated perspective on the developmental mechanism of problematic internet use behavior of children.

(J Med Internet Res 2020;22(7):e17996) doi:10.2196/17996

KEYWORDS

problematic internet use; children; cyclic value context reinforcement model; psychosocial value; environmental context; internet utility

XSL•FO
RenderX

Introduction

Background

The internet is pervading every aspect of daily life, even for children. Based on the European Union Kids Online Model of Children, Livingston et al [1] argued that the research agenda on children's internet usage needs to shift direction from "how children engage with the internet as a medium" to "how children engage with the world mediated by the internet." Furthermore, they raised a critical research question: "For which children under which circumstances does internet use lead to risk and why?" This question redirects our attention from the internet itself to children as active agents, to the environmental context in which children are located, and to children's developmental process of adaptation.

With the widespread use of the internet, problematic internet use is an important social issue in the global society. Compared to the term internet addiction or dependency, problematic internet use is broad enough to cover the negative consequences of the varying degrees experienced by an individual without the risk of social stigma as a pathological person [2-4]. Problematic internet use is defined as "use of internet that creates psychological, social, school, and/or work difficulties in a person's life" [5]. Until now, most research on problematic internet use has focused on devising diagnostic criteria or describing the factors that influence internet overuse. However, a paradigm shift in the study of the phenomenon of increased internet use is necessary, not just from a pathological point of view but also from a developmental point of view that considers children's behavior of adapting to a technology-oriented society.

In this study, we propose the Cyclic Value-Context Reinforcement Model (CVCRM) to explain problematic internet use behavior. The purpose of our study was to construct a developmental process model that provides a holistic understanding of problematic internet use behavior of children and to empirically validate the proposed model by conducting a thematic analysis of counseling data collected from 312 Korean children, aged 7 to 18 years. We explored the adaptive functions of the internet-mediated environment in achieving children's developmental tasks and satisfying their psychosocial values as well as their environmental needs.

Problematic Internet Use Behavior: Need for an Integrated Process Approach

Research topics on problematic internet use behavior have changed from a predominantly symptom-centered approach to a cause-centered approach and from individual factors to environmental contexts that impact internet use. Using the symptom-centered approach, researchers developed diagnostic criteria based on the investigation of symptoms of internet overuse and designed intervention methods to relieve the corresponding symptoms [6-10].

However, a symptom-centered approach is not enough to change the causal context and prevent symptom development. To address this issue, a cause-centered approach has been established to identify factors affecting problematic internet use. Early cause-centered studies examined individual internal factors, including comorbidity [11-13] or psychological vulnerabilities, such as loneliness and shyness [14-16], as the cause of problematic internet use. In addition to the individual internal factors, research on external variables and environmental contexts, which include the influences of family, school, and society, has begun recently. Family function, communication, and family conflicts are important factors that affect problematic internet use [17-20]. Additionally, people with internet-related problems tend to experience more real-life problems in social relationships and school life [21,22]. Considering these results, in order to gain a more comprehensive understanding of the phenomenon, we need to examine environmental contexts as important factors in models that examine problematic internet usage.

Although existing work has identified a correlation between specific individual and environmental factors and internet use, we need to look beyond individual relationships between specific factors and consider the overall development of problematic internet use behavior. In particular, certain individual factors transact cyclically with other factors, usage behavior, and problematic symptoms. For example, depression, loneliness, and self-control have been found to interact with other factors that affect problematic internet use [14,23-25]. Thus, comorbidities and psychological vulnerabilities are causes, as well as consequences, of problematic internet use [26,27]. Kardefelt-Winther [27] also reported that the correlations of loneliness and social anxiety with excessive online gaming were no longer present when controlling for stress. In addition, symptoms of psychopathology mediate irrational beliefs and internet gaming addiction [28]. Therefore, the relationship between internet use behavior and factors influencing it can be transactional and cyclic, rather than a one-way causal relationship.

Given the expanded research interest, from symptoms to causes of problematic internet use behaviors and from individual factors to environmental contexts, a process-centered approach integrating these factors is necessary to increase an overall understanding of problematic internet use. In other words, from a holistic perspective, we need to understand the intersection between the environmental context and personal characteristics and to reveal the mechanism underlying how these factors transact with each other and reinforce problematic internet use behavior.

Previous Theoretical Models on Problematic Internet Use Behavior

In order to construct an integrated developmental model and to guide our understanding in a qualitative thematic analysis, we selected 7 related theoretical models: (1) cognitive-behavioral model of pathological internet use [10], (2) a new model of media attendance [29], (3) model of compensatory internet use [30], (4) basic model of problematic internet use in youth [31], (5) differential-susceptibility to the media effects model [32], (6) reinforcing spirals model [33], and (7) player's value structure in digital games [34]. In this section, we compare these 7 models and discuss the insights gained in relation to problematic internet use behavior.

XSL•FO
RenderX

Davis [10] introduced the cognitive-behavioral model of pathological internet use. This model described the following 3 categories of factors contributing to behavioral symptoms: individual psychopathology, internet exposure experience, and social deficits. The core factor of this model is "maladaptive cognitions." Accordingly, the model suggests that individuals who have maladaptive cognitions have a negative view of themselves and use the internet to gain positive responses from others. Based on the cognitive-behavioral approach, this model was an early attempt to explain the etiology of pathological internet use. However, this model has limitations in that it focuses on the vulnerability of the individual by regarding internet use behavior as a pathological symptom. Another limitation of the model is that it considers the individual as a passive and socially deficient being that simply responds to situational cues.

LaRose and Eastin [29] proposed an internet use process model that combines the concepts of use and gratification with the social cognitive theory, by matching similar dimensions in both models. The following 6 expected internet outcomes were identified: novel, social, activity, monetary, self-reactive, and status. This model also added the factors of self-efficacy and self-regulation to broaden the understanding of use and gratification. The internet use process model explains internet use behavior using the expected outcomes of internet use, rather than focusing on problematic symptoms or individual vulnerabilities. In addition, the model regards users as active subjects that select media according to expected outcomes (ie, gratification). However, this approach does not consider the preceding factors that lead to the expected outcomes of internet use behavior.

Kardefelt‐Winther [30] presented a model of compensatory internet use and explained internet use behavior through motivation that stems from unmet real-life needs. This model suggests that people use the internet as a stress-coping strategy. Furthermore, the model considers the real-life context of internet usage by introducing the concept of "coping motivation," which regards reality-based difficulties as an important factor for inducing internet use. However, this model can only explain internet use behavior as a coping strategy for people who experience problems in a real-life context. Additionally, because motivation is relatively temporary and limited to a specific situation, the model is not suitable for explaining persistent internet use behavior.

As a more comprehensive model, Tam and Walter [31] suggested a basic model of problematic internet use for youth that is based on the literature from the past 10 years. They argued that problematic internet use is conceptualized, not as a unitary mental health condition, but rather as complex pathways of underlying psychological, developmental, ecological, and intrafamilial factors. This model distinguishes between regular internet use, problematic or heavy internet use, and pathological internet use or gaming addiction. Additionally, it separates the predisposing and protective factors that affect each state. However, while this model describes each of the subfactors affecting problematic internet use in detail, it has a limitation that it does not account for the interrelationships between the different factors.

In the literature on the effects of media, Valkenburg and Peter [32] introduced the differential-susceptibility to the media effects model. The model proposed that media effects are conditional and depend on the following 3 types of differential-susceptibility variables: dispositional, developmental, and social. These 3 differential-susceptibility variables affect the choice of media use. Additionally, the model distinguished the following 3 media response states: cognitive, emotional, and excitative. These 3 response states mediate the relationship between media use and media effects. Finally, this model adopted the transactional proposition, which suggests that media use, media response states, and differential-susceptibility variables are influenced by media effects. This model also provided insights regarding the relationship between media and nonmedia variables. However, because it deals with a microlevel analysis focused on individual media users, it is difficult to apply to design intervention methods of behavioral change on the level of user activity or environmental affordance.

Based on a system-theory perspective [35] and the social identity theory [36], Slater [33] elaborated the reinforcing spirals model. It recognized that media use serves as both an outcome and a predictor in many social processes and that media use in the contemporary society is a principal means to maintaining personal and social identities. One premise of this model is that the process of media selection and the effect of exposure to selected media are dynamic and ongoing. Therefore, this model considers that users tend to select certain types of media contents according to their social context, social identity, and prior attitudes. In addition, certain media content exposures have a subsequent impact on the strength and accessibility of social group identification, attitudes, and behaviors. In turn, the influence of subsequent media use continues to reinforce those associated elements of social identity, attitude, and behavior over time. The virtue of this model is highlighting the ongoing dynamic social process of media use behavior. However, social identity is not static and could vary according to the function of the environmental contexts and values of users.

Recently, in the domain of game studies and consumer research, Lin et al [34] proposed the structure of attribute-consequence-value chains of digital game players. Using soft-laddering interviews and a theoretical framework of means-end chains, they revealed game attributes that players direct their attention to, including attributes such as the connection system, popularity, graphic design, and diverse game genres. Depending on these attributes, players experience certain consequences, such as improved interactivity, cultivated logic and reflex, acquisition of an authentic experience, enhanced pleasure of senses, and utilized imagination. This model also identified the values that game players pursue, such as fun and enjoyment of life, sense of accomplishment, warm relationships, and excitement. This study provided us with insights in terms of the sequential process of users' values, media attributes, and experiential consequences.

Insights for Developing the Integrated Model of Problematic Internet Use Behavior

We gained two main insights from reviewing existing models on internet use behavior. First, early studies attempted to explain

XSL•FO
RenderX

the preconditions that could lead to internet use behavior, mainly in terms of an individual's psychosocial characteristics. However, recent studies have expanded the understanding of the interactions to include personal characteristics and environmental contexts. Therefore, when constructing a new model, it is necessary to cover all 3 basic aspects of internet usage behavior: the individual's psychosocial values and environmental contexts as well as the interaction between these two components.

Second, early studies looked at internet use behavior as a temporary problem and attempted to identify the cause or symptom of the problem by considering a simple causal relationship. However, recent studies have proposed a cyclic nature to the relationship between variables, one which creates persistent reinforcement and leads to the repetitive use of the internet. Therefore, it is necessary to examine the process that explains the types of internet utility that lead to stable and continuous problematic internet use. In addition, the model should link internet use back to the factors that exist within the context of daily life.

Based on these insights, this study proposes a hypothetical model that integrates existing theoretical approaches related to problematic internet use behavior from a developmental perspective, termed the CVCRM. The first research question of this study was: "Can we construct an integrated process model to understand problematic internet use behavior holistically?" The second research question is: "Can we identify factors and processes that interact with problematic internet use behavior in the CVCRM using children's counseling data?"

Methods

Counseling Data Collection

We conducted a thematic analysis using the counseling data provided by the National Information Society Agency in the Republic of Korea. The data comprised reports on counseling sessions from January 2012 to May 2014 with children who experienced internet-related problems. Each case consists of one initial interview and at least one follow-up counseling session, ranging from 1 to 5 sessions. During the first interview, counselors measured the severity of internet-related problems and examined the basic environmental and individual conditions of the client. The questions in the initial interview pertained to the client's home and school environments, time of the first internet use and motivation, feelings before and after internet use, how family and school respond to internet use, consequences of internet use, and the level of internet addiction. The optional follow-up counseling sessions after the initial interview were not strictly structured, and the counselors identified the clients' core problems and attempted interventions. During the entire counseling process, the counselors recorded the client's comments, as well as details about the client, his or her family members, and his or her family environment. All counseling records were archived as digital documents. We reviewed 312 cases that met the following 2 conditions: (1) case data included an initial interview report and had ≥1 follow-up session records and (2) clients aged between 7 and 18 years. This range is defined as the Korean school age and comprises entry into elementary school to graduation from high school.

Thematic Analysis

Thematic analysis [37] is a qualitative research method that enables us to discover and analyze patterns or themes in data. This methodology consists of 2 approaches: top-down theoretical approach and bottom-up inductive approach. The purpose of our study was to construct a new model by integrating existing theories and to verify the model using empirical data. Therefore, we performed a thematic analysis that incorporated both theoretical and inductive approaches. Using both top-down and bottom-up approaches iteratively allows researchers to reflect on both theory and data from various perspectives. As a result, one can gain a richer understanding when using such an iterative approach compared to using just a top-down or bottom-up approach. The detailed analysis process was as follows. First, we formulated the initial model based on the integration of previous theoretical approaches. Second, we verified our model through a thematic analysis with iteration between our model and empirical data. We used counseling data from children for this bottom-up process. Our research team read 312 cases of counseling records to identify the overall problem situation (Step 1: familiarization with the data). Then, we repeatedly read all the counseling records. Each sentence in a record was read to identify meaningful statements that indicate children's psychosocial values, environmental contexts, internet utility, and internet use behavior (Step 2: generation of the initial code). We repeated this process 9 times until the identified themes were saturated (Step 3: searching for and reviewing themes). Finally, all researchers reached a consensus on defining themes (Step 4: defining and naming themes). We performed a thematic analysis at the latent level to understand and theorize the underlying structures and processes of problematic internet use behavior. To establish credibility in the analysis, we used the researcher triangulation approach [38]. For each of the coding processes, 3-7 master's and doctoral student researchers participated as coders, and 2 professors supervised the coding process and results. The Gwet's AC1 interrater reliability coefficient was calculated in the ninth iteration for each of the following main themes: Children's Psychosocial Value (10 factors, mean 0.887), Environmental Context (13 factors, mean 0.827), and Internet Utility (6 factors, mean 0.912). In case of disagreement that arose after the coding process, all the coders discussed the case to reach a final agreement.

Results

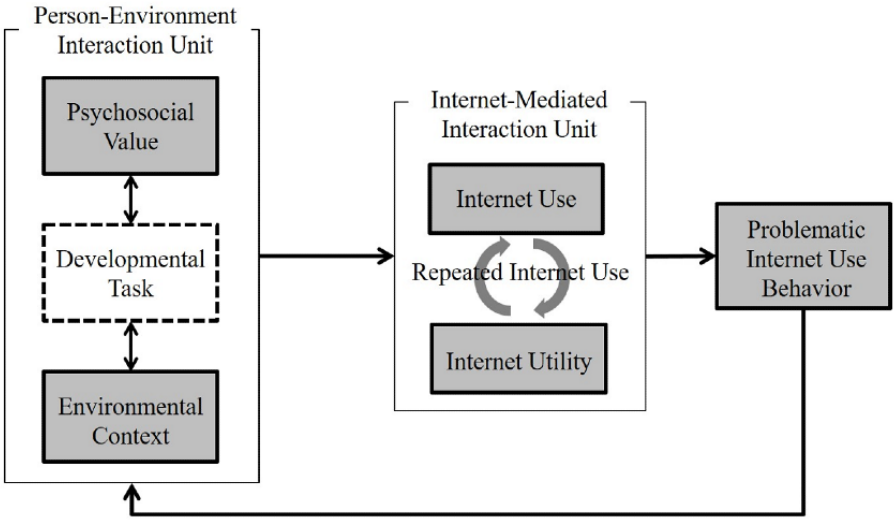
CVCRM

We formulated the CVCRM of problematic internet use behavior based on the insights drawn from previous theoretical approaches. First, we tried to cover 3 basic aspects in our model: children's psychosocial values, environmental contexts, and internet utility. Second, we incorporated repeated behavioral tendencies between internet utility and internet use in our model. Third, our model addressed the cyclic reinforcement process linkage between problematic internet use behavior and the person-environment interaction unit, which are deeply interconnected within children's daily life.

XSL•FO
RenderX

Figure 1 shows the CVCRM. The "Person-Environment Interaction Unit" is an organism in which the children's psychosocial values, developmental tasks, and environmental contexts interact. The specific conditions formed by the interaction of these elements can increase the probability of internet use. Internet use and internet utility are grouped into the "Internet-Mediated Interaction Unit." They circulate and form a repeated internet use state. Repeated internet use leads to specific problematic internet use behavior, which can give feedback to the "Person-Environment Interaction Unit" to change its state. Through this process, internet use is induced and reinforced. The states of "Person-Environment Interaction Unit" and "Internet-Mediated Interaction Unit" are not static and may vary from person to person and from time to time. We defined and described each part comprising the CVCRM as follows.

Figure 1. The Cyclic Value-Context Reinforcement Model of problematic internet use behavior.



Psychosocial Value

Psychosocial value is the intrinsic value of the child that is being met through internet use. As a result of our data analysis, 3 subthemes of psychosocial value appeared: individual, relational, and social. Psychosocial value represents relatively stable personal preferences as compared to "motivation or gratification" used in previous models. The themes and categories of the thematic analysis results are summarized in Table 1.

XSL•FO
RenderX

Table 1. Psychosocial value categories of problematic internet use behavior.

Subtheme and category	Category description	Example quotes
Individual		
To relieve stressful events	Use the internet to alleviate or eliminate psychological stress such as tension, depression, loneliness, or anxiety.	"[Using the internet] relieves my stress." "[When using the internet,] I'm not lonely anymore."
To express oneself	Evaluate the value of the internet as a means of expressing what one likes, one's emotions, and one's ideal self-image.	"[On the internet] I can express my opinions easily."
To experience a sense of accomplishment	Attach the value of the internet as a means of experiencing fulfillment, reward, and achievement.	"I feel a sense of accomplishment when I win." "I enjoy raising my ranking in a game."
To spend one's surplus time	Do not have any alternative activities other than surfing the internet and, thus, consider the internet as a convenient way to spend time.	"I get bored [when I come home], but there's nothing else I can do [to spend my time]." "I play internet games when my parents don't mind my time."
To express disobedience	Evaluate the value of the internet as a means of ignoring or actively resisting an authority's instructions.	"I played internet games more since their [parents'] actions hurt my pride." "I used the internet more heavily since my parents' regulation on internet use caused more stress."
To escape from reality	Use the internet as a place to escape from real problems that cannot be fundamentally solved.	"I feel a sense of freedom since I can escape from school." "Using the internet makes me forget about the severe conflict between my parents."
Relational/social		
To develop a sense of closeness	Evaluate the value of the internet as a means of interacting with particular subjects and maintaining familiarity.	"Internet enables me to keep in touch with my close friends who moved to other cities." "My friend and I speak the same language because of the games that both of us play."
To participate in common activities in the peer group	Evaluate the value of the internet as a way to enter peer groups, follow the current trends, and collaborate with peers.	"I need the internet to hang out with my friends." "It [the internet] makes me understand the world that my friends are in."
To be recognized as the best	Evaluate the value of the internet as a way to obtain authority, respect, and envy within a group.	"I'm proud when my friends admire me for my gaming skills." "[Using the internet] makes me happy since many people online recognize me when I upload a drawing of a game character."
To experience controllability	Evaluate the value of the internet as a place where one can control one's self, one's group, and society.	"I want to relax, have some fun, … and keep everything under control."

Developmental Task

Developmental task refers to what children must achieve at each stage of development. Thus, these tasks are fundamental psychosocial issues that can affect individuals of a certain age. Further, the developmental task at a specific age can affect and change one's environmental contexts or psychosocial values. In the model presented in Figure 1, "the developmental task" is marked with a dotted box to indicate it is a latent variable. Life's major roles are deeply interconnected with developmental tasks [39], as they refer to challenges or expectations that a culture has for individuals in different life phases [40]. The cyberspace, which is accessible via the internet, also serves as a channel for users to practice and achieve developmental tasks. For example, developing productive peer relationships is one of the important developmental tasks of school-aged children and teenagers, and it is more likely that they will have high value for social affiliation through internet use. However, the process in which these tasks interact with environmental contexts or individual value systems is not yet exposed. Therefore, the developmental task is regarded as a latent variable in this model.

Environmental Context

Environmental context includes the surrounding situation or environmental conditions that can affect the child's perception or behavior. In the counseling data, 3 subthemes of environmental context appeared: individual, family, and society. Some environmental contexts can change with time and can be influenced by the outcome of internet use. However, each of the children's experiences or perceptions of the environment, whether changed or not, may have an impact on his or her problematic internet use behavior. The themes and categories of the thematic analysis results are summarized in Table 2.

XSL•FO
RenderX

Table 2. Environmental context categories of problematic internet use behavior.

Subtheme and category	Category description	Example quotes
Individual		
Situations of physical illnesses that limit daily life	Daily life is restricted by physical disability or illness.	"After being diagnosed with tuberculosis, I began to visit the PC cafe frequently while recuperating at home."
Situations of psychological vulnerability	Mental illness requiring professional treatment (eg, ADHD[a], panic disorder) or psychological instability (eg, irritability, lethargy) is observed.	"[Mother says] client has social phobia." "I'm depressed." "I'm really worried about my school grades."
Situations of insufficient satisfaction with psychosocial needs	Complaints due to unsatisfied needs (eg, self-esteem), social exchanges (eg, social belongingness), or goal achievements (eg, realization of the ideal self) are observed.	"He doesn't know how he can relieve his disappointment when his wishes cannot be granted." "She [client] said she received art lessons but was forced to quit. She wants to pursue her career in art again."
Situations of lacking lifelong goals	A lack short-term or long-term goals or life commitments is observed.	"I don't know what I should do [for the rest of my life]."
Family		
Inadequate parenting	Caregiver's undesirable parenting style, practice, or perspective, such as excessive regulation, insufficient caregiving, or physical or verbal violence, is observed.	"[Began using the internet excessively] after my mom didn't allow me go outside." "My parents don't care [about my internet usage]." "Her [client's] mother doesn't have any faith in her."
Family communication problems	Weak family bonding or communication problems between family members are observed.	"My parents don't get along. They fight over nothing." "My parents only ask me to study. I want them to hear my thoughts about what I want to do."
Instability of the caregiver	Caregiver cannot nurture the client due to physical separation (eg, long business trips, divorce, or death).	"He [client] has a father who only visits on weekends." "He [client] lives with his father after his parents got divorced when he was 6."
Internet use of other family members	Family members show problematic internet use.	"My mother is addicted to the internet." "I cannot use the internet when my sister is at home since she is addicted to the internet."
Society		
Socially imposed stigma	Peer groups, schools, and the society insulate and isolate the client (eg, bullying, internet addiction, obesity, problem behavior).	"During middle school, he [a client] severely suffered from bullying." "My family treats me as a mental patient."
Social or school maladjustment	Difficulty in social relations or social life or problematic behavior is observed.	"I'm afraid of meeting friends at school." "He [a client] used the internet a lot because of social phobia and obsession problems."
Economic constraints	Client's activity is limited due to economic constraints (includes client's perception of economic constraints).	"The client said that he felt pity about himself, and described himself as being a beggar, several times throughout the week." "I live in poverty."
Coexisting delinquent behavior	Delinquent acts (eg, runaway, truancy, drinking, smoking, sexual activity) or crimes (eg, school violence, assault, theft, robbery, or rape) are observed.	"The client bullied other students and often questioned the purpose of studying." "His [client's] teacher said she often had to search for him because he went to the PC café instead of attending school."
Unstructured daily activities	Client's daily life is unstructured (no daily structure at home after expulsion or no regular activity plan for after-school hours).	"Since quitting the job at a restaurant, I spent time playing computer games [without any future plans]."

[a]ADHD: attention deficit hyperactivity disorder.

Internet Use

Internet use, or more specifically "repeated internet use," refers to the state of repeated engagement in internet use activity while obtaining internet utilities.

Internet Utility

Internet utility satisfies the psychosocial value of a child, including the utility of content or relationships accessible via the internet. Internet utility can mediate the influence of the "Person-Environment Interaction Unit" on problematic internet use behavior to induce repeated internet use. In the counseling data, 3 subthemes of internet utility appeared: individual, relational, and social. The themes and categories of the thematic analysis results are summarized in Table 3.

XSL•FO
RenderX

Table 3. Internet utility categories of problematic internet use behavior.

Subtheme and category	Category description	Example quotes
Individual		
To enjoy	Derive pleasure and fun from using the internet.	"Using the internet is the only pleasure in my life." "The internet is thrilling."
To engage in goal-oriented learning	Establish goals to master specific skills or subjects in real life in the short or long term and gain learning effects over the internet.	"[When I'm using the internet,] I can code programs, which is both a hobby and a way of preparing for my future career." "[On the internet,] there're a lot of things that I can learn from."
To surf the internet	Browse and collect fragmentary information through the internet.	"It [the internet] helps with my homework." "I used the internet to watch a show that a singer [who the client is a fan of] starred in, listen to his songs, and communicate with other fans."
To earn money	Acquire things that can make money or exchange money through the internet.	"I can earn a lot of money [through online gambling]." "I do online gambling because my allowances are not enough."
To satisfy one's sexual curiosity	Satisfy one's sexual curiosity through the internet.	"I sometimes watch pornography online."
Relational/social		
To be active in an online community	Communicate with new people through an online community or social networking service (including positive, neutral, and negative communication).	"It is fun to communicate with others anonymously." "Arguing with others online is fun."

Cyclic Reinforcing Process

Children's psychosocial values, environmental contexts, and internet utilities interact to induce and reinforce internet use. Afterwards, problematic internet use is fed back into individual value systems and environmental contexts, resulting in repeated and strengthened internet use. We illustrate 5 representative counseling cases in high-risk groups that vividly reveal the cyclic reinforcing process in Table 4.

XSL•FO
RenderX

Table 4. Representative counseling cases that show the cyclic reinforcing process.

Case ID	Counseling content	Corresponding cyclic pattern
Case 57	"…I experienced school violence in middle school and became obsessed with games. […] I feel like I'm out of reality and feel liberated. I also feel like I am being recognized by people. […] The interpersonal relationship was so comfortable that I don't feel the need to make friends offline. More recognition on the internet makes it more difficult to establish relationships in the real world…"	School violence (EC[a]) > escape from reality, social recognition (PV[b]) > enjoy relationship (IU[c]) > repeated and obsessive use (PIUB[d]) > weakening of offline relationships (EC) > cycle repetition
Case 155	"…Since I'm in the third year of high school, I have to study a lot, but I keep thinking about games. Eventually, I play games in the PC room with the excuse that I am going to the reading room on weekends. […] I'm not bored if I'm playing a game. I am anxious if I don't play games. I have to study, but I'm depressed and worried because I play a lot of games. […] I'm worried that, like my brother, I'm going to have a serious conflict with my parents…"	College entrance exam stress (EC) > relieve stress (PV) > enjoy fun (IU) > repeated and obsessive use, feeling of depression and worry (PIUB) > expectation of conflict with parents (EC) > cycle repetition
Case 295	"…Bored at home, I started the internet. […] I can use the internet to relieve stress and to kill time. I can meet new people in virtual space. In particular, I communicate well with people who play the same games. […] I tried to quit the game, but I was forced to play it again because of my clan friends…"	Unstructured daily life (EC) > relieve stress and spend surplus time (PV) > to be active in online community (IU) > excessive use (PIUB) > cycle repetition
Case 324	"…When I was in the third year of elementary school, my mom took me to a PC room. The internet is fun, stress-free, and pleasant. After that, I repeatedly went to the PC room to play games. […] I often went to PC rooms, and my mom's interference with games increased. Grandpa and grandma's nagging got worse. I go back to the PC room to avoid it…"	Internet use of mother (EC) > relieve stress (PV) > enjoy fun (IU) > repeated visits to PC room (PIUB) > conflict with family members (EC) > cycle repetition
Case 559	"…Life is boring. When I'm immersed in the internet, I can be happy and have fun in my life. […] I spend a lot of time playing games. […] I went to the Rescue Internet Camp last year. At first it seemed to get better, but after a while it was back…"	Lack of a lifelong goal and insufficient satisfaction of psychosocial needs (EC) > to express oneself (PV) > enjoy and feel happy (IU) > repeated use (PIUB) > intervention in Rescue Internet Camp is useless (EC) > cycle repetition

[a]EC: environmental context.

[b]PV: psychosocial value.

[c]IU: internet utility.

[d]PIUB: problematic internet use behavior.

Funding

This project was funded from October 2015 to September 2019 to analyze counseling data from 312 children who participated in counseling sessions during January 2012 to May 2014.

Discussion

Implications

Our proposal of the CVCRM of internet use behavior and our thematic analysis of the model using children's counseling data have several implications. First, the CVCRM expanded the understanding of problematic internet use behavior by introducing the concepts of children's psychosocial values and environmental contexts. In addition, the developmental task was included as a latent variable to explain the mechanisms of the developmental process of internet use behavior. In accordance with Lin et al [34], we adopted the concept of value based on the social adaptation theory by Kahle [41] and the suggested list of values [42]. Kahle listed 9 values (ie, self-respect, security, warm relationships with others, sense of accomplishment, self-fulfillment, sense of belonging, sense of being well-respected, fun and enjoyment of life, and excitement) and regarded these values as closely related to life's major roles [42]. In addition, positive adaptation in children can be defined by achieving age-salient developmental tasks, and these tasks reflect the expectations and standards that parents, teachers, and society set for them [43]. Therefore, we need to consider children's psychosocial values, developmental tasks, and environmental contexts in order to understand children's problematic internet use behavior. In the counseling data, we found 3 subdimensions in the environmental context (ie, individual, family, and society). The structure of these 3 hidden dimensions is similar to the bioecological model of human development [44]. Bronfenbrenner and Morris [44] explained that both risks and resources for positive adaptation and development stem from factors situated within individuals (genetic and hormonal systems, personality, and cognition) as well as in the proximal (family and school) and distal (societal, cultural, and institutional) contexts in which their life is embedded [43].

Second, we proposed an extended understanding that internet use behavior is not a short-term problem state, but rather a persistently evolving and cyclically reinforcing process. At the surface level, the purpose of repeated internet use is obtaining internet utilities in the short term. However, if we widen our viewpoint to include the person-environment interaction unit, we can understand that internet utility originated from the interaction between children's psychosocial values,

XSL•FO
RenderX

developmental tasks, and environmental context needs. These multiple dimensions are relatively long-lasting characteristics, and the dimensions cyclically reinforce each other. Therefore, our study can contribute to the exploration of the cyclical nature of the relationship between variables that produce persistent reinforcement mechanisms of the repetitive use or re-use of the internet.

Third, the CVCRM can be applied to inform preventive interventions for problematic internet use. The current interventions tend to focus on directly changing personal characteristics or symptoms. However, personal characteristics, such as personality, self-regulation, self-efficacy, and cognition, are difficult to intervene and affect. Conversely, it is relatively easy to induce behavioral change by modifying environmental contexts. For example, in one case, a girl was constantly home alone because her parents both worked full-time jobs. She felt sad when her mother told her to "Take care of the problem yourself," in response to trying to talk to her about a problem she faced. She used the internet for entertainment when she was home alone. Her internet usage duration increased, and she did not consider reducing it. Within a symptom-centered framework, she is at high risk of developing internet addiction. Here, therapists will attempt to intervene using a cognitive-behavioral approach, in which they will improve her self-regulation and plan her internet use time. However, from the value-context perspective, it is possible to change her parents' attitudes to meet her needs of love and belongingness or to provide her with a structured daily life by offering opportunities to engage in cultural or leisure activities. Although the individual's internal factors, such as self-control or depression, are abstract and difficult to change with short-term intervention, environmental contexts and value-seeking activities could change through practical interventions in a relatively short period.

Limitations

There are some limitations in this study. The first pertains to the structure of the counseling data. Specifically, several factors may not have been discovered in the initial interview since the questions were semistructured. Furthermore, home-visit counseling sessions were loosely structured, so that counselors could freely talk and explore the client's situation. The counseling record was not an exact transcript of the counseling session and was recorded as a summary based on the counselors' comprehensive observations. Therefore, the counselors' personal opinion may have influenced the records. Second, it is difficult to investigate cognitive factors because of the characteristics of children. Children's self-consciousness is not yet fully developed. Without verbal statements, observations alone are not enough to identify abstract cognitive factors. Therefore, we did not include cognitive factors in the proposed model. However, cognitive factors, such as maladaptive cognitions [45] and desire thinking [46], are considered important predictors of internet use problems and may mediate the effects of values and contexts. Future research needs to explore the relationship between the factors proposed by the CVCRM and cognitive factors.

Future Research

For future studies, we propose several directions to advance research in this domain. First, it is important to validate the present model using a quantitative approach given that our approach was an empirical validation. Furthermore, it is necessary to explore the interaction between environmental contexts and psychosocial values, as proposed by this study. Therefore, while we integrated these two concepts into the concept of the "person-environment interaction unit," the detailed interactive processes taking place inside the unit are yet to be investigated. Second, there is a need to compare the changes in how the environment is perceived, psychosocial values, and internet utilities, based on developmental periods. By expanding the present study to the entire lifecycle, which includes young adults, middle-aged adults, and the elderly, we can explore the interaction between differential developmental tasks, environmental contexts, and value systems throughout the lifespan. Third, research is needed to compare the structures of the environmental contexts and psychological values surrounding internet use across various cultures. Values are not culturally neutral. For example, community connectedness is a strong value in the Korean society. Thus, social recognition and acceptance are important values in Korea for children to master, compared to other cultures where the value of independence may be encouraged as a virtue of adolescence. Lastly, developing specific counseling techniques and how such techniques can reinforce problematic internet use behavior are important and interesting topics to pursue. An indepth development of counseling strategies based on the CVCRM and the effects of such strategies would be a great direction for future work.

Conclusions

We proposed the CVCRM of problematic internet use behavior from the developmental perspective using children's counseling data. By performing a thematic analysis at the latent level, we could extend the current understanding of the underlying structure of internet use behavior. The core argument of our model is that children's psychosocial values, environmental contexts, and internet utility interact to induce and reinforce problematic internet use. As a result, problematic internet use is reinforced back into the individual value system and environmental contexts, resulting in further strengthened internet use. Since the developmental task during adolescence can affect and change one's environmental contexts or psychosocial values, it was included as a latent variable.

Human behavior is shaped by various factors, including those that underlie the environment. Once we understand the mechanisms by which children start and continue to use the internet, we can gain a deeper understanding of internet use behavior. Rather than attempting to change children's internal traits or symptoms, it would be clinically useful to decipher and obtain the values that children seek and to create an environment that encourages them to change their behaviors on their own. From the perspective of developmental psychology, the value that children acquire by using the internet is related to a developmental task and is therefore very important in their growth process [40]. To help children derive value from the

XSL•FO
RenderX

internet without falling into problematic use, a value-context approach may be useful in helping to create an environment in which children can recognize their own values regarding the internet and yet ultimately control and regulate their internet use behavior.

Acknowledgments

This work was supported by the National Research Foundation of Korea: [Grant Number 2015M3C7A1065859].

Conflicts of Interest

None declared.

References

1. Livingstone S, Mascheroni G, Staksrud E. European research on children's internet use: Assessing the past and anticipating the future. New Media & Society 2017 Jan 10;20(3):1103-1122. [doi: 10.1177/1461444816685930]
2. Caplan SE. Problematic Internet use and psychosocial well-being: development of a theory-based cognitive–behavioral measurement instrument. Computers in Human Behavior 2002 Sep;18(5):553-575. [doi: 10.1016/s0747-5632(02)00004-3]
3. Yellowlees PM, Marks S. Problematic Internet use or Internet addiction? Computers in Human Behavior 2007 May;23(3):1447-1453. [doi: 10.1016/j.chb.2005.05.004]
4. Lee EWJ, Ho SS, Lwin MO. Explicating problematic social network sites use: A review of concepts, theoretical frameworks, and future directions for communication theorizing. New Media & Society 2016 Oct 06;19(2):308-326. [doi: 10.1177/1461444816671891]
5. Beard KW, Wolf EM. Modification in the proposed diagnostic criteria for Internet addiction. Cyberpsychol Behav 2001 Jun;4(3):377-383. [doi: 10.1089/109493101300210286] [Medline: 11710263]
6. Griffiths M. Technological addictions. Clinical Psychology Forum 1995;76:14-19.
7. Young KS. Psychology of computer use: XL. Addictive use of the Internet: a case that breaks the stereotype. Psychol Rep 1996 Dec;79(3 Pt 1):899-902. [doi: 10.2466/pr0.1996.79.3.899] [Medline: 8969098]
8. Scherer K. College life on-line: Healthy and unhealthy Internet use. Journal of College Student Development 1997;38(6):655-665.
9. Shapira NA, Goldsmith TD, Keck PE, Khosla UM, McElroy SL. Psychiatric features of individuals with problematic internet use. J Affect Disord 2000;57(1-3):267-272. [doi: 10.1016/s0165-0327(99)00107-x] [Medline: 10708842]
10. Davis R. A cognitive-behavioral model of pathological Internet use. Computers in Human Behavior 2001 Mar;17(2):187-195. [doi: 10.1016/S0747-5632(00)00041-8]
11. Kim K, Ryu E, Chon M, Yeun E, Choi S, Seo J, et al. Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: a questionnaire survey. Int J Nurs Stud 2006 Feb;43(2):185-192. [doi: 10.1016/j.ijnurstu.2005.02.005] [Medline: 16427966]
12. Morrison CM, Gore H. The relationship between excessive Internet use and depression: a questionnaire-based study of 1,319 young people and adults. Psychopathology 2010;43(2):121-126. [doi: 10.1159/000277001] [Medline: 20110764]
13. Yoo HJ, Cho SC, Ha J, Yune SK, Kim SJ, Hwang J, et al. Attention deficit hyperactivity symptoms and internet addiction. Psychiatry Clin Neurosci 2004 Oct;58(5):487-494 [FREE Full text] [doi: 10.1111/j.1440-1819.2004.01290.x] [Medline: 15482579]
14. Caplan SE. Preference for online social interaction: a theory of problematic internet use and psychosocial well-being. Communication Research 2016 Jun 29;30(6):625-648. [doi: 10.1177/0093650203257842]
15. Caplan SE. A social skill account of problematic internet use. Journal of Communication 2005;55(4):721-736. [doi: 10.1111/j.1460-2466.2005.tb03019.x]
16. Kim J, LaRose R, Peng W. Loneliness as the cause and the effect of problematic Internet use: the relationship between Internet use and psychological well-being. Cyberpsychol Behav 2009 Aug;12(4):451-455. [doi: 10.1089/cpb.2008.0327] [Medline: 19514821]
17. Bernardi S, Pallanti S. Internet addiction: a descriptive clinical study focusing on comorbidities and dissociative symptoms. Compr Psychiatry 2009;50(6):510-516. [doi: 10.1016/j.comppsych.2008.11.011] [Medline: 19840588]
18. Park SK, Kim JY, Cho CB. Prevalence of Internet addiction and correlations with family factors among South Korean adolescents. Adolescence 2008;43(172):895-909. [Medline: 19149152]
19. Li W, Garland EL, Howard MO. Family factors in Internet addiction among Chinese youth: A review of English- and Chinese-language studies. Computers in Human Behavior 2014 Feb;31:393-411. [doi: 10.1016/j.chb.2013.11.004]
20. Yen J, Yen C, Chen C, Chen S, Ko C. Family factors of internet addiction and substance use experience in Taiwanese adolescents. Cyberpsychol Behav 2007 Jun;10(3):323-329. [doi: 10.1089/cpb.2006.9948] [Medline: 17594255]
21. Yao B, Han W, Zeng L, Guo X. Freshman year mental health symptoms and level of adaptation as predictors of Internet addiction: a retrospective nested case-control study of male Chinese college students. Psychiatry Res 2013 Dec 15;210(2):541-547. [doi: 10.1016/j.psychres.2013.07.023] [Medline: 23896352]

XSL•FO
RenderX

22. Li D, Li X, Wang Y, Zhao L, Bao Z, Wen F. School connectedness and problematic internet use in adolescents: a moderated mediation model of deviant peer affiliation and self-control. J Abnorm Child Psychol 2013 Nov;41(8):1231-1242. [doi: 10.1007/s10802-013-9761-9] [Medline: 23695186]
23. LaRose R, Lin CA, Eastin MS. Unregulated internet usage: addiction, habit, or deficient self-regulation? Media Psychology 2003 Aug;5(3):225-253. [doi: 10.1207/s1532785xmep0503_01]
24. Özdemir Y, Kuzucu Y, Ak A. Depression, loneliness and Internet addiction: How important is low self-control? Computers in Human Behavior 2014 May;34:284-290. [doi: 10.1016/j.chb.2014.02.009]
25. Ko CH, Yen JY, Yen CF, Chen CS, Chen CC. The association between Internet addiction and psychiatric disorder: a review of the literature. Eur Psychiatry 2012 Jan;27(1):1-8. [doi: 10.1016/j.eurpsy.2010.04.011] [Medline: 22153731]
26. Lemmens JS, Valkenburg PM, Peter J. Psychosocial causes and consequences of pathological gaming. Computers in Human Behavior 2011 Jan;27(1):144-152. [doi: 10.1016/j.chb.2010.07.015]
27. Kardefelt-Winther D. Problematizing excessive online gaming and its psychological predictors. Computers in Human Behavior 2014 Feb;31:118-122. [doi: 10.1016/j.chb.2013.10.017]
28. Vukosavljevic-Gvozden T, Filipovic S, Opacic G. The mediating role of symptoms of psychopathology between irrational beliefs and internet gaming addiction. J Rat-Emo Cognitive-Behav Ther 2015 Jul 17;33(4):387-405. [doi: 10.1007/s10942-015-0218-7]
29. LaRose R, Eastin MS. A social cognitive theory of internet uses and gratifications: toward a new model of media attendance. Journal of Broadcasting & Electronic Media 2004 Oct;48(3):358-377. [doi: 10.1207/s15506878jobem4803_2]
30. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. Computers in Human Behavior 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
31. Tam P, Walter G. Problematic internet use in childhood and youth: evolution of a 21st century affliction. Australas Psychiatry 2013 Dec;21(6):533-536. [doi: 10.1177/1039856213509911] [Medline: 24151068]
32. Valkenburg PM, Peter J. The differential susceptibility to media effects model. Journal of Communication 2013 Mar 07;63(2):221-243. [doi: 10.1111/jcom.12024]
33. Slater MD. Reinforcing spirals model: conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. Media Psychol 2015 Jul 01;18(3):370-395 [FREE Full text] [doi: 10.1080/15213269.2014.897236] [Medline: 26366124]
34. Lin Y, Lin H, Yang Y. Players value structure in digital games. Games and Culture 2015 Apr 22;12(1):72-99. [doi: 10.1177/1555412015581710]
35. Von Bertalanffy L. General systems theory: Foundations, development, applications. New York: George Braziller; 1968.
36. Tajfel H, Turner J. The social identity theory of intergroup behavior. In: Worchel S, Austin WG. editors. Psychology of intergroup relations. Chicago: Nelson-Hall; 1986:7-24.
37. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. International Journal of Qualitative Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
39. Havighurst R. Developmental tasks and education. New York: David McKay Company; 1972.
40. Subrahmanyam K, Smahel D. Connecting online behavior to adolescent development: A theoretical framework. In: Digital youth. New york: Springer; 2010:27-39.
41. Kahle LR. Social values and social change. New York: Praeger; 1983.
42. Kahle LR, Beatty SE, Homer P. Alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS). J CONSUM RES 1986 Dec;13(3):405. [doi: 10.1086/209079]
43. Motti-Stefanidi F. Identity development in the context of the riskresilience framework. In: McLean KC, Syed M. editors. The Oxford handbook of identity development. Oxford: Oxford University Press; 2015:472-489.
44. Bronfenbrenner U, Morris P. The bioecological model of human development. In: Lerner RM. editor. Handbook of child psychology. Vol.1: Theoretical models of human development. New Jersey: John Wiley & Sons; 2006:793-828.
45. Forrest CJ, King DL, Delfabbro PH. Maladaptive cognitions predict changes in problematic gaming in highly-engaged adults: A 12-month longitudinal study. Addict Behav 2017 Feb;65:125-130. [doi: 10.1016/j.addbeh.2016.10.013] [Medline: 27816037]
46. Spada MM, Caselli G, Slaifer M, Nikčević AV, Sassaroli S. Desire thinking as a predictor of problematic internet use. Social Science Computer Review 2013 Nov 24;32(4):474-483. [doi: 10.1177/0894439313511318]

Abbreviations

CVCRM: Cyclic Value-Context Reinforcement Model
EC: environmental context
IU: internet utility
PIUB: problematic internet use behavior
PV: psychosocial value

XSL•FO
RenderX

Edited by G Eysenbach; submitted 28.01.20; peer-reviewed by A Roberts, H Thabrew; comments to author 21.04.20; revised version received 04.05.20; accepted 14.05.20; published 14.07.20.

Please cite as:
Doh YY, Kim B, Lee S, Gweon G
The Cyclic Value-Context Reinforcement Model of Problematic Internet Use: Empirical Validation Using a Thematic Analysis of Children's Counseling Data
J Med Internet Res 2020;22(7):e17996
URL: https://www.jmir.org/2020/7/e17996
doi:10.2196/17996
PMID:32460233

©Young Yim Doh, Bugeun Kim, Seul Lee, Gahgene Gweon. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media

Shuqing Gao[1*], MA; Lingnan He[2,3*], PhD; Yue Chen[2], BA; Dan Li[4], BA; Kaisheng Lai[4], PhD

[1]Faculty of Psychology, Beijing Normal University, Beijing, China

[2]School of Communication and Design, Sun Yat-Sen University, Guangzhou, China

[3]Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion, Guangzhou, China

[4]School of Journalism and Communication, Jinan University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Kaisheng Lai, PhD
School of Journalism and Communication
Jinan University
601 Whampoa Ave W
Guangzhou,
China
Phone: 86 020 38374980
Email: kaishenglai@126.com

Abstract

Background: High-quality medical resources are in high demand worldwide, and the application of artificial intelligence (AI) in medical care may help alleviate the crisis related to this shortage. The development of the medical AI industry depends to a certain extent on whether industry experts have a comprehensive understanding of the public's views on medical AI. Currently, the opinions of the general public on this matter remain unclear.

Objective: The purpose of this study is to explore the public perception of AI in medical care through a content analysis of social media data, including specific topics that the public is concerned about; public attitudes toward AI in medical care and the reasons for them; and public opinion on whether AI can replace human doctors.

Methods: Through an application programming interface, we collected a data set from the Sina Weibo platform comprising more than 16 million users throughout China by crawling all public posts from January to December 2017. Based on this data set, we identified 2315 posts related to AI in medical care and classified them through content analysis.

Results: Among the 2315 identified posts, we found three types of AI topics discussed on the platform: (1) technology and application (n=987, 42.63%), (2) industry development (n=706, 30.50%), and (3) impact on society (n=622, 26.87%). Out of 956 posts where public attitudes were expressed, 59.4% (n=568), 34.4% (n=329), and 6.2% (n=59) of the posts expressed positive, neutral, and negative attitudes, respectively. The immaturity of AI technology (27/59, 46%) and a distrust of related companies (n=15, 25%) were the two main reasons for the negative attitudes. Across 200 posts that mentioned public attitudes toward replacing human doctors with AI, 47.5% (n=95) and 32.5% (n=65) of the posts expressed that AI would completely or partially replace human doctors, respectively. In comparison, 20.0% (n=40) of the posts expressed that AI would not replace human doctors.

Conclusions: Our findings indicate that people are most concerned about AI technology and applications. Generally, the majority of people held positive attitudes and believed that AI doctors would completely or partially replace human ones. Compared with previous studies on medical doctors, the general public has a more positive attitude toward medical AI. Lack of trust in AI and the absence of the humanistic care factor are essential reasons why some people still have a negative attitude toward medical AI. We suggest that practitioners may need to pay more attention to promoting the credibility of technology companies and meeting patients' emotional needs instead of focusing merely on technical issues.

(J Med Internet Res 2020;22(7):e16649) doi:10.2196/16649

KEYWORDS

artificial intelligence; public perception; social media; content analysis; medical care

XSL•FO
RenderX

Introduction

Background

High-quality medical resources are in great demand worldwide. The World Health Organization estimates that there is a global shortage of about 4.3 million doctors and nurses, a problem that poses a more serious threat in developing countries [1]. The application of artificial intelligence (AI), a field of computer science that aims to mimic human cognitive functions with computer algorithms [2,3], in medical care can reduce the burden on doctors and nurses, improve patient care, and alleviate the human resource crisis in health care [4]. In recent years, AI has made breakthroughs and has begun to be widely used in various areas of daily life, such as autonomous driving [5], weather forecasting [6], and health care practices [7].

With the increasing availability of medical data and the development of algorithms, AI has become an evolving trend in the medical field [8,9]. In medical diagnosis, AI can improve accuracy through image recognition technology and semantic analysis. In 2017, a Stanford University study revealed that AI beat human doctors in skin cancer diagnosis by attaining more than 90% diagnostic accuracy [10]. In the context of medical decision making and treatment, with the help of increasingly sophisticated algorithms and delicate instruments, AI technology can provide high-quality therapeutic options and perform specific operations. For example, the first robotic hand, developed by the research team of the American Children's National Health System, could handle soft tissue automatically [11]. In health management, researchers from the Las Vegas Department of Health have applied machine learning to Twitter data and developed an AI system that can help prevent foodborne illnesses [12]. In preventive health care, increasingly rich medical and health data provide opportunities for more accurate health monitoring and disease prevention. Recently, through a deep learning approach, researchers from DeepMind developed a model that was able to predict acute kidney injuries 48 hours in advance [13].

Although medical AI has gone through many technical advances, its usage and social influence (eg, whether AI doctors would replace human doctors) have attracted the attention of the government, media, and society. Existing research mainly focuses on revealing the attitudes and views of AI experts or medical professionals toward medical AI. Previous studies indicated that the overwhelming majority of technical experts in the field of biomedical informatics believe that AI will revolutionize many medical fields [14]. Other experts even predict that human doctors are at risk of being replaced by AI as it gets closer to general human intelligence [15]. Medical professionals, including doctors and medical students, admit that AI can outperform human doctors in some areas. However, they also considered the potential of AI to be limited and are not worried about being replaced by AI [16-18]. However, it is not known how medical AI is perceived by the general public, and few studies have empirically explored the public view on medical AI.

Compared to AI experts and medical professionals, the general public has less subject knowledge, but they are the ultimate users of medical AI. The public's attitude toward AI in health care is crucial. Public perception and attitudes toward AI in medical care may affect the development progress of AI products (eg, collecting sufficient data from the public for machine learning) in the early stage. In contrast, public acceptance of AI products may exert an influence in the middle and late stages. Therefore, we suggest that the healthy development of the medical AI industry depends, to a certain extent, on whether practitioners in related fields comprehensively understand the public's views on AI in medical care.

The development of information science and the popularization of social media have made it possible to study the psychology and behavior of a large population based on the amount of behavioral data available from platforms such as Twitter and Facebook [19]. Compared with the questionnaire survey and experimental research, social media analysis in general has higher ecological validity and can reflect the opinions of large groups of people more objectively [20]. In recent years, researchers have started to use social media to study public awareness about certain illnesses, such as lung cancer [21] and cardiovascular disease [22]. It has also been used to explore public perceptions about health care (eg, cardiopulmonary resuscitation [23], vaccines [24]), and the application of new technologies in disease treatment (eg, virtual reality [25]).

Sina Weibo (Sina Corp), which is similar to Twitter, is one of the most popular social media platforms in China and boasts approximately 500 million daily active users from around China [26,27]. A growing body of literature has identified Sina Weibo as a useful platform for public health research, and its data have been used for studies on medical topics such as cancer misinformation [28], depression-related discourses [29], and organ donation awareness [30]. With the rapid development of medical AI, it has become an increasingly important topic of concern for the media and public. Over time, more posts and discussions about people's feelings, opinions, and concerns regarding medical AI have emerged on social media platforms, including Sina Weibo. This provides an excellent opportunity to study the general public's perception of AI in medical care based on social media posts.

Objectives

To the best of our knowledge, no study has comprehensively examined the views of the general public on AI in medical care. To address this gap, we explored the public perception of medical AI on social media through a content analysis of a large amount of data generated from the largest social media platform, Sina Weibo. Specifically, we focused on the following three questions: (1) What are the main medical AI–related topics about which the public is concerned? (2) What are the attitudes of the public toward AI in medical care? (3) Do people believe that medical AI can replace human doctors?

Methods

The data used in this study were collected from Sina Weibo through its application programming interface. First, we established a data set comprising 16 million Weibo users

XSL•FO
RenderX

throughout China, crawled these users' public posts from January 1 to December 31, 2017, and collected their public registration information (including gender, age, and geographic location). Second, we identified 4515 posts from this data set that contained at least one AI-related keyword (eg, "AI" or "artificial intelligence") and one medicine-related keyword (eg, "medicine," "treatment," or "health"). Third, we cleaned these posts further by manually inspecting and excluding 2000 invalid posts. These invalid texts were mainly caused by the keyword "AI," as it is usually used as the phonetic transcription of "爱," which means "love" in Chinese. Moreover, we also excluded invalid posts in which "AI" referred to Adobe Illustrator. Finally, we obtained 2315 posts for further analysis.

Based on these 2315 posts, we first outlined the general characteristics of public attention toward AI in medical care. Specifically, we examined the dynamic fluctuations of public attention over time and revealed the user profiles (such as gender, age, and region) of people who were concerned about medical AI through chi-square tests at the user account level. We then focused on public interest in and attitudes toward medical AI through content analysis. Referring to the coding methods used in previous research [30], we used a direct content analysis approach to code these posts, including the following: (1) topics of public concern regarding AI in medical care, (2) public attitudes toward AI in medical care and the reasons for them, and (3) public opinion on whether AI can replace human doctors. For example, for the thematic content analysis, two researchers screened the posts while one of them proposed a codebook for topic categories. A different researcher then evaluated the categories and discussed the framework with the first two researchers. Thereafter, the two researchers discussed and resolved the disparities in coding and reached a consensus on the classification of all topics. Subsequently, they coded approximately 10% (n=236) of posts to verify intercoder reliability. The κ score for the thematic content analysis was 0.82, which was acceptable. Finally, they independently completed the coding of the remaining posts. In this process, posts that contained more than one topic category were categorized as the more relevant of the options. The coding process of public attitudes toward AI in medical care and public opinion on whether AI can replace human doctors was consistent with the processing of the topics of public concern regarding AI. Furthermore, their intercoder reliability was also acceptable (κ scores were 0.80 for public attitudes toward AI in medical care and 0.92 for public opinion on whether AI can replace human doctors).

Results

Public Attention Toward AI in Medical Care

To explore the characteristics of public attention toward medical AI, we first examined the dynamic fluctuation in people's discussions of AI in medical care over time. Figure 1 shows the daily counts of all posts related to medical AI from January to December 2017. Public attention toward AI in medical care demonstrated obvious event-driven characteristics. Whenever a widely reported event related to AI occurred, public attention surged accordingly. We found that there were three main categories of events that could drive public attention toward AI in medical care. The first category comprised messages about improvements and breakthroughs in AI technology, such as when AI beat human doctors at diagnosing skin cancer [31]. The second category comprised information or regulations related to AI released by the government or official institutions, such as when AI was written into a government work report and the 2017 Global Cloud Computing Conference was held. The third category comprised social entertainment events related to AI, such as AlphaGo AI defeating Ke Jie, the world's number one Go player.

In addition to the temporal characteristics, we explored users' profile characteristics to identify those who were most interested in medical AI. However, since not every user had publicly disclosed their account information, we limited our analysis to those who had. Of these 1764 accounts, 1294 (73.36%) were attributed to male users and 470 (26.42%) were female users (Table 1). Males paid more attention to AI-related medical topics than did females, even after controlling for the gender ratio of the 16 million users in the data set (χ^2_1=948.5; P<.001). In terms of age, users under 30 years old paid less attention to AI medical topics than did users over 30 years old, and this was significantly lower than the proportion of young to old users in the data set (χ^2_1=491.2; P<.001). Moreover, users from the regions (provinces) with above-average incomes showed more interest in medical AI than did users from regions with below-average incomes, at a level significantly higher than the proportion of such users in the data set (χ^2_1=90.7; P<.001).

XSL•FO
RenderX

Figure 1. Number of posts about artificial intelligence in medical care from January to December 2017.

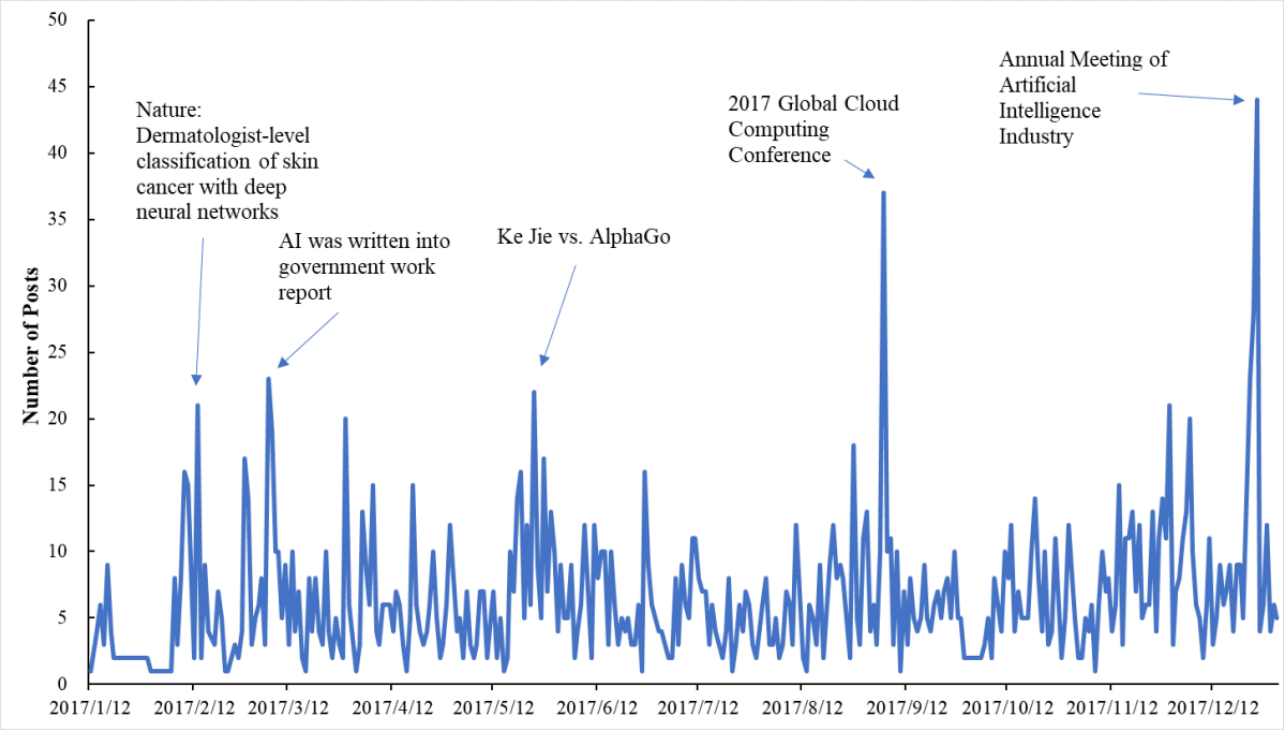


Table 1. Overview of the general demographics of the users who were interested in medical artificial intelligence.[a]

General demographics	Users in analysis, n (%)	Users of data set, n (%)
Gender		
Male	1294 (73.36)	6,267,548 (37.80)
Female	470 (26.42)	10,311,693 (62.20)
Age (years)		
≤30	231 (47.53)	6,297,207 (84.22)
>30	255 (52.47)	1,180,058 (15.78)
Region		
Above-average income	1128 (74.60)	7,858,752 (62.76)
Below-average income	384 (25.40)	4,662,573 (37.23)

[a]Our statistics were limited to those users who disclosed their demographic information on the corresponding indicators publicly, so the actual number of users on different demographic indicators may be different.

Thematic Content Analysis

We employed thematic content analysis to explore the topics of public discussion on medical AI. The results showed that such contents fall into three major categories, including 10 subcategories (Table 2).

The first major category, “Technology and application” (n=987, 42.63%), focuses mainly on what AI is and which medical fields it can be applied to; people discussed technical issues (n=109, 4.71%) related to AI technology, as well as its general uses (n=210, 9.07%) and specific uses (n=668, 28.85%). With regard to the specific uses of AI, its technological application in the medical field was the most popular topic; people showed the most interest in the application of AI in diagnosis (n=436, 18.83%) and treatment (n=146, 6.31%), and some interest in prevention (n=52, 2.25%) and recovery (n=34, 1.47%).

The second major category, “Industry development” (n=706, 30.50%), is mainly concerned with the development and trends of medical AI and related industries. People were particularly concerned with the development of related companies (n=331, 14.30%) such as Baidu, Alibaba, and Tencent. They also discussed investment and market (n=186, 8.03%) and industry expectations (n=142, 6.13%), in which they estimated the financial performance and investment possibilities in such an industry. However, few mentioned policy and law (n=47, 2.03%) in their posts. Given that medical AI is new, there are limited regulations and laws related to it for people to discuss.

XSL•FO
RenderX

Table 2. Topics of public concern regarding artificial intelligence in medical care (N=2315).

Topics	Posts, n (%)	Definition	Example post
Technology and application	987 (42.63)	What is AI[a] and in which medical fields can it be applied?	N/A[b]
Technical issues	109 (4.71)	Technical discussions on AI in medical care	"Massive medical data is indispensable for AI medical treatment"
General uses	210 (9.07)	General public discussions on AI in medical care without pointing out specific medical areas	"I am looking forward to AI being applied in medicine, autonomous vehicles, and other fields so that ultimately we will all be beneficiaries of AI"
Specific uses	668 (28.85)	Medical AI applied to specific medical fields	"Vigorous development of the application of AI in treatment and rehabilitation."
Prevention	52 (2.25)	Application of AI in medical prevention	"How amazing AI is to be able to predict Alzheimer's nine years in advance."
Diagnosis	436 (18.83)	Application of AI in medical diagnosis	"Google's deep learning AI can diagnose cancer with an accuracy rate of 89%, while the accuracy rate of human diagnosis is currently 73%."
Treatment	146 (6.31)	Application of AI in medical treatment	"Intelligent doctors can list multiple treatment plans in 10 seconds"
Recovery	34 (1.47)	Application of AI in medical recovery	"Intelligent family service robots have achieved mass production and application; robots helping in the recovery and assistance of the elderly and disabled are at the prototype production stage."
Industry development	706 (30.50)	Public attention toward and views on the medical AI industry	N/A
Development of related companies	331 (14.30)	Public discussions about companies that make advancements in medical AI	"Alibaba signed contracts with the First Affiliated Hospital and began to march toward AI medical treatment."
Investment and market	186 (8.03)	Public discussions on the financial issues associated with medical AI	"Pony Ma, Tencent's boss, will invest in artificial intelligence medicine"
Industry expectations	142 (6.13)	Public prospects for the developmental trend of AI in medical care	"I think the AI medical industry has a bright future"
Policy and law	47 (2.03)	Public discussions on privacy and legal issues in AI industry development	"The development of AI can greatly shorten the training cycle of doctors, but the relevant legal and technical norms still need improvement."
Impact on society	622 (26.87)	Influence of AI in medical care on society	N/A
Impact on doctors	422 (18.23)	Public perception of medical AI's impact on human doctors	"After the development of artificial intelligence, doctors are one of the first unemployed professions. Don't let children study medicine"
Impact on hospitals	125 (5.40)	Public perception of medical AI's impact on hospitals	"Will AI and telemedicine help hospitals control medical costs?"
Influence on public life	75 (3.24)	Public perception of medical AI's impact on public life	"When AI enters the medical field, can human beings live forever?"

[a]AI: artificial intelligence.

[b]N/A: not applicable.

The third category, "Impact on society" (n=622, 26.87%), includes the subcategories of impact on doctors (n=422, 18.23%), impact on hospitals (n=125, 5.40%), and influence on public life (n=75, 3.24%). Notably, people cared a lot about the impact on doctors, discussing both positive and negative implications, such as reduced workload and improved efficiency, as well as the need to deal with the threat of doctors being replaced. Medical AI would simultaneously impact health care modalities in hospitals and public life as well.

Public Attitudes Toward AI in Medical Care

Overview

A total of 956 posts displayed users' specific attitudes toward AI in medical care. Of these, 568 posts (59.4%) expressed positive attitudes, 329 (34.4%) conveyed neutral attitudes, and 59 (6.2%) highlighted negative attitudes. In order to understand why people hold these attitudes, we further analyzed the specific viewpoints in these posts (Table 3).

XSL•FO
RenderX

Table 3. Distributions of public attitudes toward artificial intelligence in medical care (N=956).

Attitudes	Number of posts, n (%)
Positive attitudes	568 (59.4)
Artificial intelligence's technical advantages in medical care	251 (26.3)
Optimism about industrial development	229 (24.0)
Helping human doctors	46 (4.8)
Avoiding doctor-patient conflicts	24 (2.5)
Promoting reform of medical care	11 (1.2)
New health expectations	7 (0.7)
Neutral attitudes	329 (34.4)
Information sharing, noncommittal	216 (22.6)
Hesitation	85 (8.9)
Commentary on both positive and negative aspects	28 (2.9)
Negative attitudes	59 (6.2)
Immaturity of artificial intelligence technology	27 (2.8)
Distrust of artificial intelligence companies	15 (1.6)
Fear of artificial intelligence technology	7 (0.7)
Lack of "enthusiasm" expressed by artificial intelligence	5 (0.5)
Privacy	3 (0.3)
Ethics and law	2 (0.2)

Positive Attitudes

As illustrated in Table 3, AI's technical advantages in medical care was the most mentioned reason (251/568, 44.2%) for positive attitudes, in which people identified benefits such as high diagnostic accuracy and computational efficiency. Another main reason for positive attitudes was the optimism about industrial development (n=229, 40.3%), in which people expressed confidence and expectations for medical AI's industrial development. Other reasons associated with AI making contributions to the medical field were also discussed, such as helping human doctors (n=46, 8.1%), avoiding doctor-patient conflicts (n=24, 4.2%), and promoting the reform of medical care (n=11, 1.9%). People believed that medical AI can not only improve human doctors' work quality and reduce their workload but also help avoid doctor-patient conflicts. They also believed that medical AI could maintain fairness, eliminate discrimination against patients, and would not involve under-the-table fees, thus helping reduce arguments and conflicts. In addition, people felt that medical AI would affect the development of the medical system and promote the reform of medical care, which would, in turn, boost health care efficiency and benefit more patients.

Neutral Attitudes

Posts expressing neutral attitudes conveyed noncommittal information (216/329, 65.7%). Users mainly forwarded AI-related posts or news, either summarizing the content or including ambiguous attitudes toward medical AI (eg, "AI new technology - Nature magazine reports: AI is better than doctors in early diagnosis of autism in children"). People conveyed a sense of hesitation (n=85, 25.8%) to accept medical AI, considering that its use is not widespread and it requires time to be tested. Others provided commentary on both positive and negative aspects (n=28, 8.5%) of medical AI and did not express a preference for either view (eg, "AI may eradicate disease and poverty, but it may also destroy human beings").

Negative Attitudes

For posts displaying negative attitudes toward AI in medical care, the immaturity of AI technology (27/59, 45.8%) was the leading reason for doubt. People believe AI is far from mature and must overcome many technical difficulties, including those related to obtaining high-quality medical data, such as data fraud and obstruction of hospitals (eg, "Healthcare involves huge vested interests…huge resistance needs to be overcome"), and the difficulty of standardizing medical treatment for AI (eg, "This kind of complex inspection is difficult to standardize. And it is too difficult for AI"). Distrust of AI companies (n=15, 25.4%) also accounted for negative attitudes. Some people believed that AI-related companies would use medical AI to earn money regardless of the patients' health (eg, "AI recommends incompetent hospitals that only spend their money on advertising"). Some people mentioned a fear of AI technology (7/59, 11.9%) and the lack of "enthusiasm" expressed by AI (n=5, 8.5%), which indicated that they tended to compare AI with humans and were greatly concerned with humanistic problems. Notably, only a few people mentioned fears about the problem of privacy, ethics, and law (n=2, 3.4%) in using medical AI.

XSL•FO
RenderX

Public Attitudes Toward Replacing Human Doctors With AI

For many decades, the fear that AI will cause widespread unemployment has waxed and waned; however, some researchers believe that fears of robots and computers greatly increasing secular unemployment are unwarranted [32]. However, the relationship between AI and human workers, especially the concern that AI workers can replace human workers, has become one of the most controversial issues in the development of the AI industry. Since this is also true in the field of medical AI, we paid special attention to public attitudes toward the issue of whether AI doctors can replace human doctors.

A total of 200 posts referred to the topic of AI replacing human doctors, and 80.0% of the posts conveyed the sense that AI doctors can completely or partially replace human doctors. Of these, 95 posts (47.5%) expressed the users' belief that AI would replace all human doctors. We further identified the reasons for such attitudes. The results indicated that the most important reasons for people believing that AI will replace human doctors is that AI has technical advantages in medical care, such as high accuracy, stability, and efficiency. In addition, people expressed their hope that AI doctors would have the advantage of not having conflicts with their patients. There were 65 posts (32.5%) that expressed the users' belief that AI would partially replace human doctors. Such users believed that some medical jobs, such as pathological diagnosis, are suited to AI, while others are more suitable for human doctors. To determine which kinds of doctors were considered most likely to be replaced by AI, we further coded the posts that explicitly mentioned the doctors that users believed might be completely or partially replaced by AI. Pathologists were the most frequently mentioned (n=17, 43.6%), followed by radiologists (n=8, 20.5%) and dermatologists (n=3, 7.7%). In terms of medicine and surgery, physicians (n=4, 10.3%) were considered more likely to be replaced by AI than surgeons (n=2, 5.1%).

Moreover, 40 posts (20.0%) expressed the attitude that AI will not replace human doctors. The reasons for the public holding this attitude can be classified mainly in terms of technical and humanistic concerns. First, owing to the immaturity of AI technology, people believe that it cannot adequately manage medical problems. Second, people believed patients need humanistic interactions, which AI doctors could never offer. Third, considering ethics and laws, people believed that AI doctors might invade personal privacy and lack legal supervision. Finally, a very small percentage of users opposed AI doctors based on their fear of AI technology.

Discussion

Principal Results

Applying AI to medical care is believed to be a promising solution for the global shortage of medical resources. As a new field, whether the medical AI industry can develop in a healthy and smooth way depends not only on numerous technological challenges, but also on whether the public can accept and trust it. However, we still lack sufficient empirical evidence to reveal actual public perception of medical AI. Based on social media data from Sina Weibo, this study took the lead in revealing the public perception of medical AI through content analysis. We mainly explored the general characteristics of people's attention toward medical AI, the topics of public interest in medical AI, people's attitudes toward medical AI, and their opinions on the debate regarding whether AI doctors can replace human doctors.

Overall, we found that public attention toward AI in medical care showed a noticeable event-driven trend on social media. Big social events, especially social entertainment events related to AI (eg, the world Go champion matchup between Sedol Lee and AlphaGo), generally cause a significant increase in public discussions about AI in medical care. To this end, we suggest that practitioners pay close attention to popular events related to AI and use these critical time points to promote the dissemination of knowledge related to medical AI. In addition, we revealed the profile of social media users who discussed AI in medical care; older adults, males, and people living in more affluent regions paid more attention to the issue of medical AI. These results are in accordance with previous findings that men prefer new technology products more than women [33]. As for the differences in interest across age groups, we think it may be related to the fact that older people may pay more attention to health and medical issues than younger people. Moreover, people in affluent regions may be more likely to seek information about medical AI [34] and have access to it, thus they pay more attention to it. The differential interest in medical AI between richer and relatively poorer regions in this study may support the theory of digital divide [35], and deepen the digital divide related to medical AI. How to promote the new medical technology in remote areas is also worthy of attention and future in-depth study by theorists and practitioners.

When investigating which topics interest people when discussing AI in medical care, we found that public interest mainly focused on technology and application, industry development, and societal impact. The public was most interested in AI technology and application. The specific uses of medical AI, especially in diagnosis, gained the most public attention under the category of technology and application. Our results may provide more precise guidance for the promotion and popularization of the technology and use of medical AI. In addition, the industry development and investment possibilities of medical AI (eg, development of related companies) attracted nearly one-third of public attention, among which only about 2% of posts focused on policy and law in the medical AI industry. We believe that the public's interest in industrial development is beneficial for companies that wish to rapidly promote medical AI, but the government and other regulatory authorities may need to pay attention to strengthening the formation and dissemination of relevant policies and laws in the future.

In terms of attitudes toward medical AI, our results showed that the public was optimistic, with nearly 60% of posts expressing positive attitudes. The preponderance of medical AI supporters among the general public implies that the promotion and popularization of medical AI had a relatively good public psychological basis. However, nearly 40% of public posts were neutral or opposed to AI. We found that approximately one-third of all users played the role of bystanders regarding AI in medical

XSL•FO
RenderX

care. This means that the current public attitude toward AI in medicine has considerable plasticity, and neutral people may require the most attention. Moreover, some people clearly expressed negative attitudes toward medical AI. The results showed that the immaturity of AI technology, distrust of AI companies, and fear of AI were the top three reasons for these negative attitudes. On the technical side, people questioned the ability of AI and believed that it is difficult to standardize medical care, and some even expressed a natural fear regarding the safety of AI technology. This means that it will take more time for medical AI to gain people's trust. At the organizational level, an attitude analysis revealed that people cared a lot about medical AI–related companies, and their negative attitudes could be due to their distrust of related companies (as observed with the Cambridge Analytica scandal and the #DeleteFacebook campaign). This indicates that managers from medical AI companies should not confine themselves to publicizing their technical advantages in the field of AI; rather, they should pay attention to how to actively portray a corporate image with a sense of social responsibility and trust to the public. The absence of humane care in AI was also an important reason for people's negative attitudes, which is consistent with previous studies [36]. It is noteworthy that few posts (0.5%) were concerned about issues relating to privacy, ethics, or the law in the context of medical AI. As shown in prior studies, a collectivist culture results in a lower level of privacy concerns than in an individualistic culture [37,38]. We speculate that China's collectivist culture and limited medical resources may have resulted in people from China being more interested about the benefits of AI than personal privacy.

On the controversial issue of whether AI could replace human doctors, we found that 80% of the posts indicated that AI doctors could replace human doctors completely or partially. This result was contrary to previous findings, which showed that 83% of medical students, most general practitioners, and the majority of physicians believe that AI will not replace them [16-18]. We speculate that this may be related to the psychological mechanism of self-defense of medical practitioners as stakeholders, which needs more direct evidence support from future research. It seems that the general public held a more open view about replacing doctors with AI than medical professionals, although the public is more conservative than AI experts. There is similar evidence that the general public was more enthusiastic about new medical technology than professional doctors [39]. Additionally, previous research also indicated that AI experts have made comments to the media that AI would soon replace doctors [15], and the public's attitude may be affected by these comments. However, one-fifth of the posts were against AI replacing human doctors. Their pessimism about AI's replacement of human doctors was mainly due to the immaturity of AI technology and AI's inability to express empathy and compassion. These findings correspond with those from previous studies that showed that the role of humanistic care in medical health has become increasingly prominent [40]. It is also worth noting that the public's attitudes on whether AI will replace human doctors may also affect medical enrollment and talent supply. Young people may be reluctant to go to medical school due to the fear that doctors will lose their jobs in the future. How to balance the potential conflict between the development of medical AI and the supply of medical talent merits discussion among governments, universities, and relevant practitioners.

Limitations and Future Directions

While our study contributes to understanding the public perception of AI in medical care, there are a few limitations worth noting. First, as our study focused on users of the social media platform Sina Weibo, our sample may have been younger and more educated than the broader population. Since younger and more educated people may be more open to new technologies, we may have overestimated the public's optimistic attitude toward AI in medical care. Second, the data used in this study were specific to China; the unique cultural environment may have affected people's perception of and attitudes toward medical AI; thus, the findings may not be generalizable to other countries. In this study, we found that the public in China paid much more attention to the technology and application of medical AI than to privacy risks. Whether these conclusions can be generalized to Western countries remains to be examined in future studies. However, as the most populous country in the world, the shortage of medical resources in China is severe. It is of great practical significance to explore medical AI in China. At the same time, China has the largest number of internet users in the world and has a particular advantage in the development of AI technology. Therefore, we believe that it is an important issue to clarify the Chinese public perception of and attitude toward medical AI. Nevertheless, future research can consider exploring the public perception of medical AI in other countries. Third, our analysis based on social media data has some inevitable limitations in inferring users' apparent attitude and behavior intention. In the future, such research can be supplemented and improved with research evidence, such as a questionnaire survey.

Conclusions

Our study presents a social listening method of assessing public perception and opinions on AI in medical care through social media content analysis. Our findings indicate that social events can easily drive the public's attention to medical AI, and that older adults, males, and people living in more affluent regions were more interested in it. The general public was most interested in AI technology and application, especially its specific uses in diagnosis, but showed little interest in policy and law in the medical AI industry. The characteristics of the public attention on medical AI that we found in this study may provide practical guidance for promoting this new medical technology. Still, we need to pay close attention to the present digital divide that may deepen further. In terms of attitudes toward medical AI, the majority of people held positive attitudes and believed that AI doctors would completely or partially replace human doctors. In the aggregate, the general public attitudes toward medical AI were more open than those of medical professionals but more conservative than those of AI experts.

Although they were generally optimistic about AI in medical care, some people still had negative attitudes toward medical AI owing to the immaturity of AI technology and their distrust of AI companies. Our results revealed that distrust of companies

XSL•FO
RenderX

accounted for one-quarter of all negative attitudes toward AI. Improving the public's trust in AI companies may take more time than upgrading the technology. In addition, technical and humanistic concerns were the most important reasons for some people's pessimism about AI replacing human doctors. We suggest that in the future, practitioners should pay more attention to humanistic care and try to meet patients' emotional needs, rather than only focusing on technical issues. It is worth noting that currently, AI in medical care is still in its early stages, and most people have not come into contact with it. Therefore, their attitudes are likely to change with the development of medical AI products. Future research should keep track of the changes and progress of public opinion toward medical AI in the long run. We hope this study can serve as a catalyst for the understanding of public perception on medical AI and expand the ongoing conversation to additional communities.

Acknowledgments

This work was supported by the Major Project of the National Social Science Foundation of China (award number 18ZDA165).

Conflicts of Interest

None declared.

References

1. Aluttis C, Bishaw T, Frank MW. The workforce for health in a globalized context--global shortages and international migration. Glob Health Action 2014;7:23611 [FREE Full text] [doi: 10.3402/gha.v7.23611] [Medline: 24560265]
2. Lawrence DR, Palacios-González C, Harris J. Artificial Intelligence. Camb Q Healthc Ethics 2016 Apr;25(2):250-261. [doi: 10.1017/S0963180115000559] [Medline: 26957450]
3. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
4. Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res 2018 Jul 13;18(1):545 [FREE Full text] [doi: 10.1186/s12913-018-3359-4] [Medline: 30001717]
5. Hadsell R, Sermanet P, Ben J, Erkan A, Scoffier M, Kavukcuoglu K, et al. Learning long-range vision for autonomous off-road driving. J. Field Robotics 2009 Feb;26(2):120-144. [doi: 10.1002/rob.20276]
6. Anandharajan T, Hariharan G, Vignajeth K, Jijendiran R, Kushmita. Weather monitoring using artificial intelligence. : IEEE; 2016 Jan 11 Presented at: Proceedings of the 2nd International Conference on Computational Intelligence and Networks; 2016; Bhubaneswar, India p. 106-111. [doi: 10.1109/CINE.2016.26]
7. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
8. The Lancet. Artificial intelligence in health care: within touching distance. Lancet 2018 Dec 23;390(10114):2739. [doi: 10.1016/S0140-6736(17)31540-4] [Medline: 29303711]
9. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. JAMA 2016 Dec 13;316(22):2368-2369. [doi: 10.1001/jama.2016.17217] [Medline: 27898974]
10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
11. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW. Supervised autonomous robotic soft tissue surgery. Sci Transl Med 2016 May 04;8(337):337ra64. [doi: 10.1126/scitranslmed.aad9398] [Medline: 27147588]
12. Sadilek A, Kautz H, DiPrete L, Labus B, Portman E, Teitel J, et al. Deploying nEmesis: Preventing Foodborne Illness by Data Mining Social Media. AIMag 2017 Mar 31;38(1):37. [doi: 10.1609/aimag.v38i1.2711]
13. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119. [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
14. Kohane IS, Drazen JM, Campion EW. A glimpse of the next 100 years in medicine. N Engl J Med 2012 Dec 27;367(26):2538-2539. [doi: 10.1056/NEJMe1213371] [Medline: 23268669]
15. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219. [doi: 10.1056/nejmp1606181]
16. Pinto Dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
17. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]

XSL•FO
RenderX

18. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial Intelligence and the Future of Primary Care: Exploratory Qualitative Study of UK General Practitioners' Views. J Med Internet Res 2019 Mar 20;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
19. Lazer D, Pentland A, Adamic L, Aral S, Barabasi A, Brewer D, et al. Social science. Computational social science. Science 2009 Feb 06;323(5915):721-723 [FREE Full text] [doi: 10.1126/science.1167742] [Medline: 19197046]
20. Neubaum G, Krämer NC. Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media. Media Psychology 2016 Aug 05;20(3):502-531. [doi: 10.1080/15213269.2016.1211539]
21. Sutton J, Vos SC, Olson MK, Woods C, Cohen E, Gibson CB, et al. Lung Cancer Messages on Twitter: Content Analysis and Evaluation. J Am Coll Radiol 2018 Dec;15(1 Pt B):210-217. [doi: 10.1016/j.jacr.2017.09.043] [Medline: 29154103]
22. Sinnenberg L, DiSilvestro CL, Mancheno C, Dailey K, Tufts C, Buttenheim AM, et al. Twitter as a Potential Data Source for Cardiovascular Disease Research. JAMA Cardiol 2016 Dec 01;1(9):1032-1036 [FREE Full text] [doi: 10.1001/jamacardio.2016.3029] [Medline: 27680322]
23. McGovern SK, Blewer AL, Murray A, Leary M, Abella BS, Merchant RM. Characterizing barriers to CPR training attainment using Twitter. Resuscitation 2018 Jun;127:164-167. [doi: 10.1016/j.resuscitation.2018.03.010] [Medline: 29545139]
24. Becker BFH, Larson HJ, Bonhoeffer J, van Mulligen EM, Kors JA, Sturkenboom MCJM. Evaluation of a multinational, multilingual vaccine debate on Twitter. Vaccine 2016 Dec 07;34(50):6166-6171. [doi: 10.1016/j.vaccine.2016.11.007] [Medline: 27840012]
25. Keller MS, Park HJ, Cunningham ME, Fouladian JE, Chen M, Spiegel BMR. Public Perceptions Regarding Use of Virtual Reality in Health Care: A Social Media Content Analysis Using Facebook. J Med Internet Res 2017 Dec 19;19(12):e419 [FREE Full text] [doi: 10.2196/jmir.7467] [Medline: 29258975]
26. Gao Q, Abel F, Houben G, Yu Y. A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter. : Springer; 2012 Jul 16 Presented at: International Conference on User Modeling, Adaptation, and Personalization; 2012; Montreal, Canada. [doi: 10.1007/978-3-642-31454-4_8]
27. Tian X, Yu G, He F. An analysis of sleep complaints on Sina Weibo. Computers in Human Behavior 2016 Sep;62:230-235. [doi: 10.1016/j.chb.2016.04.014]
28. Chen L, Wang X, Peng T. Nature and Diffusion of Gynecologic Cancer-Related Misinformation on Social Media: Analysis of Tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]
29. Pan J, Liu B, Kreps GL. A content analysis of depression-related discourses on Sina Weibo: attribution, efficacy, and information sources. BMC Public Health 2018 Jun 20;18(1):772 [FREE Full text] [doi: 10.1186/s12889-018-5701-5] [Medline: 29925363]
30. Jiang X, Jiang W, Cai J, Su Q, Zhou Z, He L, et al. Characterizing Media Content and Effects of Organ Donation on a Social Media Platform: Content Analysis. J Med Internet Res 2019 Mar 12;21(3):e13058 [FREE Full text] [doi: 10.2196/13058] [Medline: 30860489]
31. Reiter O, Rotemberg V, Kose K, Halpern AC. Artificial Intelligence in Skin Cancer. Curr Derm Rep 2019 Jul 6;8(3):133-140. [doi: 10.1007/s13671-019-00267-0]
32. Diamond AM. Robots and Computers Enhance Us More Than They Replace Us. The American Economist 2018 Aug 10;65(1):4-10. [doi: 10.1177/0569434518792674]
33. Schumacher P, Morahan-Martin J. Gender, Internet and computer attitudes and experiences. Computers in Human Behavior 2001 Jan;17(1):95-110. [doi: 10.1016/S0747-5632(00)00032-7]
34. Wang MP, Wang X, Viswanath K, Wan A, Lam TH, Chan SS. Digital inequalities of family life information seeking and family well-being among Chinese adults in Hong Kong: a population survey. J Med Internet Res 2014 Oct 03;16(10):e227 [FREE Full text] [doi: 10.2196/jmir.3386] [Medline: 25281831]
35. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]
36. Miller DD, Brown EW. Artificial Intelligence in Medical Practice: The Question to the Answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
37. Nemati H, Wall JD, Chow A. Privacy Coping and Information-Sharing Behaviors in Social Media: A Comparison of Chinese and U.S. Users. Journal of Global Information Technology Management 2014 Dec 11;17(4):228-249. [doi: 10.1080/1097198x.2014.978622]
38. Lowry PB, Cao J, Everard A. Privacy Concerns Versus Desire for Interpersonal Awareness in Driving the Use of Self-Disclosure Technologies: The Case of Instant Messaging in Two Cultures. Journal of Management Information Systems 2014 Dec 08;27(4):163-200. [doi: 10.2753/mis0742-1222270406]
39. Boeldt DL, Wineinger NE, Waalen J, Gollamudi S, Grossberg A, Steinhubl SR, et al. How Consumers and Physicians View New Medical Technology: Comparative Survey. J Med Internet Res 2015;17(9):e215 [FREE Full text] [doi: 10.2196/jmir.4456] [Medline: 26369254]
40. Simpkin AL, Dinardo PB, Pine E, Gaufberg E. Reconciling technology and humanistic care: Lessons from the next generation of physicians. Med Teach 2017 Apr;39(4):430-435. [doi: 10.1080/0142159X.2017.1270434] [Medline: 28013558]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence

Edited by E Borycki; submitted 10.10.19; peer-reviewed by D Pinto dos Santos, H Yang, D Chrimes, D Chrimes, A Kushniruk; comments to author 12.02.20; revised version received 02.04.20; accepted 31.05.20; published 13.07.20.

Please cite as:
Gao S, He L, Chen Y, Li D, Lai K
Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media
J Med Internet Res 2020;22(7):e16649
URL: http://www.jmir.org/2020/7/e16649/
doi:10.2196/16649
PMID:32673231

©Shuqing Gao, Lingnan He, Yue Chen, Dan Li, Kaisheng Lai. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study

Filipe Barata[1], MSc; Peter Tinschert[2], MSc; Frank Rassouli[3], MD, PhD; Claudia Steurer-Stey[4,5], MD, PhD; Elgar Fleisch[1,2], PhD; Milo Alan Puhan[4], MD, PhD; Martin Brutsche[3], MD, PhD; David Kotz[1,6,7], PhD; Tobias Kowatsch[1,2], PhD

[1]Center for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

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

[3]Lung Center, Cantonal Hospital St. Gallen, St. Gallen, Switzerland

[4]Institute of Epidemiology, Biostatistics and Prevention, University of Zurich, Zurich, Switzerland

[5]mediX Group Practice, Zurich, Switzerland

[6]Department of Computer Science, Dartmouth College, Hanover, NH, United States

[7]Center for Technology and Digital Health, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Filipe Barata, MSc
Center for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/57
Zurich, 8092
Switzerland
Phone: 41 446323509
Email: fbarata@ethz.ch

Abstract

Background: Asthma is one of the most prevalent chronic respiratory diseases. Despite increased investment in treatment, little progress has been made in the early recognition and treatment of asthma exacerbations over the last decade. Nocturnal cough monitoring may provide an opportunity to identify patients at risk for imminent exacerbations. Recently developed approaches enable smartphone-based cough monitoring. These approaches, however, have not undergone longitudinal overnight testing nor have they been specifically evaluated in the context of asthma. Also, the problem of distinguishing partner coughs from patient coughs when two or more people are sleeping in the same room using contact-free audio recordings remains unsolved.

Objective: The objective of this study was to evaluate the automatic recognition and segmentation of nocturnal asthmatic coughs and cough epochs in smartphone-based audio recordings that were collected in the field. We also aimed to distinguish partner coughs from patient coughs in contact-free audio recordings by classifying coughs based on sex.

Methods: We used a convolutional neural network model that we had developed in previous work for automated cough recognition. We further used techniques (such as ensemble learning, minibatch balancing, and thresholding) to address the imbalance in the data set. We evaluated the classifier in a classification task and a segmentation task. The cough-recognition classifier served as the basis for the cough-segmentation classifier from continuous audio recordings. We compared automated cough and cough-epoch counts to human-annotated cough and cough-epoch counts. We employed Gaussian mixture models to build a classifier for cough and cough-epoch signals based on sex.

Results: We recorded audio data from 94 adults with asthma (overall: mean 43 years; SD 16 years; female: 54/94, 57%; male 40/94, 43%). Audio data were recorded by each participant in their everyday environment using a smartphone placed next to their bed; recordings were made over a period of 28 nights. Out of 704,697 sounds, we identified 30,304 sounds as coughs. A total of 26,166 coughs occurred without a 2-second pause between coughs, yielding 8238 cough epochs. The ensemble classifier performed well with a Matthews correlation coefficient of 92% in a pure classification task and achieved comparable cough counts to that of human annotators in the segmentation of coughing. The count difference between automated and human-annotated

XSL•FO
RenderX

coughs was a mean –0.1 (95% CI –12.11, 11.91) coughs. The count difference between automated and human-annotated cough epochs was a mean 0.24 (95% CI –3.67, 4.15) cough epochs. The Gaussian mixture model cough epoch–based sex classification performed best yielding an accuracy of 83%.

Conclusions: Our study showed longitudinal nocturnal cough and cough-epoch recognition from nightly recorded smartphone-based audio from adults with asthma. The model distinguishes partner cough from patient cough in contact-free recordings by identifying cough and cough-epoch signals that correspond to the sex of the patient. This research represents a step towards enabling passive and scalable cough monitoring for adults with asthma.

(J Med Internet Res 2020;22(7):e18082) doi:10.2196/18082

KEYWORDS

asthma; cough recognition; cough segmentation; sex assignment; deep learning; smartphone; mobile phone

Introduction

Asthma is one of the most prevalent chronic respiratory diseases [1]; it is estimated to afflict 339 million people worldwide [1]. Despite increased investment in treatment, hospital admissions and mortality rates have remained stable over the last decade [2]. This stagnation highlights the lack of scalable and easy-to-manage diagnostic tools for early recognition of exacerbations [3]. Traditional asthma self-management programs include written plans of action that cover how to recognize and respond to exacerbations [4]. These pen-and-paper approaches have shown improved health outcomes such as reduced hospital admissions, better lung function, fewer asthma symptoms, and less use of rescue medication [4-6]. The implementation of these programs, however, is low in clinical practice, with only 27% of adults with asthma receiving an asthma action plan [7]. Also, patient adherence to written action plans is poor and declines over time [8], and due to the reliance on subjective information in action plans, patients may perceive their symptoms poorly, and thus, underestimate the severity of the disease [9].

Disease control (also referred to as asthma control) is defined as the degree to which symptoms are controlled by treatment [10]. Exacerbations of asthma occur gradually over several days to weeks on a background of poor asthma control [11,12]. Although there is consensus on the importance of longitudinal and objective assessment of asthma control in the home environment, few tools enable objective measurement [3]. Current guidelines for the assessment of asthma control recommend the measurement of peak expiratory flow [13]. Changes in peak expiratory flow can occur up to 2 weeks before an exacerbation [14]. This parameter, however, is dependent on effort [3], and only 5% of adults with asthma measure peak expiratory flow regularly [15]; therefore, it has limited utility. Nocturnal cough is a physiologic parameter which has shown promise for use in the assessment of asthma. Cough is a particularly burdensome asthma symptom [16] and was shown to be associated with asthma severity [17] and a worse prognosis [18]. Cough counts per night (ie, the number of coughs produced by one individual per night) have been associated with the level of asthma control [19]. Moreover, lower levels of asthma control have been correlated with poor quality of life [20] and are a predictor for cost-of-illness [21]. Coughing may also provide valuable information to predict the effects of asthma therapy early on [22]. Thus, nocturnal cough monitoring may play an important role in the prevention of exacerbations and in the personalization of treatment.

The monitoring of coughing by quantifying the number of coughs per unit of time has been attempted by researchers since the 1950s [23]. Although cough events may be counted manually from sound and video recordings, this process is extremely laborious. In consequence, many semi and fully automated cough monitoring systems have been in development since the 1950s. Among the more well-known systems are the Hull Automatic Cough Counter [24], the Leicester Cough Monitor [25], and LifeShirt [26]. The Hull Automatic Cough Counter is a software program written in MATLAB with the ability to differentiate cough from noncough sounds using Sony Walkman tape recordings. Leicester Cough Monitor is based on a free-field microphone necklace that records sound continuously onto a digital recorder. Recordings can then be uploaded onto a computer where an automated cough detection algorithm analyzes them. LifeShirt is an ambulatory cardiorespiratory monitoring system with a unidirectional contact microphone. Though these systems lacked scalability and cost-effectiveness, they showed feasibility with respect to automatic cough detection and counting from audio recordings.

Smartphones are now ubiquitous [27] and are equipped with sensors capable of many types of monitoring with clinically valuable accuracy [28]. Their widespread adoption in all age groups enables them to be used for measurements within different population samples [29]. In addition, they can be used to passively monitor the health status of patients without an additional task of activating a monitor. Recent advances have used only a smartphone and its built-in microphone for cough monitoring [30-32].

To develop an automatic cough monitoring system is challenging due to the rare occurrence of coughing in comparison to the occurrence of other sounds. This natural imbalance of cough and noncough sounds poses two problems. First, it demands high specificity from the cough monitoring system to avoid false alarms from other similar and more frequently occurring sounds. Second, existing classification methods tend to perform poorly on minority-class examples when the data set is highly imbalanced [33]. In addition, experts have suggested sensitivities greater than or equal to 90% as necessary for clinical use [34]. An even greater challenge is to collect sufficient data in the intended context and timeframe to allow realistic assessment of the monitoring system, particularly

XSL•FO
RenderX

with respect to respiratory conditions. Some cough monitoring systems have been developed based on data collected under lab conditions by recording voluntary coughs [35-37]. While this may be a valid approach to compare the performance of different classifiers, it may not represent the soundscape of a real-use case. No cough monitoring systems exist that have undergone longitudinal overnight testing nor do any exist that have been specifically evaluated in the context of asthma. Also, depending on the intended mode of use, new challenges arise for cough monitoring—distinguishing partner coughs from patient coughs in a room with two or more people using contact-free recordings is a difficult task. To date, there are no standardized methods, and there are no sufficiently validated cough monitors for general use that are commercially available and clinically acceptable [23].

The study was designed to mimic a real-world use case; data were collected from a smartphone placed on the bedside table in the participant's bedroom. The aim was to build a cough classifier and to evaluate its performance on unseen data. Further, we aimed to use the classifier to segment and count cough events over the course of the night. Building upon previous work [32], we adopted a convolutional neural network architecture , which performed best in comparison to that of other machine learning approaches when using voluntary cough data from different smartphone recordings. We altered the learning part of the algorithm by combining three different techniques from literature to combat the high class imbalance encountered in this real-life data set—ensemble learning [38], minibatch balancing [39], and decision thresholding [40].

Though a smartphone-based nocturnal cough monitoring service may enable passive monitoring in theory, its utility for application in practice depends on whether coughs can be correctly assigned to individuals. Prior research has shown that humans are able to determine whether the source of a cough is male or female based on sound alone [41]. In addition, sex-based differences in signal properties have been measured in cough signals [42]. Epidemiological research suggests that less than 10% of the general population identify as homosexual [43]. Assuming that most people either sleep alone or share their bedroom with a partner, for the vast majority of patients, correctly classifying cough by sex-based properties of the cough signal could allow a cough monitoring system to disregard the coughs that are not from the individual of interest. Therefore, in addition to cough detection, this work examines to what extent coughs can be correctly be classified by sex. This research represents a step towards enabling passive scalable monitoring for people with asthma.

Methods

Overview

This study involved the collection of smartphone-recorded audio and daily questionnaire data, the definition and quantification of coughs within that data (data annotation and automated cough recognition and segmentation), sex-based classification, and model performance evaluation.

Data Collection

We used data collected in a multicenter, longitudinal observational study over a 29-day period (28 nights) [41]. On the first and last day, participants underwent medical examination by health professionals at the study centers. At the start of the study, participants were equipped with a smartphone (Samsung Galaxy A3 2017, SM-A320FL) on which Clara—the study's chat-based app—was installed. This app was a study-specific enhancement of the mobile app in the open-source MobileCoach behavioral intervention platform [44,45]. At night, the app recorded audio data using the smartphone's microphone. It also delivered daily questionnaires to the patients, asking them (among other things) whether the participant slept alone.

All participant data were collected by the physician (asthma evaluation data) or the nurse (lung function evaluations) in the study centers and were transferred to an electronic format and stored online on the study server. Nightly sensor data were stored locally on the smartphone. Data were backed up to external hard drives and secure online storage once the participant had completed the study and had returned the smartphone.

The study protocol was reviewed and approved by the *Ethikkommission Ostschweiz*, which is responsible for research on humans in Eastern Switzerland (Business Management System for Ethics Committees ID: 2017–01872).

Cough Definition and Quantification

With respect to cough monitoring systems, the definition of cough depends on the modality used for monitoring [23]. In this study, we aimed to recognize coughs from sound recordings which can be done in several ways [23]. We focused on two methods: (1) counting explosive cough sounds and (2) counting cough epochs (continuous coughing sounds without a 2-second pause [23]). The latter is derived from the first metric by computing the duration between explosive cough sounds.

Data Annotation

Before annotation, silence was marked by applying a decibel filter by means of the Audacity software to the recordings. The Sound Finder filter marked sounds below –26 dB as silence with the constraint that the minimum duration of silence between sounds was 1 second. These periods marked as silence served as visual aids for the remainder of the annotation process. Human annotators listened to the smartphone recordings and labeled the periods that were not marked as silence as a cough if an explosive cough sound was identified [23,41].

We used two approaches to verify the quality of the labeling. First, we instructed human annotators to label an acoustic event if they were unsure that it was a cough. If annotators were unsure, the event was discarded and was not considered in the analysis. The remainder of acoustic events were classified as noncoughs. Second, the interrater reliability for the annotators was determined using intraclass correlation. A zero-inflated generalized mixed-effects model with a Poisson response was used [46]. Additional details of the annotation method can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Cough Recognition and Segmentation

Data Set Partitioning for Cough Recognition

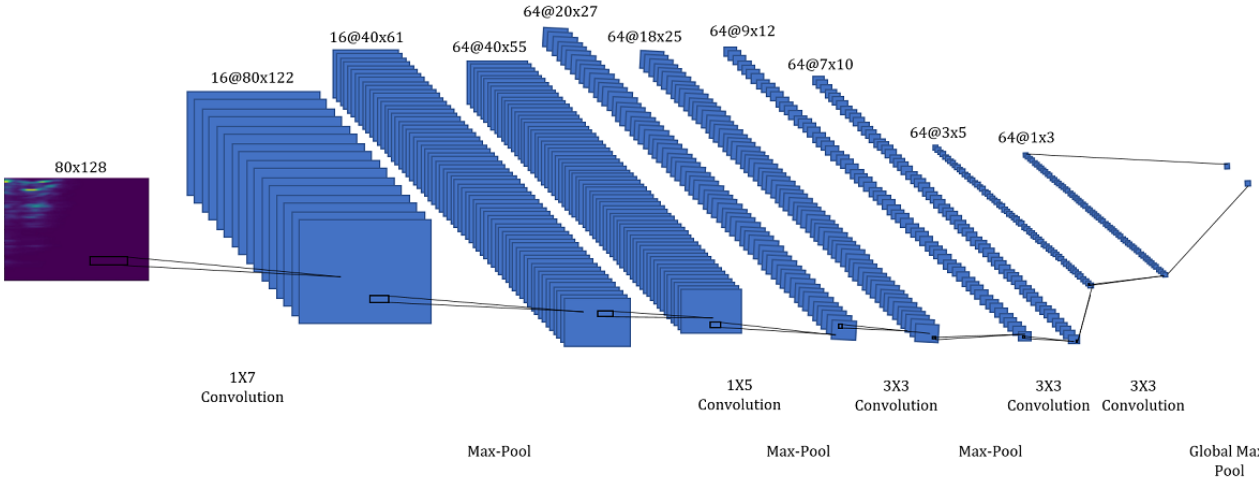
When developing neural networks, the split into sets for training, validation, and testing of the model is favored over other approaches that involve cross-validation because of the long training phases of the models; however, this comes at the risk of overfitting the model to the specific data set and a lack of generalizability to unseen samples. To mitigate these effects, we split our data into disjunct data sets which contained a different set of participants in the training, validation, or test sets. Furthermore, the large number of cough samples and participants in comparison to former studies [7,24,25,30,31] may help mitigate the risk of overfitting.

Training, validation, and test sets were created in the following way. From participants with complete data sets, roughly 20% were drawn at random; these nocturnal audio recordings constituted the testing corpus for our evaluations. From the remaining participants, roughly 15% were drawn at random to be included in the validation set. The remaining participants then comprised the training set. Thus, data were roughly split into a ratio of 65:15:20. For model evaluation, the neural network was first trained on the training set. Hyperparameter tuning and model selection were performed on the validation set. Once the best performing parameters and model were selected, the final training was completed on the unified training and validation set. Model results were derived from the test set. The data split proportions were motivated by the fact that larger amounts of training data improve the performance of the classifier [47].

Neural Network Architecture for Cough Recognition

This work was built upon a convolutional neural network architecture for cough recognition that we introduced in previous work [32] which recognized coughs in Mel spectrograms. Mel-scaled spectrograms are visual representations of audio signals with respect to time and frequency. Signal frequency ranges are Mel-scaled to represent the human perception of sound. In conjunction with convolutional neural network architectures, they have been reported to perform better than other time-frequency representations [48]. The detailed calculation of the Mel spectrograms used can be found in Multimedia Appendix 1. We evaluated this approach against different approaches for smartphone-based cough recognition in previous work on voluntary coughs and found that it performed best [32]. Our approach produced stable results across recordings of five different devices with different hardware and service life duration. Moreover, it was designed to be lightweight and its deployment and energy-efficiency were tested on smartphones. Figure 1 depicts the architecture.

Figure 1. The architecture consists of 5 convolutional layers with alternating max-pooling layers followed by a global max-pooling layer. The annotation "16@80x122" refers to a feature map with dimensions (height x width) and 16 channels. The annotation "1x7 Convolution" refers to a convolutional filter with spatial dimensions (height x width).



Network Training for Cough Recognition

Overview

A common problem in real data sets is that some classes have more samples than others. This class imbalance can have a considerable detrimental effect on convergence during the training phase and generalization of a model on the test set [40]. To counter the high imbalance in our data set, we employed 3 different techniques: ensemble learning, balanced minibatch learning, and decision thresholding. Our training approach can be summarized as follows. First, equally sized windows were extracted from the labeled acoustic cough and noncough events. From the resulting windows, we computed Mel spectrograms and employed data partitioning on the cough and noncough Mel spectrograms. Second, we created 5 folds out of the training data. Third, from each of those folds, a separate convolutional neural network model was trained using balanced minibatch training. These models had a probability score as output that indicated the likelihood of a cough. Finally, a threshold was determined to determine the averaged predicted probability of the trained models. The trained ensemble convolutional neural network classifier including thresholding is referred to as ensemble convolutional neural network throughout this paper.

Window Extraction

To train the classifier, nonoverlapping 650 ms windows from the noncough acoustic events were extracted. The duration of the window was based upon previous work. The same

XSL•FO
RenderX

approximate duration has performed best in other cough monitoring approaches [49]. From the cough events, a single 650 ms window centered around the maximum amplitude was extracted since most cough events were shorter. Mel spectrograms were computed for these windows.

Ensemble Learning

In ensemble learning, combining individual models may help to improve generalization performance, if the individual models are dissimilar [50]. In previous work [32], we employed this approach by varying the subset of the devices from which single classifiers are trained, and thus created dissimilar classifiers. In this data set, we had a vast amount of recordings from one device model at our disposal. We, therefore, aimed to create dissimilar classifiers by training disjunct folds of the data instead. At the same time, by using disjunct folds, we reduced the training duration for a single classifier. Ensemble classifiers have also been applied successfully to other imbalanced data sets in prior work [38]. In this work, we subsampled the noncough class of the training data set into 5 participant-disjunct parts. For each of these subsamples and for the total sample of coughs in the training data, we trained a separate convolutional neural network classifier. In such a manner, we reduced the level of imbalance and still benefited from the increased prediction performance of an ensemble classifier. We limited our approach to 5 classifiers in order to limit computation, which increased linearly with the number of classifiers. To execute one of the convolutional neural network classifiers, 10.74 million floating-point operations are needed. The amount of computation may be of importance if the model is to be deployed on smartphones with limited computational resources. The output of each classifier is a probability score computed in the sigmoid neuron of the architecture.

Balanced Minibatch Training

Although the multilayer architecture of neural networks is essential, it is the back-propagation algorithm [51]that solves the optimization problem of minimizing the error between the samples and the predictions during the training phase using gradient descent. It computes the gradients of this error with respect to the neural network weights and propagates it back over the different layers defined by the architecture.

This process was done iteratively using the learning rule defined as follows:



where n was the number of samples, $\theta^{\tau+1}$ was the new updated weight, θ^{τ} was the current weight, η was the learning rate, E was the error, and ∂ was the differential operator. A more efficient technique for the evaluation of the learning rule is minibatch training [39] which introduces a trade-off when updating the weights. While using all training samples at once would have allowed for a precise update of the weights, it would also have taken longer and would have required excessive memory, and using a single instance to update the weights would have introduced noisy updates and would have been computationally inefficient. Minibatch training, however, splits the training data into small batches and computes weight updates from those batches. We employed balanced minibatch training, which also balanced the amount of majority and minority class instances per batch. Finally, we trained each of the convolutional neural network classifiers using balanced minibatch training and the Adam adaptive learning rate optimization algorithm [52]. This training approach relied on a number of hyperparameters, which were tuned and selected on the validation set, learning rate, batch size, and the number of iterations. Furthermore, weight initialization was accomplished using Xavier initialization [53] and dropout which is a regularization technique to prevent overfitting [54]. A dropout rate of 50% was used.

Thresholding

Thresholding adjusts the decision threshold of a classifier. It is typically applied in the test phase and involves changing the output class probabilities [40]. Since we used balanced minibatch training, in the training phase, we implicitly assumed that cough and noncough windows were each as likely to occur as the other. To account for the class imbalance and to find the best decision threshold, we employed a grid search on probabilities in the range of [0.5, 1) and tuned it on the validation set. The single classifier is one randomly selected classifier out of the 5 classifiers that constitute the ensemble classifier. We computed the threshold-based decision rule for the ensemble classifier as follows:



where $Q_i(x|\lambda_{cough})$ is the probability score of the classifier i for the Mel spectrogram x and the class cough λ_{cough}, and t is the threshold.

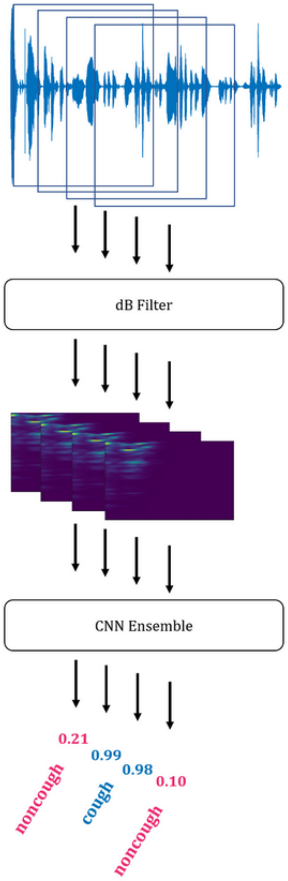
Cough and Cough-Epoch Segmentation

In cough segmentation, the objective is to segment coughs from continuous audio recordings by employing a trained convolutional neural network ensemble classifier. We extracted 650 ms window with an overlap of 65 ms. To discard silent windows, each of these windows passes through a decibel filter that removes windows with sounds below –26 dB as in the annotation process. Subsequently, preprocessed into Mel spectrograms for each window were computed (as described in Multimedia Appendix 1). From these Mel spectrograms, cough probabilities were computed using the trained ensemble classifier. The continuous probabilities that were output were then transformed into segmentations by applying 3 postprocessing rules: (1) Only consecutive probabilities above the derived threshold were considered to be coughs, to reduce the number of false detections. (2) Single probabilities above the threshold were then considered, when the mean of the probability above the threshold and the following probability was above 0.9. (3) When more than eight consecutive detections occurred, they were recognized as two coughs. This was done to compensate for the limited resolution caused by the size of the overlap. These rules were derived by observation and yielded the best results on the validation set. Figure 2 illustrates the cough segmentation process. From the segmentation of cough, cough counts were computed. Cough epochs were recognized when two (or more) coughs occurred without a 2-second pause

XSL•FO
RenderX

in between. An annotated cough epoch was considered to be correctly identified when at least one of the predicted coughs of the cough epoch corresponded with the annotated cough in the cough epoch.

Figure 2. The steps for the segmentation of coughs from continuous audio recordings (from top to bottom): First, the continuous extraction of overlapping windows from continuous audio recordings; second, the discarding of silent windows by applying a dB filter; third, the computation of Mel spectrograms; fourth, the computation of the prediction probability of cough by the convolutional neural network ensemble; last, the recognition of cough by applying the postprocessing rules. CNN: convolutional neural network.



Sex Classification

Data Set Partitioning

For the data set partitioning for determining the source of each cough by sex, we used the complete data set and did not consider the partitioning used for cough recognition. The reason for this lay in the limited amount of data that fulfilled the filtering requirements for the analysis. Since the data collection study included couples or multiple people in one room, we filtered the annotated cough data based on the information collected daily regarding whether the participant slept alone or not. We then filtered the data of the corresponding nights to create a balanced data set of male and female coughs that included 19 female and 19 male participants. We conducted our analysis on both extracted cough and cough-epoch signals. In both cases, we partitioned the data set into a disjunct training set of 10 female and 9 male and a test set of 9 female and 10 male participants.

Gaussian Mixture Models

Gaussian mixture models in combination with Mel-frequency cepstral coefficients [55] are a known method for tackling several different recognition tasks in the audio domain, such as text-independent speaker recognition [56] or gender recognition from speech [57]. For the sex classification of cough signals, we used the 650 ms windows that were labeled as coughs and the resulting cough epochs to compute the Mel-frequency cepstral coefficients. In addition, their first time-derivative estimate and the first time-derivative estimate of the zero-crossing rate computed over the signal were used as features. The idea of training Gaussian mixture models was to approximate the probability distribution of a class by a linear combination of K Gaussian distributions. The likelihood function of feature vector X given class λ can be described as follows:



where $P_k(X|\mu_k, \Sigma_k)$ is the Gaussian distribution. The parameters mean μ_k, covariance Σ_k, and weights ω of the distributions are determined during the training of the features X of class λ.

Considering the equal distribution of both sexes in the partitioned data set, a feature vector x_i of a cough or cough epoch can then be evaluated as follows:



XSL•FO
RenderX

Mel-frequency cepstral coefficients (n=20) were computed with 256 samples between successive frames and a 4096-point fast Fourier transform. Analogously, the zero-crossing rate was computed over frames of 4096 samples with 256 samples between successive frames. These features were then vertically concatenated, which resulted in a matrix where the first dimension contained 41 entries. Feature selection and specific parameters were determined by employing 5-fold cross-validation on the training set. Hyperparameters of the Gaussian mixture models were also determined by employing 5-fold cross-validation on the training set which resulted in 30 Gaussian distributions each for female and male classes. Further hyperparameters were the number of initializations (n=3), number of expectation-maximization iterations (n=200), and the use of diagonal-type covariance.

Performance Evaluation

For the evaluation of the performance of the different models, we reported several metrics such as sensitivity (true positive rate), specificity (true negative rate), accuracy, Matthews correlation coefficients, predictive positive value, negative predictive value, receiver operating characteristic curve, precision-recall curve, and Bland-Altman plot. These metrics are commonly used in machine learning and research in the context of clinical decision-support systems. For the segmentation of cough and cough epoch, we reported the number of false positives, true positives, and false negatives per night. These metrics are defined in Multimedia Appendix 1.

Results

Data

Participant Data

A total of 94 participants (female: 54/94, 57%; male: 40/94, 43%) were recruited for the study. Ages of the participants ranged from 18 to 89 years with a mean of 43 (SD 16) years. Fifteen of the 94 participants were excluded from the analysis; 2 participants withdrew, 3 participants were not involved in the study procedures for more than 5 days, and 10 participants had more than 5 nights of missed audio recordings. Some of the missed audio recordings were due to technical difficulties (such as the app crashed) while some were participant-related (such as the participant's smartphone had been turned off).

Data Set Partitioning

Cough Recognition

Of the 79 participants whose data were included for analysis, 15 participants were initially drawn at random to be included in the test set. From the remaining 64 participants, 12 additional participants were drawn at random and included in the validation set. Data from the remaining 52 participants comprised the training set.

Window Extraction

Of a total of 704,697 acoustic events, 30,304 were clearly classified as coughing and 0.11% (767/704,697) were discarded. A total of 2,547,187 noncough and 30,304 cough Mel spectrograms were computed yielding a 0.015 class ratio.

Thresholds

Thresholds of 0.98 and 0.95 yielded best results in terms of Matthews correlation coefficient for the single and ensemble convolutional neural networks of 84% and 91% on the validation set, respectively.

Annotator Intraclass Correlation

Two of the annotators together accounted for 90.23% of all nights. These two annotators had an intraclass correlation of 95.79% (mean absolute error: 0.44 coughs per night). We calculated the intraclass correlation based on 65 nights. The intraclass correlation was interpreted as excellent.

Evaluation Cough Recognition Classifier Performance

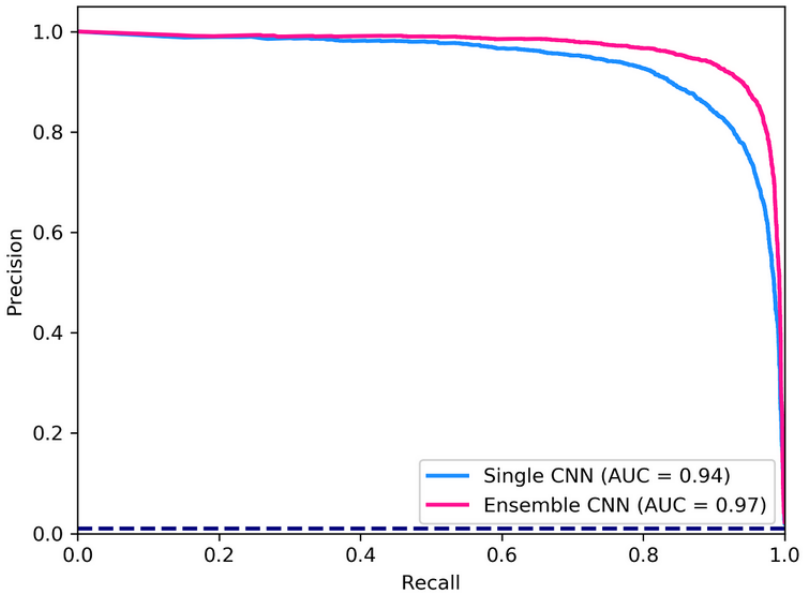
We evaluated the classifiers on the testing set, which consisted of 5489 cough and 541,972 noncough events. The test set represented the soundscape encountered in the bedrooms of 15 participants over the course of 28 nights. As shown in Table 1, the performance of the ensemble classifier was better than that of the single classifier. The difference was especially notable for true negative rate, negative predictive value, Matthews correlation coefficient values, and the area under the curve of the precision-recall curve. Both classifiers showed better performance for true positive rate compared to that for true negative rate which indicated a superior capability to recognize cough whenever a cough sound was presented compared to the capability to recognize noncough sounds. As a consequence, the false positive rate was higher than the false negative rate (Figure 3).

Table 1. Results of the convolutional neural network classifier for cough recognition.

Model type	True positive rate, %	True negative rate, %	Accuracy, %	Matthews correlation coefficient, %	Positive predictive value, %	Negative predictive value, %
Single	99.9	87.5	99.7	87.2	99.9	87.1
Ensemble	99.9	91.5	99.8	92.0	99.9	92.6

XSL•FO
RenderX

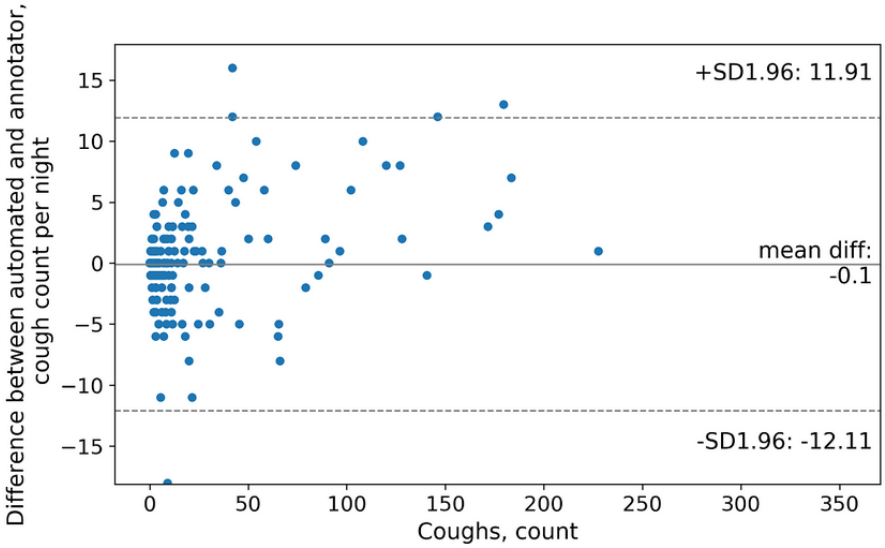
Figure 3. Precision-recall curves with the corresponding area-under-the-curve values, for the single and ensemble convolutional neural network models for the recognition of coughing. The dashed line represents the curve for a random classifier showing the proportion of cough-class instances to the total amount of instances. AUC: area under the curve; CNN: convolutional neural network.



Evaluation Cough and Cough-Epoch Segmentation

The test set included 15 participants and 421 nights; human annotators counted as few as zero and as many as 368 coughs in one night. The mean count difference between automated and annotator coughs was –0.1 (95% CI –12.11, 11.91) coughs per night (Figure 4). Further, our classifier produced a mean of 1.76 false positives per night and 1.66 false negatives per night. In total, 241 nights were identified with a count difference of 0 coughs (Figure 5) and 5 nights had count differences greater than 20 coughs. Human annotators counted as few as zero and as many as 101 cough epochs in one night. The mean count difference between automated and observer cough epochs was 0.24 (95% CI –3.67, 4.15) epochs per night (Figure 6). Our classifier produced a mean of 0.33 false positives per night and 0.57 false negatives per night. In total, 312 nights were identified with a count difference of 0 cough epochs (Figure 7) and 6 nights had count differences greater than 6 cough epochs.

Figure 4. Bland-Altman plot of the automated and annotator cough counts per night.



XSL•FO
RenderX

Figure 5. Histogram of the differences between automated and annotator cough counts per night.

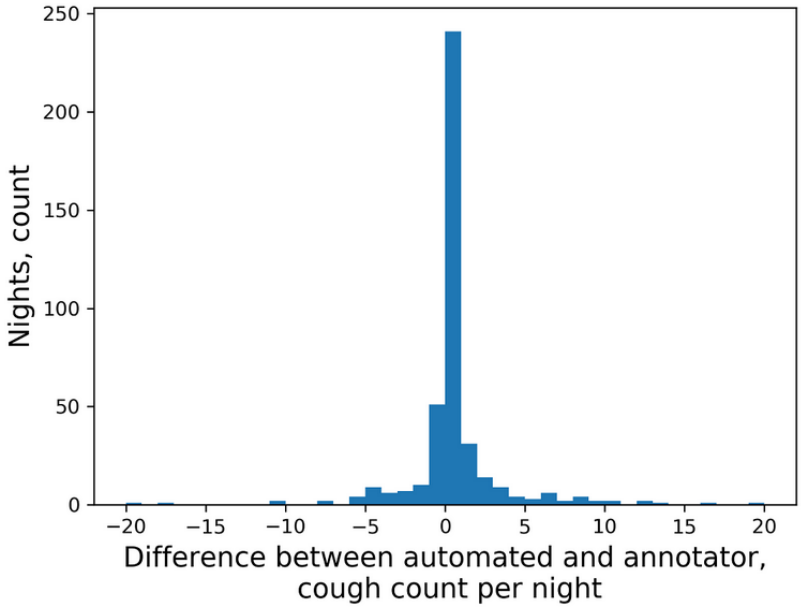
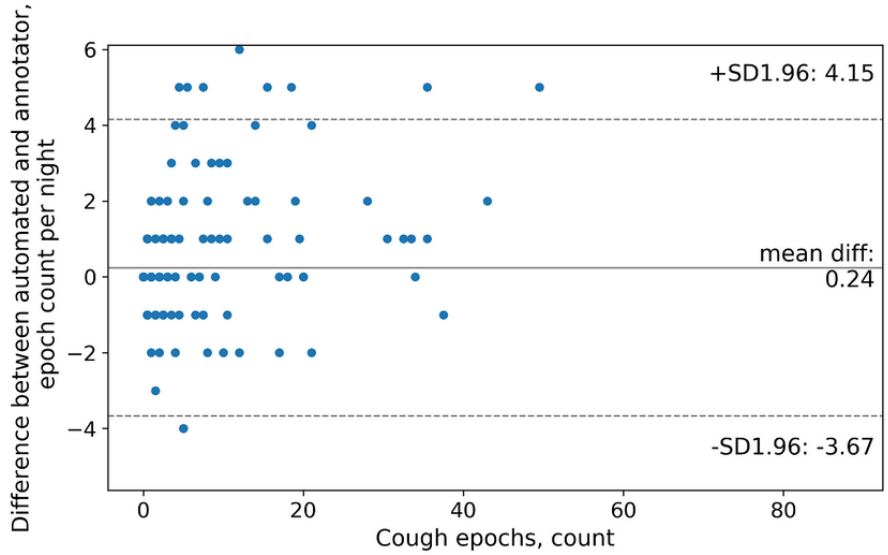
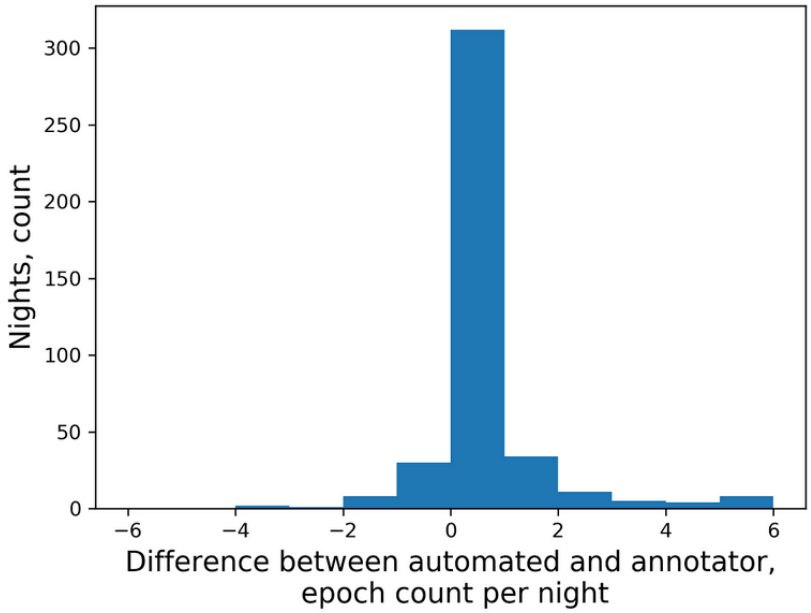


Figure 6. Bland-Altman plot of the automated and annotator cough-epoch counts per night.



XSL•FO
RenderX

Figure 7. Histogram of the differences between automated and annotator cough-epoch counts per night.



Evaluation of Sex Classifier Performance

Using the data set that included 19 female and 19 male participants which had been selected for a balanced set of male and female coughs, both extracted cough and cough-epoch signals were analyzed. The partitioning resulted in 1532 female and 1527 male coughs for training and 500 female and 498 male coughs for testing. In the case of cough epochs, this partitioning led to 366 female and 351 male cough epochs for training, and 194 male and 134 female cough epochs for testing.

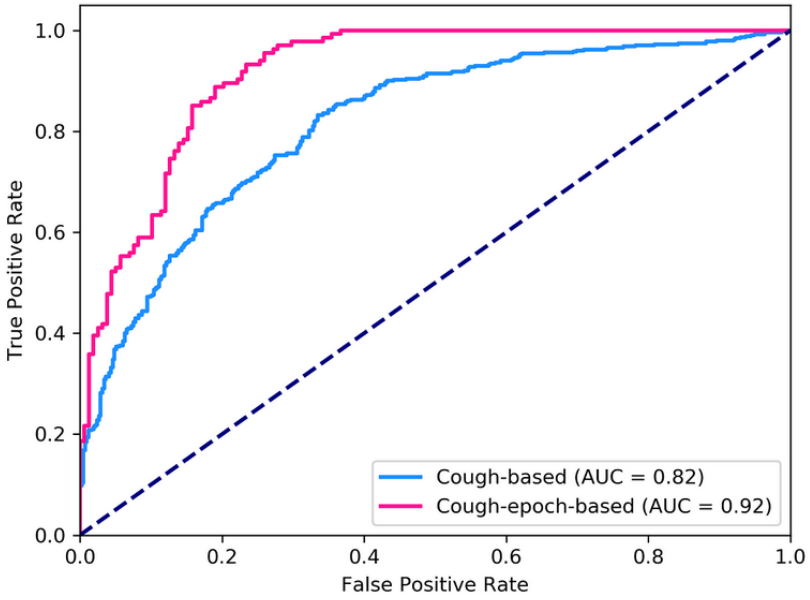
As shown in Table 2 and Figure 8, the performance of the classifier that was based on cough-epoch signals outperformed the classifier based on cough signals. The difference was especially notable on true positive rate, accuracy, Matthews correlation coefficient, positive predictive values, negative predictive values, and the area under the curve of the receiver operating characteristic curve. Both classifiers showed better performance for true positive rate than for that of true negative rate, which indicated a superior capability to recognize female cough or cough-epoch signals whenever a female cough or cough-epoch signal presented itself, in comparison to the corresponding capability to recognize a male cough or cough-epoch signal.

Table 2. Gaussian mixture model results of sex recognition for coughs and cough epochs.

Model for	True positive rate, %	True negative rate, %	Accuracy, %	Matthews correlation coefficient, %	Positive predictive value, %	Negative predictive value, %
Cough	81.0	71.8	74.8	49.6	57.8	88.8
Cough epochs	95.0	74.9	83.2	69.1	72.8	95.5

XSL•FO
RenderX

Figure 8. Receiver operating characteristic curves with corresponding area-under-the-curve values for cough and cough epoch–based sex assignment. The dashed line represents the curve for a random classifier. AUC: area under the curve; ROC: receiver operating characteristic.



Discussion

Principal Findings

To the best knowledge of the authors, the data set in this paper is the largest, real-life cough data set with published recognition and segmentation results, not only for adults with asthma but across all respiratory conditions. Given the data set of continuous overnight recordings of 79 adults with asthma in different soundscapes (excluding dropouts), our results demonstrate that cough recognition from smartphone-based audio recordings is feasible. The ensemble classifier performed well with values greater than 90% across different metrics for the pure classification task and achieved comparable cough counts to that of human annotators in the segmentation of coughing from continuous overnight audio recordings. In specific cases (for example, the 6 nights with a difference in cough counts of 20 and above), a need for further development was demonstrated. We listened to the original recording of these cases and believe these failures were caused by strong background noise, peculiar chuckle and laughter sounds, and a specific type of music, among others. These sounds, however, strongly suggest that the participant was not asleep.

We also provided a first step towards distinguishing partner cough from patient cough by determining the source of cough signals classifying those that corresponded to sex of the patient as patient coughs. This can be applied to cough recordings from the bedrooms of opposite-sex couples, even when both are coughing. Our results further indicate that cough epoch–based sex classification has greater potential than that of cough-based sex classification. This may be explained by the fact that cough epochs are longer and may contain more periodic information, rather than the limited amount of periodic information contained in the short bursts of the explosive cough sound. Speech signals of a typical adult male contain a fundamental frequency from 85 to 180 Hz and those of a typical adult female from 165 to 255 Hz [58]. This discrepancy gives rise to sex classification from speech with greater than 90% of accuracy [59]. Finally, we also investigated the automatic segmentation of cough epochs from continuous audio recordings which yielded results that were comparable to those of annotated cough epochs. These results and the fact that 86% of the coughs in our study originated from cough epochs provided a foundation for cough epoch–based sex classification. Our classifier determined the source of cough epochs based on sex with 83% of accuracy.

Comparison With Prior Work

Cough-monitoring systems that are capable of detecting reflex coughs in audio recordings have been proposed in previous work [24,25,30,60], and some of them have achieved sensitivity (true positive rate) and specificity (true negative rate) values greater than 90% [25,30,60]; however, these data sets contain coughs that have been recorded in various conditions and that have been applied in different contexts which makes a comparison with our work difficult. To our knowledge, no models or systems have been trained and evaluated on such an extensive asthmatic cough database; our data set contained 30,304 coughs. None of these systems underwent longitudinal evaluation (for more than one night). For instance, the Leicester Cough Monitor system was one of the few systems that was evaluated over a longer period of time; 6-hour and 24-hour cross-sectional recordings of patients with chronic cough [25]. Only a few approaches proposed modes of use that were comparable to our mode of use where the microphone was not attached to the patient [24,31,60]. Among those, only one involved a smartphone, where 1-hour recordings were collected for each participant in a laboratory setting [31,61]. None of these approaches addressed the problem of distinguishing the participant's cough from the coughs of other people in contact-free recordings.

Limitations

There were several limitations in our study regarding the generalization of our results. First, we only used data collected

XSL•FO
RenderX

by one specific model of smartphone. It has previously been demonstrated [32] that noisy or low-quality recordings from a different device can have a detrimental effect on the performance of the classifiers. Second, the data set may limit the generalizability of our sex-classification results. For the analysis, we included data from different male and female participants, who slept alone but were recorded in different rooms, in contrast to a real scenario of a couple, where both are sleeping and coughing in the same room. Due to the amount and length of the recordings, the annotation process was extremely laborious. As a consequence, the majority of the recordings were only annotated by one annotator. This gives rise to the possibility that certain coughs were missed or wrongly annotated.

Conclusions

Our study proposed a combined approach to combat the detrimental effect of learning from highly imbalanced data sets by combining techniques such as ensemble learning, balanced minibatch training, and decision thresholding. We showed that automated methods can recognize nocturnal coughs and cough epochs in smartphone-based audio recordings. The model addressed distinguishing subject coughs from those of a bed partner in contact-free recordings by classifying cough and cough-epoch signals to the corresponding sex of the participant. This research enables smartphone-based cough monitoring of individuals and of couples of different sexes in their bedrooms. It represents a step towards passive, scalable cough monitoring for people with asthma, and thus contributes to the development of a scalable diagnostic tool for early recognition of exacerbations.

Acknowledgments

MB, MAP, PT, FB, CS-S, FR, and TK contributed to the study design. FR and CS-S were responsible for the study execution. PT and FB provided the technological support of the study execution. FB developed the study app and embedded the content created by PT and TK. The data annotation process was designed by PT and FB, and evaluated by PT. Data preprocessing, analysis, and machine learning were conducted by FB. Writing and editing of the manuscript was done by FB. Critical review and revision of the manuscript were done by DK, MB, MAP, TK, PT, CS-S, FR, and EF. FR was the principal investigator of the clinical study. This study was funded by CSS Insurance, Switzerland. The CSS insurance supported the recruitment of participants but had no role in study design, app design, data management plans, or in reviewing and approving the manuscript for publication. DK's participation in this research was funded by Dartmouth College and by the US National Institute of Drug Abuse through the Center for Technology and Behavioral Health at Dartmouth College. The views and conclusions contained in this document are those of the authors and do not necessarily represent the official policies, either expressed or implied, of the sponsors.

Conflicts of Interest

FB, PT, EF, and TK are affiliated with the Center for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at Eidgenössische Technische Hochschule Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways and has used the open-source MobileCoach platform for that purpose; however, Pathmate Technologies is not involved in the study app described in this paper.

EF and TK are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways and has used the open-source MobileCoach platform for that purpose; however, Pathmate Technologies is not involved in the study app described in this paper.

Multimedia Appendix 1
Additional information.
[DOCX File , 19 KB - jmir_v22i7e18082_app1.docx]

References

1. The Global Asthma Report 2018. Auckland, New Zealand: Global Asthma Network; 2018. URL: http://www.globalasthmareport.org/ [accessed 2020-06-03]
2. Pavord ID, Beasley R, Agusti A, Anderson GP, Bel E, Brusselle G, et al. After asthma: redefining airways diseases. Lancet 2018 Jan 27;391(10118):350-400. [doi: 10.1016/S0140-6736(17)30879-6] [Medline: 28911920]
3. Huffaker MF, Carchia M, Harris BU, Kethman WC, Murphy TE, Sakarovitch CCD, et al. Passive Nocturnal Physiologic Monitoring Enables Early Detection of Exacerbations in Children with Asthma. A Proof-of-Concept Study. Am J Respir Crit Care Med 2018 Aug 01;198(3):320-328 [FREE Full text] [doi: 10.1164/rccm.201712-2606OC] [Medline: 29688023]
4. Gibson P, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003 Jul 22;3(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]

XSL•FO
RenderX

5. Gibson PG, Powell H. Written action plans for asthma: an evidence-based review of the key components. Thorax 2004 Feb 01;59(2):94-99 [FREE Full text] [doi: 10.1136/thorax.2003.011858] [Medline: 14760143]
6. Adams R, Smith B, Ruffin R. Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax 2000 Jul;55(7):566-573 [FREE Full text] [doi: 10.1136/thorax.55.7.566] [Medline: 10856316]
7. Honkoop PJ, Taylor DR, Smith AD, Snoeck-Stroband JB, Sont JK. Early detection of asthma exacerbations by using action points in self-management plans. Eur Respir J 2013 Jan;41(1):53-59 [FREE Full text] [doi: 10.1183/09031936.00205911] [Medline: 22653768]
8. Kaya Z, Erkan F, Ozkan M, Ozkan S, Kocaman N, Ertekin BA, et al. Self-management plans for asthma control and predictors of patient compliance. J Asthma 2009 Apr;46(3):270-275. [doi: 10.1080/02770900802647565] [Medline: 19373635]
9. Harkins MS, Fiato KL, Iwamoto GK. Exhaled nitric oxide predicts asthma exacerbation. J Asthma 2004 Jun;41(4):471-476. [doi: 10.1081/jas-120033990] [Medline: 15281333]
10. Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: reliability, validity, and responsiveness in patients not previously followed by asthma specialists. J Allergy Clin Immunol 2006 Mar;117(3):549-556. [doi: 10.1016/j.jaci.2006.01.011] [Medline: 16522452]
11. Chan-Yeung M, Chang JH, Manfreda J, Ferguson A, Becker A. Changes in peak flow, symptom score, and the use of medications during acute exacerbations of asthma. Am J Respir Crit Care Med 1996 Oct;154(4 Pt 1):889-893. [doi: 10.1164/ajrccm.154.4.8887581] [Medline: 8887581]
12. Turner MO, Noertjojo K, Vedal S, Bai T, Crump S, Fitzgerald JM. Risk factors for near-fatal asthma. A case-control study in hospitalized patients with asthma. Am J Respir Crit Care Med 1998 Jun;157(6 Pt 1):1804-1809. [doi: 10.1164/ajrccm.157.6.9708092] [Medline: 9620909]
13. Global Strategy for Asthma Management and Prevention, 2019. Global Initiative for Asthma. URL: https://ginasthma.org/ [accessed 2020-06-03]
14. Tattersfield AE, Postma DS, Barnes PJ, Svensson K, Bauer CA, O'Byrne PM, et al. Exacerbations of asthma: a descriptive study of 425 severe exacerbations. The FACET International Study Group. Am J Respir Crit Care Med 1999 Aug;160(2):594-599. [doi: 10.1164/ajrccm.160.2.9811100] [Medline: 10430734]
15. Murphy KR, Meltzer EO, Blaiss MS, Nathan RA, Stoloff SW, Doherty DE. Asthma management and control in the United States: results of the 2009 Asthma Insight and Management survey. Allergy Asthma Proc 2012;33(1):54-64. [doi: 10.2500/aap.2011.32.3518] [Medline: 22309716]
16. Osman L, McKenzie L, Cairns J, Friend J, Godden D, Legge J, et al. Patient weighting of importance of asthma symptoms. Thorax 2001 Feb;56(2):138-142 [FREE Full text] [doi: 10.1136/thorax.56.2.138] [Medline: 11209103]
17. de Marco R, Marcon A, Jarvis D, Accordini S, Almar E, Bugiani M, European Community Respiratory Health Survey Therapy Group. Prognostic factors of asthma severity: a 9-year international prospective cohort study. J Allergy Clin Immunol 2006 Jun;117(6):1249-1256. [doi: 10.1016/j.jaci.2006.03.019] [Medline: 16750983]
18. Thomson NC, Chaudhuri R, Messow CM, Spears M, MacNee W, Connell M, et al. Chronic cough and sputum production are associated with worse clinical outcomes in stable asthma. Respir Med 2013 Oct;107(10):1501-1508 [FREE Full text] [doi: 10.1016/j.rmed.2013.07.017] [Medline: 23927851]
19. Marsden PA, Satia I, Ibrahim B, Woodcock A, Yates L, Donnelly I, et al. Objective Cough Frequency, Airway Inflammation, and Disease Control in Asthma. Chest 2016 Jun;149(6):1460-1466. [doi: 10.1016/j.chest.2016.02.676] [Medline: 26973014]
20. Vollmer WM, Markson LE, O'Connor E, Sanocki LL, Fitterman L, Berger M, et al. Association of asthma control with health care utilization and quality of life. Am J Respir Crit Care Med 1999 Nov;160(5 Pt 1):1647-1652. [doi: 10.1164/ajrccm.160.5.9902098] [Medline: 10556135]
21. Accordini S, Corsico AG, Braggion M, Gerbase MW, Gislason D, Gulsvik A, et al. The cost of persistent asthma in Europe: an international population-based study in adults. Int Arch Allergy Immunol 2013;160(1):93-101. [doi: 10.1159/000338998] [Medline: 22948386]
22. Fukuhara A, Saito J, Birring SS, Sato S, Uematsu M, Suzuki Y, et al. Clinical Characteristics of Cough Frequency Patterns in Patients with and without Asthma. J Allergy Clin Immunol Pract 2020 Feb;8(2):654-661. [doi: 10.1016/j.jaip.2019.08.053] [Medline: 31541769]
23. Morice A, Fontana G, Belvisi M, Birring S, Chung K, Dicpinigaitis ROTAOC, European Respiratory Society (ERS). ERS guidelines on the assessment of cough. Eur Respir J 2007 Jun;29(6):1256-1276 [FREE Full text] [doi: 10.1183/09031936.00101006] [Medline: 17540788]
24. Barry SJ, Dane AD, Morice AH, Walmsley AD. The automatic recognition and counting of cough. Cough 2006 Sep 28;2:8 [FREE Full text] [doi: 10.1186/1745-9974-2-8] [Medline: 17007636]
25. Birring S, Fleming T, Matos S, Raj A, Evans D, Pavord I. The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough. Eur Respir J 2008 May;31(5):1013-1018 [FREE Full text] [doi: 10.1183/09031936.00057407] [Medline: 18184683]
26. Wilhelm FH, Roth WT, Sackner MA. The lifeShirt. An advanced system for ambulatory measurement of respiratory and cardiac function. Behav Modif 2003 Oct;27(5):671-691. [doi: 10.1177/0145445503256321] [Medline: 14531161]

XSL•FO
RenderX

27. Fox S, Duggan M. Tracking for health. Pew Research Center's Internet & American Life Project. 2013. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_TrackingforHealth-with-appendix.pdf [accessed 2020-06-03]
28. Juen J, Cheng Q, Schatz B. A natural walking monitor for pulmonary patients using mobile phones. IEEE J Biomed Health Inform 2015 Jul;19(4):1399-1405. [doi: 10.1109/JBHI.2015.2427511] [Medline: 25935052]
29. R Schatz B. Population measurement for health systems. Digit Med 2018;1:20174.
30. Larson E, Lee T, Liu S, Rosenfeld M, Patel S. Accurate and privacy preserving cough sensing using a low-cost microphone. 2011 Presented at: Proceedings of the 13th international conference on Ubiquitous computing; 2011; Beijing, China p. 375-384. [doi: 10.1145/2030112.2030163]
31. Monge-Alvarez J, Hoyos-Barcelo C, Lesso P, Casaseca-de-la-Higuera P. Robust Detection of Audio-Cough Events Using Local Hu Moments. IEEE J Biomed Health Inform 2019 Jan;23(1):184-196. [doi: 10.1109/JBHI.2018.2800741] [Medline: 29994432]
32. Barata F, Kipfer K, Weber M, Tinschert P, Fleisch E, Kowatsch T. Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks. 2019 Presented at: 7th IEEE International Conference on Healthcare Informatics (ICHI); 2019; Xi'an, China p. 1-11. [doi: 10.1109/ICHI.2019.8904554]
33. Rahman MM, Davis D. Addressing the Class Imbalance Problem in Medical Datasets. IJMLC 2013;3(2):224-228. [doi: 10.7763/IJMLC.2013.V3.307]
34. Smith J, Woodcock A. New developments in the objective assessment of cough. Lung 2008;186 Suppl 1:S48-S54. [doi: 10.1007/s00408-007-9059-1] [Medline: 18066694]
35. Vizel E, Yigla M, Goryachev Y, Dekel E, Felis V, Levi H, et al. Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions. Cough 2010 May 27;6:3 [FREE Full text] [doi: 10.1186/1745-9974-6-3] [Medline: 20504377]
36. Drugman T, Urbain J, Dutoit T. Assessment of audio features for automatic cough detection. 2011 Presented at: 19th European Signal Processing Conference; 2011; Barcelona p. 1289-1293.
37. Amoh J, Odame K. Deep Neural Networks for Identifying Cough Sounds. IEEE Trans Biomed Circuits Syst 2016 Oct;10(5):1003-1011. [doi: 10.1109/TBCAS.2016.2598794] [Medline: 27654978]
38. Feng W, Huang W, Ren J. Class Imbalance Ensemble Learning Based on the Margin Theory. Applied Sciences 2018 May 18;8(5):815. [doi: 10.3390/app8050815]
39. Li M, Zhang T, Chen Y, Smola A. Efficient mini-batch training for stochastic optimization. 2014 Presented at: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014; New York, USA p. 661-670. [doi: 10.1145/2623330.2623612]
40. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 2018 Oct;106:249-259. [doi: 10.1016/j.neunet.2018.07.011] [Medline: 30092410]
41. Tinschert P, Rassouli F, Barata F, Steurer-Stey C, Fleisch E, Puhan MA, et al. Prevalence of nocturnal cough in asthma and its potential as a marker for asthma control (MAC) in combination with sleep quality: protocol of a smartphone-based, multicentre, longitudinal observational study with two stages. BMJ Open 2019 Jan 07;9(1):e026323 [FREE Full text] [doi: 10.1136/bmjopen-2018-026323] [Medline: 30617104]
42. Rosenberry K, Goldsmith W, Reynolds J, McKinney W, Frazer D. Gender differences in voluntary cough sound spectra demonstrated by an inverse power law analysis. 2002 Presented at: Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society; 2002; Houston, USA p. 222-223. [doi: 10.1109/iembs.2002.1134463]
43. Diamond M. Homosexuality and bisexuality in different populations. Arch Sex Behav 1993 Aug;22(4):291-310. [doi: 10.1007/BF01542119] [Medline: 8368913]
44. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. 2015 Presented at: Wireless Telecommunications Symposium (WTS); 2015; New York, USA. [doi: 10.1109/WTS.2015.7117255]
45. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. 2017 Presented at: International Conference on Design Science Research in Information System and Technology; 2017; Karlsruhe, Germany p. 485-489. [doi: 10.1007/978-3-319-59144-5_36]
46. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
47. Banko M, Brill E. Scaling to very very large corpora for natural language disambiguation. 2001 Presented at: Proceedings of the 39th annual meeting on association for computational linguistics; 2001; Toulouse, France p. 26-33. [doi: 10.3115/1073012.1073017]
48. Huzaifah M. Comparison of time-frequency representations for environmental sound classification using convolutional neural networks. arXiv preprint arXiv. 2017. URL: https://arxiv.org/abs/1706.07156 [accessed 2020-06-03]

XSL•FO
RenderX

49. Miranda I, Diacon A, Niesler T. A Comparative Study of Features for Acoustic Cough Detection Using Deep Architectures. In: Conf Proc IEEE Eng Med Biol Soc. 2019 Jul Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019; Berlin, Germany p. 2601-2605. [doi: 10.1109/EMBC.2019.8856412]
50. Zhou Z. Ensemble methods: foundations and algorithms. Boca Raton, Florida, USA: CRC press; 2012.
51. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. Nature 1986 Oct;323(6088):533-536. [doi: 10.1038/323533a0]
52. Kingma D, Ba J. Adam: A method for stochastic optimization. 2015 Presented at: 3rd International Conference for Learning Representations; May 7-9, 2015; San Diego USA.
53. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. 2010 Presented at: Proceedings of the thirteenth international conference on artificial intelligence and statistics; 13-15 May 2010; Chia Laguna Resort, Sardinia, Italy p. 249-256.
54. Hinton G, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv. 2012. URL: https://arxiv.org/abs/1207.0580 [accessed 2020-06-03]
55. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust., Speech, Signal Process 1980 Aug;28(4):357-366. [doi: 10.1109/TASSP.1980.1163420]
56. Reynolds DA, Rose RC. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Process 1995;3(1):72-83. [doi: 10.1109/89.365379]
57. Metze F, Ajmera J, Englert R, Bub U, Burkhardt F, Stegmann J. Comparison of four approaches to age and gender recognition for telephone applications. 2007 Presented at: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '07; 15-20 April 2007; Honolulu, USA. [doi: 10.1109/ICASSP.2007.367263]
58. Titze IR, Martin DW. Principles of voice production. The Journal of the Acoustical Society of America 1998 Sep;104(3):1148-1148. [doi: 10.1121/1.424266]
59. Childers DG, Wu K. Gender recognition from speech. Part II: Fine analysis. J Acoust Soc Am 1991 Oct;90(4 Pt 1):1841-1856. [doi: 10.1121/1.401664] [Medline: 1755877]
60. Amrulloh YA, Abeyratne UR, Swarnkar V, Triasih R, Setyati A. Automatic cough segmentation from non-contact sound recordings in pediatric wards. Biomedical Signal Processing and Control 2015 Aug;21:126-136. [doi: 10.1016/j.bspc.2015.05.001]
61. Monge PR, Contractor NS. Theories of Communication Networks. New York, USA: Oxford University Press; 2003.

Edited by G Eysenbach; submitted 03.02.20; peer-reviewed by L Zhang, G Lim; comments to author 03.03.20; revised version received 27.04.20; accepted 30.04.20; published 14.07.20.

Please cite as:
Barata F, Tinschert P, Rassouli F, Steurer-Stey C, Fleisch E, Puhan MA, Brutsche M, Kotz D, Kowatsch T
Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study
J Med Internet Res 2020;22(7):e18082
URL: https://www.jmir.org/2020/7/e18082
doi:10.2196/18082
PMID:32459641

©Filipe Barata, Peter Tinschert, Frank Rassouli, Claudia Steurer-Stey, Elgar Fleisch, Milo Alan Puhan, Martin Brutsche, David Kotz, Tobias Kowatsch. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Artificial Intelligence in Health Care: Bibliometric Analysis

Yuqi Guo[1], PhD; Zhichao Hao[2], MSW; Shichong Zhao[3], BSc; Jiaqi Gong[4], PhD; Fan Yang[3], PhD

[1]School of Social Work, University of North Carolina at Charlotte, Charlotte, NC, United States

[2]School of Social Work, The University of Alabama, Tuscaloosa, AL, United States

[3]Social Welfare Program, School of Public Administration, Dongbei University of Finance and Economics, Dalian, China

[4]Department of Information Systems, University of Maryland, Baltimore, MD, United States

Corresponding Author:
Fan Yang, PhD
Social Welfare Program
School of Public Administration
Dongbei University of Finance and Economics
217 Jianshan Street
Shahekou District
Dalian
China
Phone: 86 411 84710562
Email: fyang10@dufe.edu.cn

Abstract

Background: As a critical driving power to promote health care, the health care–related artificial intelligence (AI) literature is growing rapidly.

Objective: The purpose of this analysis is to provide a dynamic and longitudinal bibliometric analysis of health care–related AI publications.

Methods: The Web of Science (Clarivate PLC) was searched to retrieve all existing and highly cited AI-related health care research papers published in English up to December 2019. Based on bibliometric indicators, a search strategy was developed to screen the title for eligibility, using the abstract and full text where needed. The growth rate of publications, characteristics of research activities, publication patterns, and research hotspot tendencies were computed using the HistCite software.

Results: The search identified 5235 hits, of which 1473 publications were included in the analyses. Publication output increased an average of 17.02% per year since 1995, but the growth rate of research papers significantly increased to 45.15% from 2014 to 2019. The major health problems studied in AI research are cancer, depression, Alzheimer disease, heart failure, and diabetes. Artificial neural networks, support vector machines, and convolutional neural networks have the highest impact on health care. Nucleosides, convolutional neural networks, and tumor markers have remained research hotspots through 2019.

Conclusions: This analysis provides a comprehensive overview of the AI-related research conducted in the field of health care, which helps researchers, policy makers, and practitioners better understand the development of health care–related AI research and possible practice implications. Future AI research should be dedicated to filling in the gaps between AI health care research and clinical applications.

(J Med Internet Res 2020;22(7):e18228) doi:10.2196/18228

KEYWORDS

health care; artificial intelligence; bibliometric analysis; telehealth; neural networks; machine learning

Introduction

From its birth in the 1950s to present, artificial intelligence (AI) and its application in modern health care have boomed with the advancement of science and technology [1-3]. As a critical driving power that promotes the coming and development of industry 4.0, AI has become an indispensable component of the advancement and innovation of health care and medical diagnosis. Medical AI technologies provide algorithms and programs that analyze symbolic models of diseases and their relationships to patient signs and symptoms [4-6].

In the field of health care, the implementation of AI technologies fosters the prediction, diagnosis, and treatment of diseases, which benefits both patients and health care providers [7]. The

XSL•FO
RenderX

promise of improving diagnostic accuracy is one of AI's most exciting health care applications. AI can effectively assist health care providers in diagnosing symptoms at a faster rate than most medical professionals [8]. AI can mimic the predictive power of human doctors to improve the accuracy of diagnosis by horizontally and vertically assessing the electronic health records of patients in a short period of time [9]. A meta-analysis reported that the diagnostic sensitivity of AI was higher than that of dermoscopy (91% vs 88%) [10]. Additionally, AI can help patients keep track of complex symptoms, improve patients' quality of life, and increase medication adherence [11]. For example, a randomized clinical trial study indicated that an AI platform successfully increased medication adherence in stroke patients on anticoagulation therapy by 50% [12].

The widespread application of AI in health care advances the processing of related research. The body of relevant literature grows rapidly. As a result, health care–related AI studies are thriving in recent health care literature. Although health care–related AI research has gained popularity, only a few bibliometric analyses focus on AI applications in specific types of health problems, such as depression [13]. A bibliometric analysis of general health care–related AI studies can depict a map that helps researchers better understand the development of health care–related AI research and the direction of patterns and trends in the future. Keeping abreast of the fast-growing body of health care–related AI studies helps practitioners and policy makers to seize the opportunities of applying AI interventions to promote the well-being of patients and their caregivers.

Bibliometrics is a measurable informatic method that analyzes the emerging trends and the knowledge structure within a certain field to obtain quantifiable, reproducible, and objective data [11]. Bibliometric analysis provides researchers and related stakeholders the opportunity to gain an informative understanding of the field of study and promotes interdisciplinary collaboration [14]. This study aims to provide a holistic view of health care–related AI research and the directions of future work to benefit patients and health care providers. Through an extensive and global review of literature on AI in health care, the purpose of this analysis is to examine the AI research focused on promoting health care. This analysis is an indispensable resource for researchers to have as an overview of the AI field in health care, which will help them to develop health care–related AI studies. This analysis is also an essential resource for someone who is less familiar with this field but is interested in AI applications in health care.

Methods

Search Strategy

Based on bibliometric indicators, we developed a search strategy in an iterative manner, starting from search terms used in the literature already known to us. This bibliometric analysis provided critical insights into the current state of health care–related AI research up to September 2019. The Web of Science (WoS) (Clarivate PLC) Core Collection was used to search all existing and highly cited AI publications. We opted for using the Science Citation Index (SCI) and the Social Science Citation (SSCI) databases in Web of Science, and then we conducted the temporal and spatial analysis, analysis of word co-occurrence, coauthorship analysis, and cocountry analysis.

Search keywords related to (1) AI technologies and (2) health care and medicine were identified from a preliminary literature review and consultation with a librarian. We entered the retrieval search string by combing keywords with Boolean operators: TI=((artificial intelligence) OR TI=("data learning") OR TI=("machine learning") OR TI=("expert systems") OR TI=("fuzzy logic") OR TI=("computer vision") OR TI=("automatic programming") OR TI=("speech understanding") OR TI=("autonomous robots") OR TI=("intelligent tutoring") OR TI=("intelligent agents") OR TI=("neural network") OR TI=("voice recognition") OR TI=("text mining") OR TI=("electronic health record")) AND (TS=(health) OR TS=(healthcare) OR TS=(medicine) OR TS=(mental health) OR TS=(behavior health)), Indexes=SCI-EXPANDED, SSCI Timespan=All years. The final search was conducted on September 09, 2019, in the WoS. A total of 5235 papers that were registered between January 1995 and September 2019 in the SCI Expanded and the SSCI Index databases from the Web of Science.

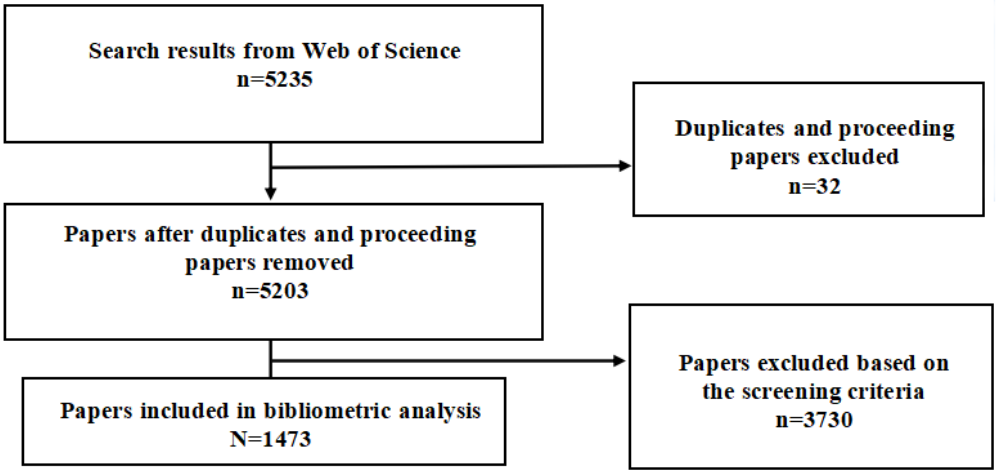
Screening Strategy

In this analysis, all journal papers about AI in health care were included for screening. The papers for analysis were restricted to those that (1) were written in the English language, (2) focused on promoting health or health care, and (3) involved AI technologies. As AI technology is a leading-edge and rapid update research area, papers published in peer-reviewed journals, conference proceedings, and early access articles were included. Book chapters and books were excluded from this bibliometric analysis.

The coauthors received training of bibliometric analysis screening by watching the video of Müller's study screening guide [15]. The screening procedure was conducted based on the screening guide [15]. Two coauthors (YG and ZH) conducted a pilot screening with the same set of 50 papers. During the pilot screening process, papers from the preliminary searches were categorized into include, exclude, or unsure. Papers marked as unsure were screened by 3 of the authors (YG, ZH, and FY) and discussed until consensus was reached in team meetings. The practical inclusion and exclusion criteria were determined. After that, 2 coauthors (ZH and YG) each received a unique set of papers for title and abstract screening. Different from a systematic review, a bibliometric analysis only requires screening of the abstract and full text when it is necessary. First, 32 inapplicable and duplicate papers were removed. According to the screening criteria, 3730 papers were excluded either because they did not focus on promoting health care or did not involve AI technologies. Finally, 1473 papers were included for bibliometric analysis (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart detailing the paper collection and screening process.



Bibliometric Analysis

Bibliometrics is the quantitative study of literature and a measurable method used to identify the developmental trends within a certain field to obtain quantifiable, reproducible, and objective data. In this analysis, we computed the growth rate of publications, characteristics of research activities (topics and keywords), publication patterns (countries and journals), and research hotspot tendencies (citation bursts and timeline map).

Growth Rate of Publications

The growth rate of publications over time was computed by raising the rate of the number of publications in 2019 over the number of publications in 1995 to the power of 1/24, as shown below. The publication trends of the number of publications each year were also reported.

Growth rate = ((number of publications in the last year ÷ number of publications in the first year)$^{1/(\text{last year - first year})}$ - 1) × 100

Publication Patterns

Citation trends of the top 10 countries, top 10 journals, and top 10 research domains were analyzed for the publication patterns. Frequencies and percentages of publications in each journal and each country were computed based on the publication years. The countries were identified by the affiliations of the listed authors. This information was provided by WoS. The rank of research impact of each country and each journal was provided based on the citation rates.

Characteristics of Research Activities

The characteristics of research activities were analyzed according to the topics and the keywords of publications. Top topics, including health problems, AI technology, function, and population, were identified and described by frequency, percentage, and citation rate of keywords listed by the authors. The centrality of a keyword was a combination of statistic equations that measure the representativeness of selected words for the text content according to betweenness, closeness, degree, eigenvector, PageRank (Google LLC), eccentricity, coreness, clustering coefficient, and term frequency scores [16]. The centralities of keywords were computed using HistCite. .

Research Hotspot Tendencies

Citation bursts and a timeline map were developed using the HistCite software. The cluster view was generated based on publications between 1995 and 2019, and each cluster was labeled by the keywords used by the paper. The time slice was set as 1 year and the threshold interpolation citations, cocitations, and cocitation coefficient were set as 4, 1, and 20, respectively. A minimum tree calculation formula was adopted to tailor the network.

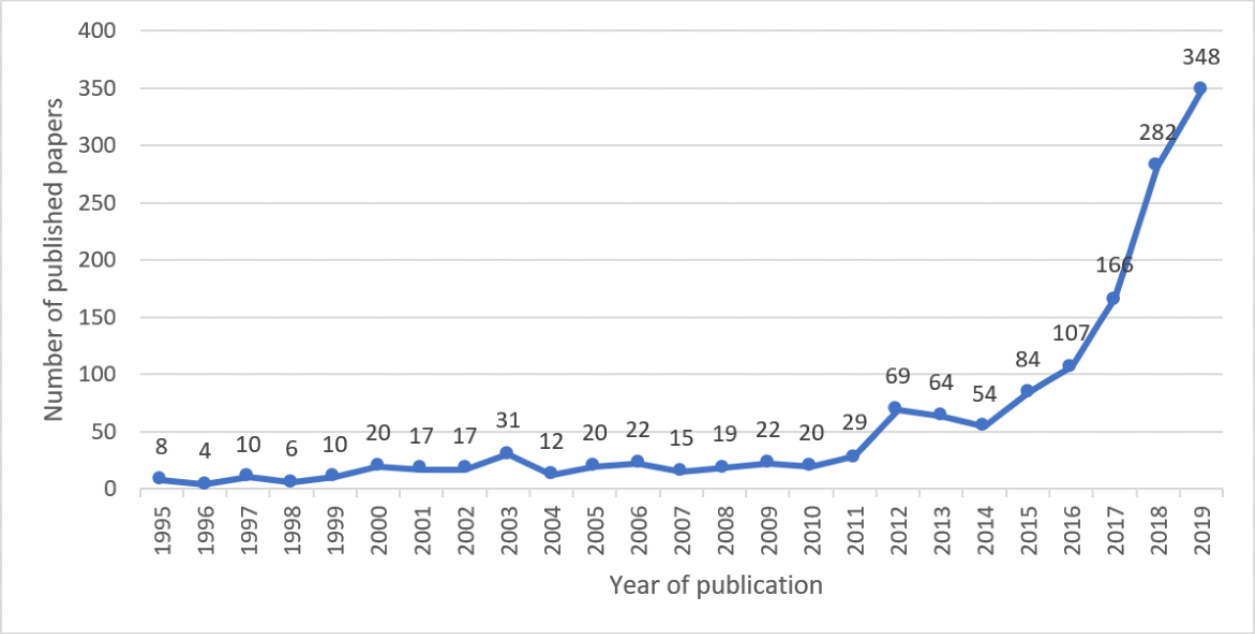
Results

Growth Rate of Publications

Figure 2 plots the annual trends of publications about AI in health care. From 1995 to 2019, the average growth rate of scientific research papers on health care–related AI research was 17.02%. The growth rate from 1995 to 2010 was 6.33%, the growth rate from 2011 to 2014 was 23.02%, and the growth rate from 2015 to 2019 was 42.67%. The number of publications increased steeply between 2014 and 2019, accounting for 70.67% (1041/1473) of all included papers.

XSL•FO
RenderX

Figure 2. The distribution of the bibliographic records per year.



Publication Patterns

Overall, 95.59% (1408/1473) of the included papers were published from 10 countries, shown in Table 1. About 45.42% (669/1473) of the included studied were from the United States. China was the next leading country (183/1473, 12.42%), followed by England (113/1473, 7.67%). The papers were published in 715 different journals. As demonstrated in Table 2, *PLOS One* published the most papers (57/1473, 3.87%), followed by *Artificial Intelligence in Medicine* (24/1473, 1.63%), and *Expert Systems with Applications* (1.63%, 24/1473). WoS subject categories were employed to indicate the research domains of included publications, as seen in Table 3. The research domains of computer science (252/1473, 17.11%), engineering (192/1473, 13.03%), and medical informatics (169/1473, 11.47%) are the top research domains of health care–related AI research.

Table 1. The distribution of the bibliographic records by top 10 (by quantity) countries.

Countries	Ranking based on total output	Output[a], n (%)	Ranking based on citations	Citations[b], n (%)
United States	1	669 (47.79)	1	10,794 (51.11)
China	2	183 (13.07)	2	2568 (12.16)
England	3	113 (8.07)	3	1969 (9.32)
India	4	82 (5.86)	9	598 (2.83)
Italy	5	74 (5.29)	5	924 (4.37)
Germany	6	73 (5.21)	4	1462 (6.92)
Canada	7	63 (4.50)	6	823 (3.90)
Japan	8	49 (3.50)	6	823 (3.90)
Spain	9	48 (3.43)	8	724 (3.43)
Iran	10	46 (3.29)	10	438 (2.07)

[a]N=1400.

[b]N=21,123.

XSL•FO
RenderX

Table 2. The distribution of the bibliographic records by top 10 (by quantity) journals.

Journals	Ranking based on total output	Output[a], n (%)	Ranking based on citations	Citations[b], n (%)
PLOS One	1	57 (23.55)	5	395 (9.85)
Artificial Intelligence in Medicine	2	24 (9.92)	1	923 (23.01)
Expert Systems with Applications	3	24 (9.92)	3	566 (14.11)
Scientific Reports	4	24 (9.92)	10	71 (1.77)
Journal of the American Medical Informatics Association	5	22 (9.09)	4	426 (10.62)
Journal of Medical Systems	6	21 (8.68)	7	262 (6.53)
Computer Methods and Programs in Biomedicine	7	20 (8.26)	8	222 (5.53)
Medical Physics	8	19 (7.85)	2	689 (17.17)
Journal of Biomedical Informatics	9	16 (6.61)	6	319 (7.95)
Computers in Biology and Medicine	10	15 (6.20)	9	139 (3.46)

[a]N=242.

[b]N=4012.

Table 3. The distribution of the bibliographic records by top 10 (by quantity) research domains.

Research domains	Ranking based on total output	Output[a], n (%)	Ranking based on citations	Citations[b], n (%)
Computer science	1	252 (18.42)	1	15,706 (21.01)
Engineering	2	192 (14.04)	6	5468 (7.32)
Medical informatics	3	169 (12.35)	8	4893 (6.55)
Oncology	4	153 (11.18)	2	11,467 (15.34)
Radiology, nuclear medicine, and medical imaging	5	142 (10.38)	4	6989 (9.35)
Health care sciences services	6	132 (9.65)	5	6729 (9.00)
Science, technology, and other topics	7	99 (7.24)	7	5207 (6.97)
General internal medicine	8	85 (6.21)	10	2565 (3.43)
Mathematical and computational biology	9	78 (5.70)	3	10,894 (14.57)
Biochemistry and molecular biology	10	66 (4.82)	9	4831 (6.46)

[a]N=1368.

[b]N=74,749.

Characteristics of Research Activities

Keywords are the core word extractions provided by researchers in the studies. Table 4 shows information about the frequency and centrality of keywords. The top 5 health problems are cancer, depression, Alzheimer disease, heart failure, and diabetes. The top 5 AI technologies are machine learning, artificial neural networks, deep learning neural networks, electronic health records, and support vector machines. The top 5 functions are case classification, diagnosis, prediction, risk estimate, and chronic condition management. The top 5 populations focused on in health care–related AI studies are children, adults, women, men, and elderly persons.

HistCite intelligently classified the research topic into 12 clusters, labeled from 0 to 11 in Figure 3. The modularity (Q) was 0.423, which was higher than 0.3, indicating that the cluster results were significant. Cluster 0 is the largest cluster (coronary artery disease) and cluster 11 is the smallest one (diabetes mellitus). Each cluster was generated based on the number of keywords under one research domain, not the frequency of keywords.

XSL•FO
RenderX

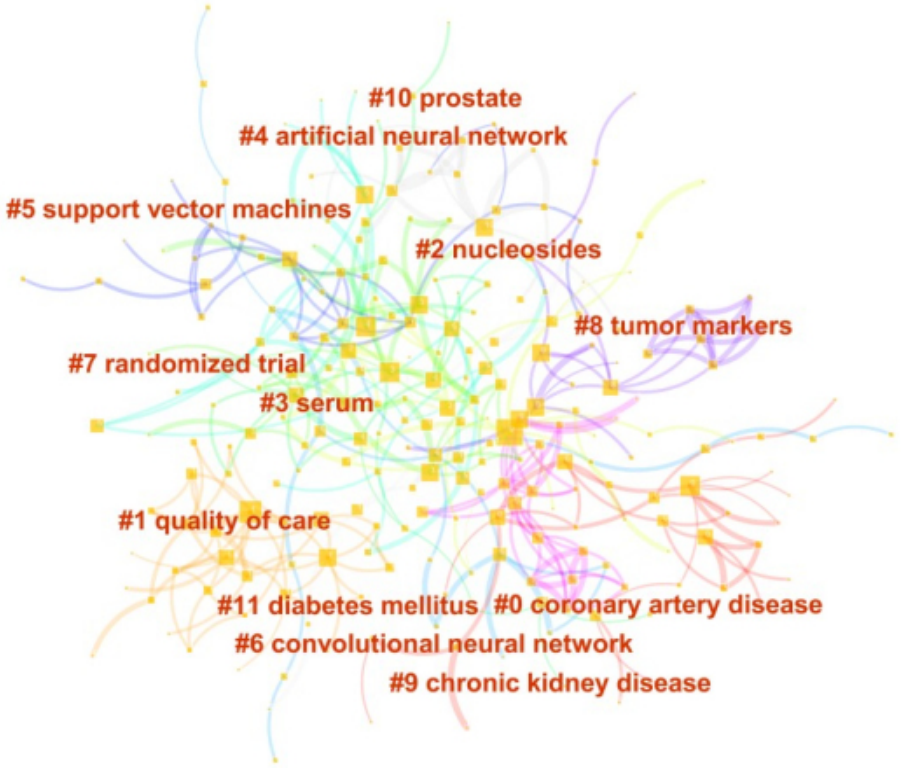
Table 4. The top keywords of artificial intelligence health care publications.

Category	Frequency (as identified by title, keywords, or manuscript)	Centrality
Health problem		
Cancer[a]	273	0.13
Depression	16	0.02
Alzheimer disease	7	0.00
Heart failure	5	0.00
Diabetes	3	0.00
Technology		
Machine learning	288	0.09
Artificial neural network	270	0.13
Deep learning neural network	95	0.01
Electronic health record	87	0.06
Support vector machine	62	0.03
Function		
Case classification	269	0.11
Diagnosis	165	0.14
Prediction	149	0.06
Risk estimate	116	0.10
Chronic condition management	71	0.02
Population		
Children	25	0.01
Adult	15	0.00
Women	11	0.00
Men	9	0.00
Elderly persons	7	0.00

[a]Breast: n=124; carcinoma: n=46; prostate: n=45; lung: n=44; other: n=14.

XSL•FO
RenderX

Figure 3. Cluster analysis of artificial intelligence health care publications.



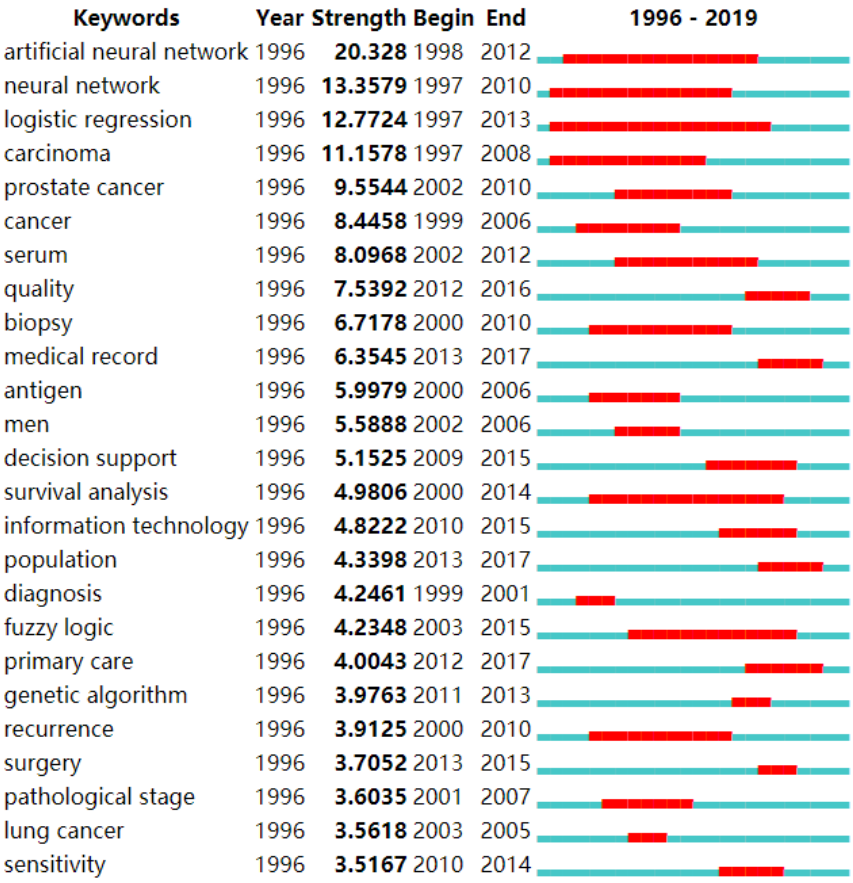
Research Hotspot Tendencies

We presented the major milestones in the development of AI in health care by analyzing the list of keywords that had strong citation bursts between 1996 and 2019, as seen in Figure 4. The first milestone keywords in the studies were neural network, logistic regression, and carcinoma. The next milestone was the artificial neural network (highly cited in 1998-2002). The most recent milestone was survival analysis, highly cited until 2014.

We also generated a timeline visualization to depict clusters along a horizontal axis, seen in Multimedia Appendix 1. The clusters we analyzed in Figure 3 were vertically listed in descending order of their sizes on the right of Multimedia Appendix 1. Each cluster was analyzed on each horizontal timeline from left to right, with the year shown on top of the view. The colored curves represent cocitation links added in the year of the corresponding color. Large-sized nodes are of particular interest because they are either highly cited, have citation bursts, or both.

XSL•FO
RenderX

Figure 4. Top 15 keywords with the strongest citation bursts.



Discussion

By using bibliometric data, this study analyzed the health care–related AI research through examining the growth rate of publications, characteristics of research activities, publication patterns, and research hotspot tendencies.

Growth Rate of Publications

Since the first publication in 1995, health care–related AI research had a slow increase in the following 17 years. However, since 2012, the field has started to move forward at a fast pace. In the past 5 years, the growth rate of publication reached 45.1%, which is more than 4 times the publication growth rate from 1995 to 2014. The recent rapid growth is due to multiple reasons. The technological breakthroughs of AI in this time period contributed to the explosive growth of AI adoption in health care–related AI research [17]. From 2015 to 2018, the appearance of autonomous robots, voice recognition, neural networks, and machine learning provided unprecedented opportunities for the prediction, diagnosis, and treatment or management of diseases [18,19]. According to the exponential growth pattern of health care–related AI research, the publications in health care–related AI will continue to grow in the future. Applying the growth rate of the past 5 years to the next 5 years, we can estimate that the publication volume in health care–related AI research will double about every two years.

Publication Patterns

The field of AI health care has attracted people from all around the world, but the high-income countries are the main force in health care–related AI research. The United States itself contributed about half of the research in the field of health care–related AI research. Compared with the rapid advancement of health care–related AI research, this research field in non–high-income countries is still meager. This observation of the small outputs in middle-income and low-income countries causes concern because many low-income countries have limited health care resources, but their public health issues are nonetheless increasing dramatically due to rapid globalization and urbanization [20]. Some developed countries have launched federal AI health care policies incorporating guidance on the development and regulation of AI in health care, such as the UK Code of Conduct for Data-Driven Health and Care Technology [21]. Consequently, about 85% of the research outputs were generated in developed countries, but 80% of the world population lives in developing countries. Multiple barriers, including funding, prioritization, research capacity, infrastructure, and language contribute to these disproportionate results [22]. From a global perspective, AI technologies are promising in terms of promoting health outcomes in low-income countries with limited health care resources.

Based on the output and citation counts, health care–related AI research is generally favored by large-scale journals related to health care. However, the sizes of the research effect are

XSL•FO
RenderX

field-specific. The improvement and sophistication of the fields of computer science and engineering have paved the way for the development of AI. It is worth noting that the field of oncology received the second most citations. These results are consistent with the findings of the top keywords of included papers, implying that published AI papers on cancer care accumulated a high number of citations.

Characteristics of Research Activities

AI is redefining and disrupting the way health care is being carried out across diverse levels, based on the results of the characteristics of AI publications. According to the top keywords in the identified categories, the top domains of disease in AI research are cancer, coronary artery disease, chronic kidney disease, prostate cancer, and diabetes mellitus. The diseases of the leading causes of death received the major attention of AI research. By 2030, chronic diseases will contribute 80% of the human deaths globally and result in severe global burden of disease [23-25]. Specifically, the cancer mortality rate has steadily risen by 6% during the past 10 years [26]. Therefore, researchers direct their hopes and efforts on early detection and condition management by using advanced technology [27,28]. AI is making its way into cancer treatment from diagnostic classification to tumor behavior prediction [29,30]. Furthermore, cardiovascular diseases are also highly related to behaviors, especially among older adults [31,32]. AI can exert a proactive role in predicting risk factors for cardiovascular diseases related to older adults' behaviors in order to remarkably reduce the hospitalization rate, readmission rate, cost of care, and even mortality rate [33-35].

Extensive functions of AI have assisted modern health care by providing smart medical data analysis and developing accurate and efficient prediction for treatment [36]. Machine learning is the most commonly used AI technique in health care. Artificial neural networks are the next biggest research area related to health care. There are promising applications of these AI technologies to the development of interventions for other diseases in future research, such as mental health interventions, health education, and chronic disease management.

AI research is witnessing widespread adoption in the prediction, detection, diagnosis, classification, treatment, and survival prediction of diseases [30,37]. The most common application of AI technologies is reflected in the domains of medical classification and quality of care. AI is investing significantly in improving the quality of care in the health care system [38]. The potential of AI in reviewing medical images and analyzing large-scale data has led to significant improvement in the quality of care [39]. Additionally, AI-based risk prediction models can investigate the complex relationships between clinical data and disease treatment [40].

Regarding the research populations, researchers seem more interested in child populations. Due to the development of cognitive aids to support diagnosis, treatment, care coordination, surveillance and prevention, and health maintenance, improvement of AI in clinical pediatric health care is remarkable [41]. However, cancer and various chronic diseases are the main focus of the current health care–related AI publications. Considering that elderly populations are the primary sufferers of chronic diseases, future studies may shift their angle to further explore AI implications on health in older adult populations.

Research Hotspot Tendencies

In the last three decades, the times and lengths of citation bursts of each health care–related AI research topic have varied. Certain keywords were extraordinary consistent for a long period of time, while some keywords only briefly surged in the field of AI health care. The period of 1997 to 2014 contained a concentrated outbreak of multiple health care–related AI study bursts (Figure 4). AI technology research involved more health care fields and produced more diverse keywords in recent years. It is possible that the more diverse keywords of health care–related AI research diluted the citation bursts.

According to the citation rates, artificial neural networks, support vector machines, and convolutional neural networks have the highest impact on health care. These techniques have been widely adopted by oncology research to predict disease and survival rates [42]. Historically, logistic analysis has been a widely used research method of health care–related AI. However, survival analysis has had a recent citation burst in this field.

The studies in these clusters focus on testing AI techniques and translating these techniques into practical settings [37,43]. The next step of health care–related AI research may transform from lab-based research to the development of clinically validated and safe regulated systems. The current health care–related AI research faces the challenges of deploying lab-based AI intervention into clinical practice [44]. These challenges include ethical issues brought up by AI application [45] and the quality and quantity of medical data and cases [46], which will decide the diagnosis and treatment ability of AI.

Health care–related AI research experienced several paradigm shifts during the past decades and continues to shift. As shown in the timeline overview (Multimedia Appendix 1), the sustainability of health care–related AI research clusters is influenced by the increasing availability of health care data and the rapid progression of AI techniques. The research of cancer, convolutional neural networks, and nucleosides sustained a hotspot over the past twenty years, whereas some clusters were relatively short-lived, such as chronic kidney disease, prostate cancer, and diabetes mellitus. Particularly, nucleosides, convolutional neural networks, and tumor markers have remained research hotspots through 2019. These research domains exert strong impacts on the field and these analysis results suggest that their influence will likely continue in the next few years.

Limitations

Limitations of our work need to be acknowledged. First, although we are confident that a single database—WoS—is a large enough database to offer a wide variety of publications vital for our analyses, future studies will apply other databases, such as Scopus, to explore more potential papers. Second, the search keywords (eg, mental health, behavioral health, and health care) related to health care were quite general, which may not be able to identify AI-related studies in all aspects of health care. For example, AI-related studies of HIV might have

XSL•FO
RenderX

been omitted. Third, some keywords, although they were ranked as top keywords, were uninformative by themselves (eg, risk, model, and system) and could not be analyzed. Fourth, we did not include gray literature (eg, books), and we did not include papers published in languages other than English. Therefore, we may have missed relevant studies conducted in different forms, languages, and countries. Future studies can broaden the search scope to explore more relevant research to enrich the literature.

Conclusions

We aimed to provide a bird's-eye view of the entirety of the health care–related AI research. This analysis provides a comprehensive overview of the AI-related research conducted in the field of health care. With multiple searching and screening rounds, 21 finalized search terms, and a 25-year timespan from 1995 to 2019, we are confident that we have identified inclusive health care–related AI studies.

This analysis also depicted research trends of AI-related health research: (1) the growth rate of health care–related AI publications has grown rapidly in the past decade and the rate showed a trend of continuous growth; (2) high-income countries are the main force of health care–related AI research; (3) most AI research was focused on chronic diseases, particularly on cancer; (4) machine learning and neural networks are the most commonly used AI techniques in classification, diagnosis, and prediction; and (5) the research domains of nucleosides, convolutional neural networks, and tumor markers are currently research hotspots of health care–related AI research. AI research on health care is accelerating rapidly, with potential applications being demonstrated across various domains of medicine. However, there are currently limited examples of such techniques being successfully deployed into clinical practice. Future AI research should be dedicated to filling the gap between health care–related AI research and clinical applications.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The keywords timeline view of artificial intelligence health care publications.
[PNG File , 231 KB - jmir_v22i7e18228_app1.png]

References

1. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 2018 Feb 22;172(5):1122-1131.e9 [FREE Full text] [doi: 10.1016/j.cell.2018.02.010] [Medline: 29474911]
2. Lomazzo J. Fintech's Impact on Wealth Management. Ivey Business School. 2016. URL: https://www.ivey.uwo.ca/cmsmedia/3775893/ifp-presentation-for-scotia-digital-banking-lab-dec-20-2016-lomazzo-j.pdf [accessed 2020-06-22]
3. McFarland M. Google's artificial intelligence breakthrough may have a huge impact on self-driving cars and much more. Washington Post. URL: https://www.washingtonpost.com/news/innovations/wp/2015/02/25/googles-artificial-intelligence-breakthrough-may-have-a-huge-impact-on-self-driving-cars-and-much-more/ [accessed 2020-02-15]
4. Koh H, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005;19(2):64-72. [Medline: 15869215]
5. Wimmer H, Yoon VY, Sugumaran V. A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy. Decision Support Systems 2016 Aug;88:51-66. [doi: 10.1016/j.dss.2016.05.008]
6. Ramesh A, Kambhampati C, Monson J, Drew P. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004 Sep;86(5):334-338 [FREE Full text] [doi: 10.1308/147870804290] [Medline: 15333167]
7. Reddy S. Use of Artificial Intelligence in Healthcare Delivery. EHealth-Making Health Care Smarter. IntechOpen 2018:E. [doi: 10.5772/intechopen.74714]
8. Saifi S, Taylor AJ, Allen J, Hendel R. The use of a learning community and online evaluation of utilization for SPECT myocardial perfusion imaging. JACC Cardiovasc Imaging 2013 Jul;6(7):823-829 [FREE Full text] [doi: 10.1016/j.jcmg.2013.01.012] [Medline: 23643281]
9. Choi E, Bahadori M, Schuetz A, Stewart W, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
10. Rajpara S, Botello A, Townend J, Ormerod A. Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma. Br J Dermatol 2009 Sep;161(3):591-604. [doi: 10.1111/j.1365-2133.2009.09093.x] [Medline: 19302072]
11. Kreps GL, Neuhauser L. Artificial intelligence and immediacy: designing health communication to personally engage consumers and providers. Patient Educ Couns 2013 Aug;92(2):205-210. [doi: 10.1016/j.pec.2013.04.014] [Medline: 23683341]
12. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using Artificial Intelligence to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. Stroke 2017 May;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]

XSL•FO
RenderX

13. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HLT, et al. The Current Research Landscape on the Artificial Intelligence Application in the Management of Depressive Disorders: A Bibliometric Analysis. Int J Environ Res Public Health 2019 Jun 18;16(12):2150 [FREE Full text] [doi: 10.3390/ijerph16122150] [Medline: 31216619]
14. Niu B, Hong S, Yuan J, Peng S, Wang Z, Zhang X. Global trends in sediment-related research in earth science during 1992–2011: a bibliometric analysis. Scientometrics 2013 Jun 22;98(1):511-529. [doi: 10.1007/s11192-013-1065-x]
15. Study screening guide. Youtube. 2020 Mar 01. URL: https://www.youtube.com/watch?v=7p6w0zaTGH [accessed 2020-03-01]
16. Vega-Oliveros DA, Gomes PS, E. Milios E, Berton L. A multi-centrality index for graph-based keyword extraction. Information Processing & Management 2019 Nov;56(6):102063. [doi: 10.1016/j.ipm.2019.102063]
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
18. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg 2018 Jul;268(1):70-76 [FREE Full text] [doi: 10.1097/SLA.0000000000002693] [Medline: 29389679]
19. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
20. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: A systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
21. Morley J, Floridi L. NHS AI Lab: Why We Need to Be Ethically Mindful About AI for Healthcare. SSRN Journal 2019 Jan 01:1-1. [doi: 10.2139/ssrn.3445421]
22. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348 [FREE Full text] [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
23. Nuño R, Coleman K, Bengoa R, Sauto R. Integrated care for chronic conditions: the contribution of the ICCC Framework. Health Policy 2012 Apr;105(1):55-64. [doi: 10.1016/j.healthpol.2011.10.006] [Medline: 22071454]
24. Rosen JM, Kun L, Mosher RE, Grigg E, Merrell RC, Macedonia C, et al. Cybercare 2.0: meeting the challenge of the global burden of disease in 2030. Health Technol (Berl) 2016;6:35-51 [FREE Full text] [doi: 10.1007/s12553-016-0132-8] [Medline: 27358760]
25. Samb B, Desai N, Nishtar S, Mendis S, Bekedam H, Wright A, et al. Prevention and management of chronic disease: a litmus test for health-systems strengthening in low-income and middle-income countries. Lancet 2010 Nov 20;376(9754):1785-1797. [doi: 10.1016/S0140-6736(10)61353-0] [Medline: 21074253]
26. ReFaey K, Tripathi S, Grewal SS, Quinones DJ, Chaichana KL, Antwi SO, et al. Current Leading Cause of Death in the World: Cancer versus Cardiovascular Disease. SSRN Journal 2019 Jan 01:1-1. [doi: 10.2139/ssrn.3396025]
27. Houssami N, Kirkpatrick-Jones G, Noguchi N, Lee CI. Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert Rev Med Devices 2019 May;16(5):351-362. [doi: 10.1080/17434440.2019.1610387] [Medline: 30999781]
28. Martinez-Millana A, Zettl A, Floch J, Calvo-Lerma J, Sevillano JL, Ribes-Koninckx C, et al. The Potential of Self-Management mHealth for Pediatric Cystic Fibrosis: Mixed-Methods Study for Health Care and App Assessment. JMIR Mhealth Uhealth 2019 Apr 18;7(4):e13362 [FREE Full text] [doi: 10.2196/13362] [Medline: 30998222]
29. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018 Jul;21(4):653-660. [doi: 10.1007/s10120-018-0793-2] [Medline: 29335825]
30. Leatherdale ST, Lee J. Artificial intelligence (AI) and cancer prevention: the potential application of AI in cancer control programming needs to be explored in population laboratories such as COMPASS. Cancer Causes Control 2019 Jul;30(7):671-675. [doi: 10.1007/s10552-019-01182-2] [Medline: 31093860]
31. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, American Heart Association Statistics Committee Stroke Statistics Subcommittee. Heart disease and stroke statistics--2013 update: a report from the American Heart Association. Circulation 2013 Jan 01;127(1):e6-e245 [FREE Full text] [doi: 10.1161/CIR.0b013e31828124ad] [Medline: 23239837]
32. Jernigan VBB, Duran B, Ahn D, Winkleby M. Changing patterns in health behaviors and risk factors related to cardiovascular disease among American Indians and Alaska Natives. Am J Public Health 2010 Apr;100(4):677-683. [doi: 10.2105/AJPH.2009.164285] [Medline: 20220114]
33. Abdel-Motaleb I, Akula R. Artificial intelligence algorithm for heart disease diagnosis using phonocardiogram signals. 2012 Presented at: IEEE International Conference on Electro/Information Technology; May 6-8, 2012; Indianapolis, IN p. 1-6. [doi: 10.1109/eit.2012.6220714]
34. Srivastava C. Machine Learning Based Framework for Prediction of Diabetic Patient Readmission Rate in Hospitals. Trans Sci Technol. 2018. URL: http://tost.unise.org/pdfs/vol5/no3/5x3x204x209.pdf [accessed 2020-06-09]
35. Yeh J, Wu T, Tsao C. Using data mining techniques to predict hospitalization of hemodialysis patients. Decision Support Systems 2011 Jan;50(2):439-448. [doi: 10.1016/j.dss.2010.11.001]

XSL•FO
RenderX

36. Petersen CL, Weeks WB, Norin O, Weinstein JN. Development and Implementation of a Person-Centered, Technology-Enhanced Care Model For Managing Chronic Conditions: Cohort Study. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11082 [FREE Full text] [doi: 10.2196/11082] [Medline: 30892274]
37. Zhang H, Hung C, Chu W, Chiu P, Tang C. Chronic kidney disease survival prediction with artificial neural networks. 2018 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 3-6, 2018; Madrid, Spain. [doi: 10.2196/11082]
38. Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res 2018 Jul 13;18(1):545 [FREE Full text] [doi: 10.1186/s12913-018-3359-4] [Medline: 30001717]
39. Gupta A, Slater JJ, Boyne D, Mitsakakis N, Béliveau A, Druzdzel MJ, et al. Probabilistic Graphical Modeling for Estimating Risk of Coronary Artery Disease: Applications of a Flexible Machine-Learning Method. Med Decis Making 2019 Nov;39(8):1032-1044. [doi: 10.1177/0272989X19879095] [Medline: 31619130]
40. Sandstedt M, Henriksson L, Janzon M, Nyberg G, Engvall J, De Geer J, et al. Evaluation of an AI-based, automatic coronary artery calcium scoring software. Eur Radiol 2020 Mar;30(3):1671-1678. [doi: 10.1007/s00330-019-06489-x] [Medline: 31728692]
41. Shu L, Sun Y, Tan L, Shu Q, Chang AC. Application of artificial intelligence in pediatrics: past, present and future. World J Pediatr 2019 Apr;15(2):105-108. [doi: 10.1007/s12519-019-00255-1] [Medline: 30997653]
42. Burt JR, Torosdagli N, Khosravan N, RaviPrakash H, Mortazi A, Tissavirasingham F, et al. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. Br J Radiol 2018 Sep;91(1089):20170545 [FREE Full text] [doi: 10.1259/bjr.20170545] [Medline: 29565644]
43. Wong T, Xu D, Ting D. Artificial intelligence deep learning system for predicting chronic kidney disease from retinal images. Invest Ophthalmol Vis Sci 2019;60(9):1468.
44. O'Connor PJ, Sperl-Hillen JM, Rush WA, Johnson PE, Amundson GH, Asche SE, et al. Impact of electronic health record clinical decision support on diabetes care: a randomized trial. Ann Fam Med 2011;9(1):12-21 [FREE Full text] [doi: 10.1370/afm.1196] [Medline: 21242556]
45. Yu Y, Xu H, Hao J, Qiao Y. [Ethical Issues of Medical Artificial Intelligence]. Zhongguo Yi Xue Ke Xue Yuan Xue Bao 2020 Feb 28;42(1):128-131 [FREE Full text] [doi: 10.3881/j.issn.1000-503X.10961] [Medline: 32131952]
46. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford) 2020 Jan 01;2020:E [FREE Full text] [doi: 10.1093/database/baaa010] [Medline: 32185396]

Abbreviations

AI: artificial intelligence
SCI: Science Citation Index
SSCI: Social Science Citation
WoS: Web of Science

Edited by G Eysenbach; submitted 25.02.20; peer-reviewed by Z Ni, W Weeks; comments to author 11.04.20; revised version received 22.04.20; accepted 14.05.20; published 29.07.20.

Please cite as:
Guo Y, Hao Z, Zhao S, Gong J, Yang F
Artificial Intelligence in Health Care: Bibliometric Analysis
J Med Internet Res 2020;22(7):e18228
URL: http://www.jmir.org/2020/7/e18228/
doi:10.2196/18228
PMID:32723713

©Yuqi Guo, Zhichao Hao, Shichong Zhao, Jiaqi Gong, Fan Yang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis

Alaa Ali Abd-Alrazaq[1], PhD; Asma Rababeh[2], MSc; Mohannad Alajlani[3], PhD; Bridgette M Bewick[4], PhD; Mowafa Househ[1], PhD

[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]Jordan Health Aid Society International, Amman, Jordan

[3]Institute of Digital Healthcare, University of Warwick, Warwick, United Kingdom

[4]Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Mowafa Househ, PhD
College of Science and Engineering
Hamad Bin Khalifa University
Liberal Arts and Sciences Building, Education City
Ar Rayyan
Doha
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: The global shortage of mental health workers has prompted the utilization of technological advancements, such as chatbots, to meet the needs of people with mental health conditions. Chatbots are systems that are able to converse and interact with human users using spoken, written, and visual language. While numerous studies have assessed the effectiveness and safety of using chatbots in mental health, no reviews have pooled the results of those studies.

Objective: This study aimed to assess the effectiveness and safety of using chatbots to improve mental health through summarizing and pooling the results of previous studies.

Methods: A systematic review was carried out to achieve this objective. The search sources were 7 bibliographic databases (eg, MEDLINE, EMBASE, PsycINFO), the search engine "Google Scholar," and backward and forward reference list checking of the included studies and relevant reviews. Two reviewers independently selected the studies, extracted data from the included studies, and assessed the risk of bias. Data extracted from studies were synthesized using narrative and statistical methods, as appropriate.

Results: Of 1048 citations retrieved, we identified 12 studies examining the effect of using chatbots on 8 outcomes. Weak evidence demonstrated that chatbots were effective in improving depression, distress, stress, and acrophobia. In contrast, according to similar evidence, there was no statistically significant effect of using chatbots on subjective psychological wellbeing. Results were conflicting regarding the effect of chatbots on the severity of anxiety and positive and negative affect. Only two studies assessed the safety of chatbots and concluded that they are safe in mental health, as no adverse events or harms were reported.

Conclusions: Chatbots have the potential to improve mental health. However, the evidence in this review was not sufficient to definitely conclude this due to lack of evidence that their effect is clinically important, a lack of studies assessing each outcome, high risk of bias in those studies, and conflicting results for some outcomes. Further studies are required to draw solid conclusions about the effectiveness and safety of chatbots.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019141219; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019141219

(J Med Internet Res 2020;22(7):e16021) doi:10.2196/16021

KEYWORDS

chatbots; conversational agents; mental health; mental disorders; depression; anxiety; effectiveness; safety

XSL•FO
RenderX

Introduction

Background

Mental illness is a growing public health concern worldwide [1]. One in 4 adults and 1 in 10 children are likely to be affected by mental health problems annually [2]. Mental illness has a significant impact on the lives of millions of people and a profound impact on the community and economy. Mental disorders impair quality of life and are considered one of the most common causes of disability [3]. Mental disorders are predicted to cost $16 trillion globally between 2011 and 2030 due to lost labor and capital output [4].

There is a shortage of mental health human resources, poor funding, and mental health illiteracy globally [5,6]. This lack of resources is especially evident in low-income and middle-income countries where there are 0.1 psychiatrists per 1,000,000 people [7], compared to 90 psychiatrists per 1,000,000 people in high-income countries [8]. According to the World Health Organization, mental health services reach 15% and 45% of those in need in developing and developed countries, respectively [9]. This could be a major factor contributing to the increase in suicidal behavior in recent decades [10].

The demand for better mental health services has increased, and meeting these demands has become increasingly difficult and costly due to a lack of resources [4]. Therefore, new solutions are needed to compensate for the deficiency of resources and promote patient self-care [4]. Distance can impede the reach of traditional mental health services to populations in remote areas in both high-income and low-income countries. Technology-based treatment, such as mobile apps, can overcome most of these barriers and engage hard-to-reach populations [11]. A World Health Organization survey of 15,000 apps revealed that 29% focus on mental health diagnosis or support [10].

One technology that offers a partial solution to the lack of capacity within the global mental health workforce is mobile apps. They have the potential to improve the quality and accessibility of mental health [12]. Chatbots are one of the main mobile apps used for mental health [13]. Chatbots, also known as conversational agents, conversational bots, and chatterbots, are computer programs able to converse and interact with human users [5,14]. Chatbots use spoken, written, and visual languages [5,14]. The use of chatbots has grown tremendously over the last decade and has become pervasive in fields such as mental health [13]. It is expected that chatbots will make a positive contribution to addressing the shortfall of mental health care [15]. Chatbots can facilitate interactions with those who are reluctant to seek mental health advice due to stigmatization [5] and allow more conversational flexibility [16].

Research Problem and Aim

Adoption of new technology, especially those heavily related to artificial intelligence and machine learning, relies on first ascertaining the levels of safety and effectiveness [17]. There has been a steady rise in the number of studies assessing the effectiveness and safety of using chatbots for mental health [5]. There is a need to critically evaluate and statistically combine findings to inform policy and practice. Previously conducted reviews [5,18,19] did not assess the effectiveness and safety of chatbots in mental health. Accordingly, the current systematic review aimed to assess the effectiveness and safety of using chatbots in mental health through summarizing and pooling the results of previous studies. The review question is "what is the effectiveness and safety of using chatbots for improving mental health?"

Methods

Overview

A systematic review of the literature was conducted to accomplish the objective. This review is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Multimedia Appendix 1) [20]. The protocol for this systematic review is registered at PROSPERO (number CRD42019141219).

Search Strategy

Search Sources

The following bibliographic databases were searched in this review: MEDLINE, EMBASE, PsycINFO, IEEE Xplore, ACM Digital Library, Scopus, and Cochrane Central Register of Controlled Trials. The search engine "Google Scholar" was also searched. As Google Scholar retrieved a large number of studies ordered by their relevance to the search topic, we screened the first 100 hits (10 pages). The search started on the June 8, 2019 and finished on June 11, 2019. We carried out backward reference list checking, where reference lists of the included studies and reviews were screened for further studies of relevance to the review. In addition, we conducted forward reference list checking, where we used the "cited by" function available in Google Scholar to identify studies that cited the included studies.

Search Terms

Search terms in this review were related to population (eg, mental disorder, mood disorder, and anxiety disorder) and intervention (eg, conversational agent, chatbot, chatterbot, and virtual agent). The search terms were derived from previous reviews and informatics experts interested in mental health issues [13]. Further, search terms related to mental disorders were derived from the Medical Subject Headings index in MEDLINE. The search strings utilized for searching each bibliographic database are shown in Multimedia Appendix 2.

Study Eligibility Criteria

The population of interest was individuals who use chatbots for their mental health, but not physicians or caregivers who use chatbots for their patients. Eligible interventions were chatbots operating as standalone software or via a web browser. Chatbots that were integrated into robotics, serious games, SMS, or telephone systems were excluded. The current review also excluded chatbots that relied on human-operator generated dialogue. There were no restrictions regarding the type of dialogue initiative (ie use, system, mixed) and input and output modality (ie spoken, visual, written). There were no limitations related to the comparator (eg, information, waiting list, usual

XSL•FO
RenderX

care). This review focused on any outcome related to effectiveness (eg, severity or frequency of any mental disorders and psychological wellbeing) or safety (eg, adverse events, deaths, admissions to psychiatric settings) of chatbots. Regarding the study design, we included only randomized controlled trials (RCTs) and quasiexperiments. The review included peer-reviewed articles, dissertations, conference proceedings, and reports. The review excluded reviews, conference abstracts, proposals, and editorials. Only studies written in English were included in the review. There were no restrictions regarding study setting, year of publication, and country of publication.

Study Selection

Two steps were followed for selecting studies. First, the titles and abstracts of all retrieved studies were screened independently by two reviewers (AA, MA). Second, the full texts of studies included from the first step were read independently by the same reviewers. Any disagreements between the reviewers were resolved by discussion or by consulting a third reviewer (MH). Cohen κ [21] was calculated to assess interrater agreement between reviewers, which was 0.85 and 0.89 in the first and second step of the selection process, respectively, indicating a very good level of agreement [22].

Data Extraction

Before extracting data, we developed a data extraction form and piloted it using three included studies to conduct a systematic and precise extraction of data (Multimedia Appendix 3). Two reviewers (AA, MA) independently extracted data from the included studies, and disagreements were resolved by discussion or by consulting the third reviewer (MH). Interrater agreement between the reviewers was very good (Cohen κ=0.84) [22].

Assessment of Risk of Bias

Two Cochrane tools were used to assess the risk of bias in the included studies. Risk of bias in RCTs was assessed using the Risk of Bias 2 (RoB 2) tool [23], and risk of bias in quasi-experiments was examined using the Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) tool [24]. The results of the risk of bias are presented as a graph showing the reviewers' judgments about each "risk of bias" domain. Further, they are presented as a figure showing the reviewers' judgments about each "risk of bias" domain for each included study. Two reviewers (AA, AR) independently assessed the risk of bias, and disagreements were resolved by discussion or by consulting the third reviewer (MH). Interrater agreement between the reviewers was good (Cohen κ=0.75) [22].

Data Synthesis

Data extracted from studies were synthesized using narrative and statistical methods. The statistical approach was used when there was more than one RCT for a certain outcome and the study reported enough data for the analysis. Where statistical findings were not available, a narrative approach was used to synthesize the data. Findings of studies were grouped and synthesized according to the measured outcome.

Statistical analysis was carried out using Review Manager (RevMan 5.3). As all extracted data were continuous, the effect of each trial and the overall effect were measured using either the mean difference (MD) or the standardized mean difference (SMD). To be more precise, when the outcome was measured using the same method between studies, the MD was utilized. The SMD was used when, between studies, the outcome was assessed using different measurement tools. A random-effects model was used for combining results because there was clinical heterogeneity between studies in terms of population (eg, clinical versus nonclinical samples), intervention (eg, rule-based versus artificial intelligence chatbots), and comparator (eg, treatment as usual versus information).

When there was a statistically significant difference between groups, we assessed how this difference was clinically important. A minimal clinically important difference refers to the smallest change in a measured outcome that a patient would deem as worthy and significant and which mandates a change in a patient's treatment [25]. Boundaries of a minimal clinically important difference for each outcome were calculated as ±0.5 times the SD of the control arms of the studies at baseline.

Clinical heterogeneity of the meta-analyzed trials was assessed by checking their participants, interventions, comparators, and measured outcomes. Statistical heterogeneity was assessed by calculating the statistical significance of heterogeneity (chi-square *P* value) and I^2. A chi-square *P* value >.05 indicates that the studies are homogenous [26]. I^2 was used to quantify the heterogeneity of studies, where I^2 of 0%-40%, 30%-60%, 50%-90%, and 75%-100% represents unimportant, moderate, substantial, and considerable heterogeneity, respectively [26].

When the evidence was synthesized statistically, the overall quality of that evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation [17]. Two reviewers (AA & AR) assessed the quality of the evidence, and disagreements were resolved by discussion or by consulting the third reviewer (MH). There was considerable interrater agreement between the reviewers (Cohen κ= 0.86) [22].

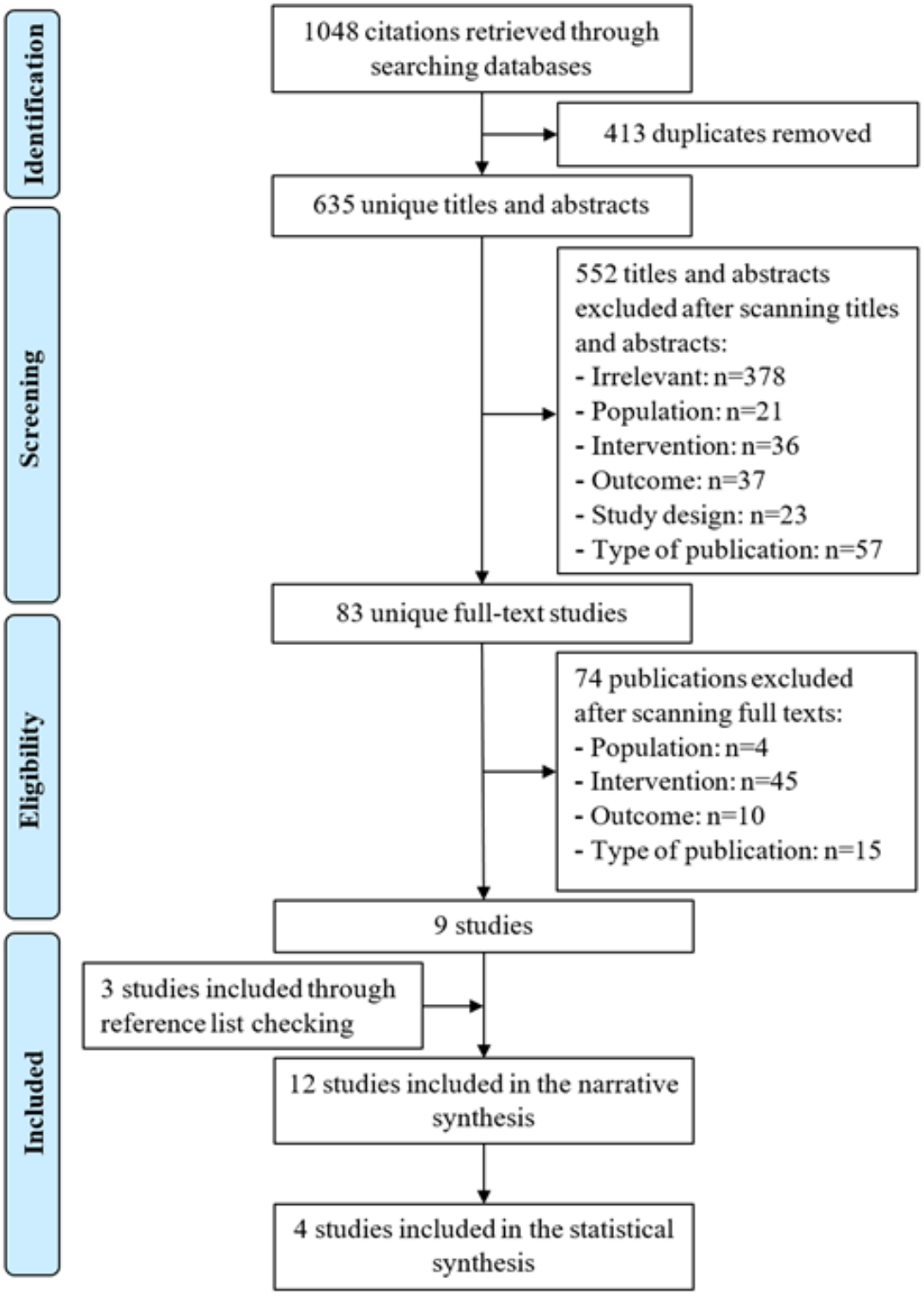
Results

Search Results

The search retrieved 1048 citations (Figure 1). After removing 413 duplicates, 635 unique titles and abstracts remained. By screening those titles and abstracts, 552 citations were excluded. Of the remaining 83 studies, 9 studies were included after reading the full text. Two additional studies were identified from forward reference list checking, and one study was identified by backward reference list checking. Overall, 12 studies were included in the narrative synthesis, but only 4 of those studies were meta-analyzed.

XSL•FO
RenderX

Figure 1. Flow chart of the study selection process.



Description of Included Studies

As shown in Table 1, half of the studies (6/12) were RCTs, while the other half were quasiexperimental. Two-thirds of studies (8/12) were journal articles. Studies were conducted in more than 11 countries. Studies were published between 2015 and 2018. The majority of studies was published in 2018 (7/12). The sample size was <100 in 6 studies (6/12, 50%), and sample sizes ranged from 10 to 454 participants, with a median of 71.5 participants. The age of participants was reported in 10 studies; the mean age of participants in those studies was 31.3 years. The sex of participants was reported in 9 studies; the mean percentage of male participants in those studies was 35%. Half of the studies (6/12) recruited nonclinical samples. Participants were recruited from either community (6/12), educational (4/12), or clinical (3/12) settings. The characteristics of each included study are shown in Multimedia Appendix 4.

XSL•FO
RenderX

Table 1. Characteristics of the included studies (n=12).

Characteristics	Number of studies
Study design	
Quasiexperiment	6
Randomized controlled trial	6
Type of publication	
Journal article	8
Conference proceedings	3
Thesis	1
Country	
United States	4
Japan	1
Sweden	1
Turkey	1
Australia	1
United Kingdom	1
China	1
Romania, Spain, and Scotland	1
Global population	1
Year of publication	
2018	7
2017	2
2016	1
2015	2
Sample size	
<100	6
100-200	3
>200	1
Age (years), mean (range)[a]	31.3 (22-45)
Sex (male), %[b]	35
Sample type	
Clinical sample	6
Nonclinical sample	6
Setting[c]	
Community	6
Educational	4
Clinical	3
Intervention purpose	
Therapy	10
Self-management	2
Intervention platform	
Web-based	6
Standalone software	6

XSL•FO
RenderX

Characteristics	Number of studies
Intervention response generation	
Rule-based	8
Artificial intelligence	4
Intervention dialogue initiative	
Chatbot	9
Both	3
Intervention input modality	
Written	9
Spoken	2
Written and spoken	1
Intervention output modality	
Written	6
Written, spoken, and visual	3
Spoken and visual	2
Written and visual	1
Embodiment	
Yes	6
No	6
Targeted disorders[d]	
Depression	7
Anxiety	4
Any mental disorder	3
Acrophobia	1
Stress	1
Comparator	
Pretest vs posttest	
No intervention	4
Education	3
High users vs low users	1
Measured outcomes[e]	
Severity of depression	6
Psychological wellbeing	3
Severity of anxiety	3
Positive and negative affect	2
Distress	2
Stress	2
Safety	2
Severity of acrophobia	1
Measures[f]	
PHQ-9[g]	4
GAD-7[h]	2

XSL•FO
RenderX

Characteristics	Number of studies
PANAS[i]	2
K10[j]	2
PSS-10[k]	2
AQ[l]	1
HAD-S[m]	1
OASIS[n]	1
WHO-5-J[o]	1
HIQ[p]	1
BDI-2[q]	1
Adverse events	2
Follow-up period[r]	
2 weeks	6
4 weeks	6
6 weeks	1
12 weeks	1

[a]Mean age was reported in 10 studies.

[b]Sex was reported in 9 studies.

[c]Numbers do not add up as one study recruited the sample from more than one setting.

[d]Numbers do not add up as 4 chatbots focused on both depression and anxiety.

[e]Numbers do not add up as most studies assessed more than one outcome.

[f]Numbers do not add up as some studies used more than one tool to assess a single outcome and several studies have more than one outcome.

[g]PHQ-9: Patient Health Questionnaire.

[h]GAD-7: Generalized Anxiety Disorder scale.

[i]PANAS: Positive and Negative Affect Schedule.

[j]K10: Kessler Psychological Distress Scale.

[k]PSS-10: Perceived Stress Scale.

[l]AQ: Acrophobia Questionnaire.

[m]HAD-S: Hospital Anxiety and Depression Scale.

[n]OASIS: Overall Anxiety Severity and Impairment Scale.

[o]WHO-5-J: World Health Organization-5 Well-Being Index.

[p]HIQ: Heights Interpretation Questionnaire.

[q]BDI-2: Beck Depression Inventory II.

[r]Numbers do not add up as two studies assessed outcomes at 2 different points of time.

The included studies investigated the effect of 11 different chatbots. In most studies (10/12) chatbots were used for delivering therapy (Table 1). Chatbots were implemented using standalone software (6/12, 50%) and in web-based platforms (6/12, 50%). Chatbot responses were based on predefined rules or decision trees (rule-based) in two-thirds of studies (8/12). Chatbots in the remaining one-third of studies (4/12) utilized machine learning and natural language processing (artificial intelligence) to understand users' replies and generate responses. Chatbots led and controlled the conversation in 75% (9/12) of the studies. Users could interact with the chatbots using only written language via keyboards and mouse (9/12), only spoken language via microphones (2/12), or a combination of written and spoken languages (1/12). Chatbots used the following modalities to interact with users: only written language via text on the screen (6/12); a combination of written, spoken (via speakers), and visual languages (via embodiment) (3/12); a combination of spoken and visual languages (2/12); and a combination of written and visual languages (1/12). In half of the studies, chatbots contained virtual representations (eg, avatar). Chatbots in 58% of the studies targeted depression (7/12). Multimedia Appendix 5 shows the characteristics of the intervention in each study.

There was no comparator in the 4 one-arm quasiexperiments; these quasiexperimental studies assessed outcomes before and after the intervention (Table 1). In 4 additional studies, an intervention was not provided to the control group. In 3 additional studies, chatbots were compared with providing information or education. In the remaining study, the comparison

XSL•FO
RenderX

was between high users (more engaged app users) and low users (less engaged app users). The most common outcome assessed by the included studies was severity of depression (6/12). The Patient Health Questionnaire (PHQ-9) was the most used outcome measure in the included studies (4/12). The follow-up periods were 2 weeks (6/12), 4 weeks (6/12), 6 weeks (1/12), and 12 weeks (1/12). Characteristics of the comparators and measured outcomes in each included study are presented in Multimedia Appendix 6.

Risk of Bias in the Included Studies

Most of the RCTs (5/6) used an appropriate random allocation sequence, concealed that allocation sequence, and had comparable groups. These studies were rated as having a low risk of bias in the randomization process (Figure 2). Although participants, carers, and people delivering the interventions were aware of the assigned intervention during the trial in most studies (this may be normal due to the nature of the intervention), there were no deviations from the intended intervention because of the experimental context in all studies. Given the lack of deviation and using an appropriate analysis to estimate the effect of assignment to the intervention, a risk of bias due to deviations from the intended interventions was considered low for all studies (Figure 2). The domain of missing outcome data was judged as having a low risk of bias in 4 studies while it was rated as having a high risk of bias in the remaining 2 studies due to a high attrition rate, lack of analysis methods used to correct for bias, and presence of differences between intervention groups in the proportions of missing outcome data.

Although the methods of measuring the outcomes were appropriate and they were comparable between intervention groups (in terms of tools, thresholds, and timing), the risk of bias in the measurement of the outcome was high in 5 studies (Figure 2). This is attributed to the fact that assessors of the outcome were aware of the intervention received by study participants and this knowledge could affect the outcome assessment in those 5 studies. Five studies were judged to raise some concerns in the selection of the reported result (Figure 2). This judgment was due to a discrepancy between studies and their protocols in planned outcome measurements and analyses, unavailability of their protocols, or insufficient details in their protocols regarding outcome measurements and analyses. The overall risk of bias was rated as high for all studies because 5 studies were assessed as high risk in at least one domain, while the remaining study had some concerns in two domains. Multimedia Appendix 7 shows the reviewers' judgments about each "risk of bias" domain for each included RCT.

There was moderate risk of bias due to confounding in all quasiexperimental studies (Figure 3). This judgment was based on a potential for confounding of the effect of intervention in all studies, and it was not clear whether authors in all studies used an appropriate analysis method to control for all confounding domains. The selection of participants was not based on participant characteristics observed after the start of the intervention in 5 studies, and the start of follow-up and start of intervention coincided for most participants in all studies. Accordingly, the "risk of bias due to selection of participant" domain was judged as low in the 5 studies (Figure 3). Although all studies clearly defined the intervention groups at the start of the intervention, it was not clear whether classification of intervention status could be affected by knowledge of the outcome in 3 studies. Therefore, the risk of bias in the classification of the interventions was rated as high in those 3 studies. Further, the risk of bias in this domain was judged as serious in one study, as the classification of the intervention status could be affected by knowledge of the outcome in that study.

Given that there were no deviations from the intended intervention beyond what would be expected in usual practice in all studies, the risk of bias from the deviations from the intended interventions was considered low in all studies (Figure 3). The risk of bias due to missing outcome data was judged as low in 3 studies while it was rated as moderate in the remaining 3 studies due to availability of less than 95% of the participants' data. The risk of bias in the measurement of the outcomes was serious in all studies (Figure 3); assessors of the outcome were aware of the intervention received by study participants, and this could affect the assessment of outcomes. In 5 studies, there was moderate risk of bias in the selection of the reported results (Figure 3); this is because there were insufficient details about the analyses used in the study. While the overall risk of bias was rated as critical in 1 study, it was judged as moderate and serious in 3 and 2 studies, respectively. Multimedia Appendix 8 shows the reviewers' judgments about each "risk of bias" domain for each included quasiexperiment.

XSL•FO
RenderX

Figure 2. Risk of bias graph for randomized controlled trials, showing the review authors' judgments about each risk of bias item.

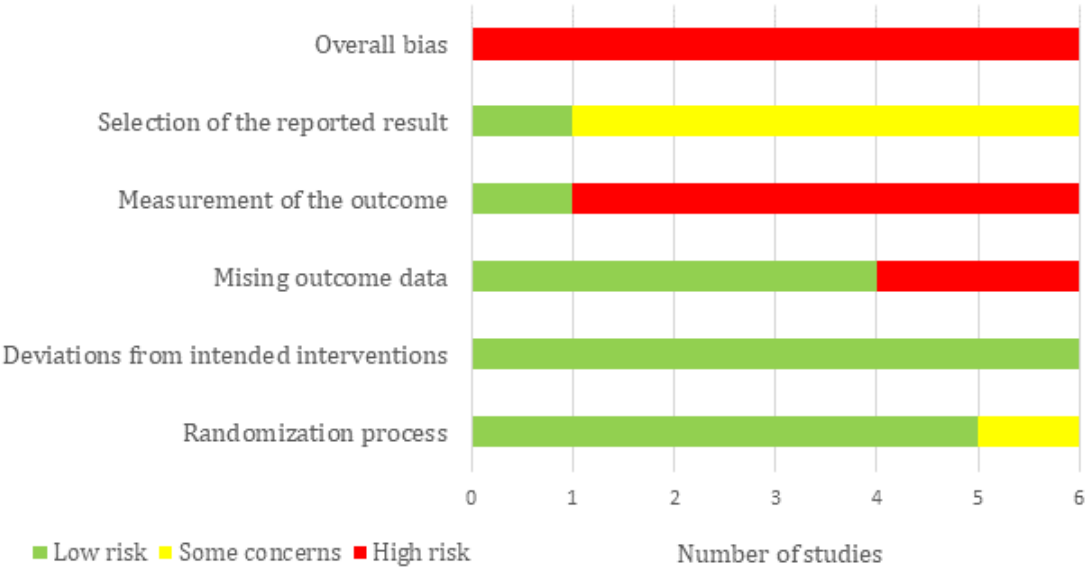


Figure 3. Risk of bias graph for quasiexperiements, showing the review authors' judgments about each risk of bias item.



Results of Studies

Depression

Half of the included studies (6/12) examined the effect of using chatbots on the severity of depression [27-32]. Of these 6 studies, 4 studies were RCTs [27-30], and the remaining 2 studies were pretest-posttest quasiexperiments [31,32]. Four studies were conducted in the United States [28-30,32], and each of the 2 remaining studies was conducted in multiple countries [27,31]. The severity of depression was measured using PHQ-9 [28,29,31,32], Beck Depression Inventory II [27], and Hospital Anxiety and Depression Scale [30].

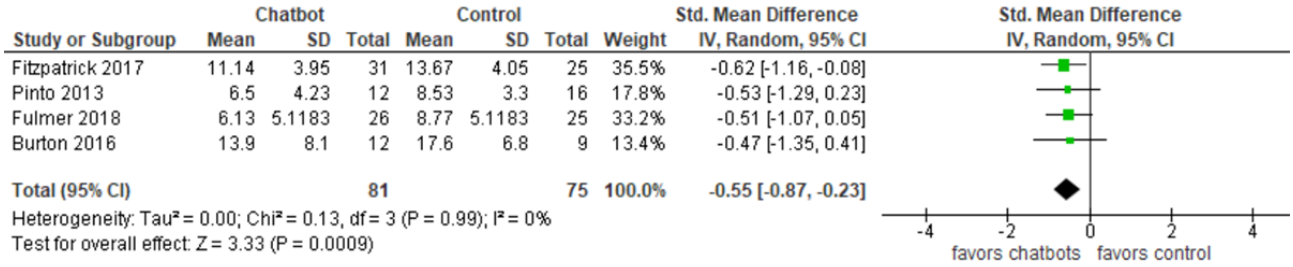
We meta-analyzed the results of only 4 RCTs. However, the results of the 2 quasiexperiments were synthesized narratively

XSL•FO
RenderX

because such a study design has a greater risk of bias than RCTs, and some data required for the meta-analysis was missing from 1 of the 2 studies. The meta-analysis showed a statistically significant difference (*P*<.001) favoring chatbots over treatment as usual or information on the severity of depression (SMD –0.55, 95% CI –0.87 to –0.23; Figure 4). However, this difference was not clinically important, as the total effect (–0.55) was within the boundaries of a minimal clinically important difference (–4.7 to 4.7); the boundaries of a minimal clinically important difference for this outcome was calculated as ±0.5 times the median SD of the control arms of the studies at baseline. The heterogeneity of the evidence was not a concern (*P*=.99; I^2= 0%). The quality of the evidence was low because it was downgraded by 2 levels for a high risk of bias (Multimedia Appendix 9).

Of the 2 quasiexperiments that measured depression, 1 study concluded that the severity of depression decreased significantly postintervention in the high user (*P*<.001) and low user (*P*=.01) groups [31]. Further, the improvement in depression was significantly higher in the high user group than in the low user group (*P*=.03). The second study found a statistically significant decrease in the severity of depression after the intervention (mean 9.78) compared to before the intervention (mean 13.03) [32].

Figure 4. Forest plot of the 4 studies assessing the effect of using chatbots on the severity of depression.

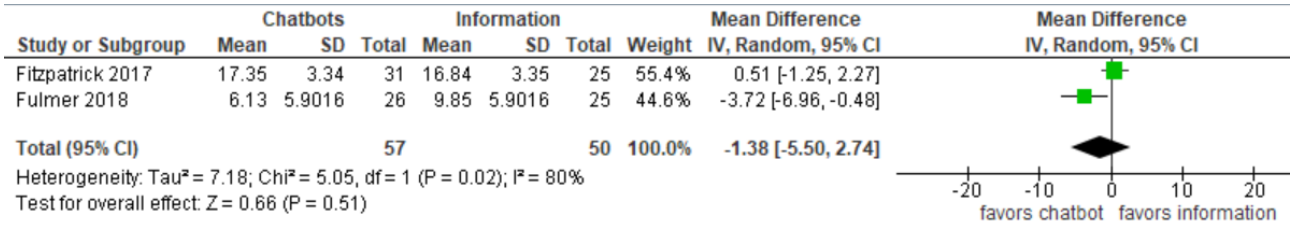


Anxiety

Of the 12 studies, 3 studies assessed the influence of using chatbots on the severity of anxiety [28,29,32]. All studies were conducted in the United States. The severity of anxiety was measured using the Generalized Anxiety Disorder scale [28,29] and Overall Anxiety Severity and Impairment Scale [32]. While 2 studies were RCTs [28,29], the third study was a pretest-posttest quasiexperiment [32]. In contrast to the 2 RCTs, the quasiexperiment was a one-arm trial [32]. For this reason, only the findings of the 2 RCTs were meta-analyzed.

As shown in Figure 5, no statistically significant difference (*P*=.55) in the severity of anxiety was found between those allocated to receive the chatbot intervention compared to those receiving information only (MD –1.38, 95% CI –5.5 to 2.74). There was substantial heterogeneity (*P*=.02; I^2=80%). The quality of the evidence was very low because it was downgraded by 3 levels due to a high risk of bias and heterogeneity (Multimedia Appendix 9). The third study concluded that there was a statistically significant decrease in anxiety level among participants after using chatbots (mean 10.45 versus 7.89) [32].

Figure 5. Forest plot of the 2 studies assessing the effect of using chatbots on the severity of anxiety.



Positive and Negative Affect

The effect of using chatbots on positive and negative affect, which is an indicator of depression and anxiety, was examined in 2 studies [28,29]. Both studies were RCTs conducted in the United States [28,29]. The outcome in the 3 studies was measured using the Positive and Negative Affect Schedule. Meta-analysis could not be executed as only 1 study reported enough data for the analysis [28].

The first study found no statistically significant difference between chatbot use and information on positive affect (*P*=.71) and negative affect (*P*=.91) [28]. In contrast, Fulmer et al [29] found a statistically significant difference favoring chatbot use over information on positive and negative affect at the 2-week follow-up (*P*=.03).

Subjective Psychological Wellbeing

The effect of using chatbots on subjective psychological wellbeing was examined by 3 studies [33-35]. Those studies were conducted in Sweden, Turkey, and Japan, respectively [33-35]. Of the 3 studies, 2 studies were quasiexperiments [34,35], and the remaining study was an RCT [33]. The Flourishing Scale was used to measure subjective psychological wellbeing in 2 studies [33,34], whereas the WHO-5 Well-Being Index was used by the third study [35]. Given that the high risk associated with quasiexperiments and availability of only one RCT, the results of the 3 studies were synthesized narratively.

In the first study [33], the intention-to-treat analysis showed that subjective psychological wellbeing was not statistically different (*P*=.97) after treatment between the chatbot (mean

XSL•FO
RenderX

45.14) and waiting list (mean 45.07) groups. Further, when analyzing data from only the participants who adhered to the intervention, there was no statistically significant difference (*P*=.72) between the chatbot and waiting list groups on subjective psychological wellbeing after treatment (mean 45.07 versus 45.85) [33]. The second study demonstrated a slight improvement in subjective psychological wellbeing after using chatbots, but this improvement was not statistically significant (*P*=.06) [34]. Similarly, the third study found no statistically significant difference (*P*=.32) between the chatbot and control groups on subjective psychological wellbeing after treatment [35].

Psychological Distress

The influence of using chatbots on psychological distress was examined by 2 studies, conducted in Japan and Australia [35,36]. Distress was measured using the Kessler Psychological Distress Scale. While 1 study was a one-group quasiexperiment [36], the other study was a two-group quasiexperiment [35]. Therefore, a narrative approach was used to analyze their results.

According to Suganuma et al [35], there was a statistically significant difference (*P*=.005) favoring chatbot use (mean 21.65) over no intervention (mean 23.97) on distress levels after treatment. Further, there was a statistically significant improvement in distress level among the chatbot group after treatment (mean 21.65) compared with before treatment (mean 23.58). Likewise, the other study found a statistically significant decrease (*P*<.001) in distress from a pre-intervention score of 33.27 to a post-intervention score of 28.90 [36].

Stress

Stress was an outcome in 2 studies [33,37]. The first was an RCT conducted in Sweden [33], and the second was a quasiexperimental study conducted in China [37]. The Perceived Stress Scale was utilized to measure stress in both studies. A meta-analysis was not carried out for this outcome as 1 study [37] did not report data required for the analysis.

Ly and colleagues [33] found a statistically significant difference favoring chatbots over the waiting list on stress when they analyzed data from all participants (*P*=.03) and from those who only adhered to the intervention (*P*=.01). Huang et al [37] concluded that stress status improved over time when using a chatbot.

Acrophobia

The effect of using chatbots on acrophobia (ie, fear of height) was examined by 1 RCT conducted in the United Kingdom [38]. The outcome was measured using two tools: Heights Interpretation Questionnaire and Acrophobia Questionnaire. Compared with participants who received usual care, the chatbot significantly decreased the severity of acrophobia as measured by both tools at 2-week and 4-week follow-ups (*P*<.001) [38].

Safety

Safety of chatbots was assessed in 2 RCTs [30,38]. While 1 study was conducted in the United States [30], the other study was conducted in the United Kingdom [38]. The former study concluded that the chatbot was safe because users did not report any harm, distress, adverse events, or worsening of depressive symptoms resulting from using the chatbot during the study [30]. Similarly, Freeman et al [38] concluded that the chatbot was safe because no serious adverse events (eg, suicide attempts, death, serious violent incidents) or discomfort caused by the chatbot were reported.

Discussion

Principal Findings

This study systematically reviewed the evidence regarding the effectiveness and safety of using chatbots to improve mental health. We identified 12 studies examining the effect of using chatbots on 8 outcomes. For the first outcome (depression), low-quality evidence from 4 RCTs showed a statistically significant difference favoring chatbots over treatment as usual or information on the severity of depression, but this difference was not clinically important. Two quasiexperiments concluded that the level of depression decreased after using chatbots. As evidence from the 2 studies was synthesized narratively, we could not identify whether this decrease in depression was clinically important. Findings in the 2 studies may be affected by serious bias in the measurement of outcomes. Given that no reviews assessed the effectiveness of chatbots in mental health, the results were compared with other reviews regarding similar interventions (ie, internet-based psychotherapeutic interventions). The overall effect on depression in this review (–0.55) was comparable to other reviews. Specifically, while the overall effect of internet-based and computerized psychological interventions of depression without therapist support was 0.25 (95% CI 0.14-0.35) in a meta-analysis conducted by Andersson and Cuijpers [39], another meta-analysis showed that the total effect of internet-based psychotherapeutic interventions of depression was 0.32 [40].

With regards to anxiety, very low-quality evidence from 2 RCTs showed no statistically significant difference between chatbots and information on the severity of anxiety. In contrast, one quasiexperiment concluded that anxiety levels considerably decreased after using chatbots. These contradictory findings may be attributed to 2 reasons. First, pretest-posttest quasiexperiments are not as reliable as RCTs for finding the effect of an intervention due to low internal validity resulting from selection bias [35,41]. Second, in contrast to the 2 RCTs, the chatbot in the quasiexperiment [32] contained a virtual representation (ie, embodiment), which enables chatbots to communicate with users verbally and nonverbally (through body movements and facial expressions). It is purported that embodiment makes conversations with chatbots more empathetic and facilitates effective rapport with users [19,42,43]. Results of the meta-analyses in this review and another review related to smartphone mental health interventions were contradictory. A meta-analysis of 9 RCTs showed a considerable reduction in the anxiety level after using smartphone mental health interventions compared to no intervention (SMD 0.325, 95% CI 0.17-0.48) [44]. These conflicting results may result from either differences in interventions (chatbots versus different mobile interventions) in both reviews or the number of meta-analyzed studies (2 versus 9).

XSL•FO
RenderX

Findings regarding the effect of chatbots on positive and negative affect were conflicting. While one study concluded that chatbots improved the positive and negative affect at the 2-week follow-up [29], another study did not find any significant influence of chatbots at the 2-week follow-up [28]. Although the 2 studies were very homogenous in terms of study design, sample characteristics, comparator characteristics, and outcome measures, they were different in the type of chatbots and data analysis, and these differences may have led to contradictory findings. Specifically, the chatbot in the first study [29] was more advanced than the one in the second study [28]; it depended on artificial intelligence and machine learning to generate responses to users, and this makes it more humanlike and lets users feel more socially connected [5]. With regards to the second difference, while the first study assessed the effect of the chatbot on positive and negative affect together [29], the other study examined the effect of the chatbot on positive affect and negative affect separately [28].

A narrative synthesis of 3 studies showed no statistically significant difference between chatbots and control group on subjective psychological wellbeing. The justification for the nonsignificant difference is the use of a nonclinical sample in the 3 studies. In other words, as participants already had good psychological wellbeing, the effect of using chatbots may be less likely to be significant.

According to the 2 studies synthesized in a narrative approach, chatbots significantly decreased the levels of distress. Both studies had a high risk of bias; therefore, this finding should be interpreted with caution. Studies in a similar context reported findings comparable to our findings. To be more precise, an RCT concluded that online chat counselling significantly improved psychological distress over time [45].

In this review, chatbots significantly decreased stress levels over time. Unfortunately, we cannot draw a definitive conclusion regarding the effect of chatbots due to the high risk of bias in the evidence.

Chatbots were effective in decreasing the severity of acrophobia according to one RCT. The effect size of chatbots on acrophobia in this RCT [38] was substantially higher than the total effect size of therapist-assisted exposure treatment on phobias reported by a meta-analysis (2.0 versus 1.1) [46]. This indicates that chatbots may be equivalent to, if not better, exposure treatment delivered by a therapist in treating phobias.

Of the 2 RCTs measuring the safety of chatbots, both concluded that chatbots are safe for use in mental health, as no adverse events or harm were reported when chatbots were used to treat users with depression and acrophobia. However, this evidence is not sufficient to conclude that chatbots are safe, given the high risk of bias in the 2 studies.

Strengths and Limitations

Strengths

This study is the first review of the literature that assessed the effectiveness and safety of chatbots in mental health. The findings are of importance for users, providers, policymakers, and researchers. This review was developed, executed, and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [20]. Accordingly, this enabled us to produce a high-quality review.

In this review, the most popular databases in health and information technology were used to run the most sensitive search possible. The review minimized the risk of publication bias as much as possible through searching Google Scholar and conducting backward and forward reference list checking to identify grey literature. The search was not restricted to a certain type of chatbots, comparators, outcomes, year of publication, nor country of publication, and this makes the review more comprehensive.

To reduce selection bias, two reviewers independently selected studies, extracted data, and assessed the risk of bias in the included studies and quality of the evidence. Agreement between reviewers was very good, except for the assessment of the risk of bias (which was good). When possible, findings of the included studies were meta-analyzed; thereby, we were able to increase the power of the studies and improve the estimates of the likely size of effect of chatbots on a variety of mental health outcomes.

Limitations

The intervention of interest in this review was restricted to chatbots that work within standalone software or via a web browser (but not robotics, serious games, SMS, nor telephones). Further, we excluded studies that contained chatbots controlled by human operators. Accordingly, this review cannot comment on the effectiveness of chatbots that involve human-generated content or those that use alternative modes of delivery. It was necessary to apply those restrictions because these features are not part of ordinary chatbots. For this reason, 3 previous reviews about chatbots applied these restrictions [5,13,19].

Owing to practical constraints, the search was restricted to English studies. Therefore, it is likely that we missed some non-English studies. The overall risk of bias was high in most of the included studies. The quality of evidence in the meta-analyses ranged from very low to low. Accordingly, the high risk of bias and low quality of evidence may reduce the validity of the findings and their generalizability.

Ideally, the difference between pre-intervention and post-intervention data for each group should be used in a meta-analysis [47]. However, we used only post-intervention data in each group for the meta-analysis because studies did not report enough data (eg, change in SD or SE of the mean between the pre-intervention and post-intervention for each group). In this review, it was possible to meta-analyze pre-intervention and post-intervention data from one-group trials (ie, did not include comparison groups). However, such analysis was not carried out in this review as such trials are very vulnerable to several threats of internal validity, such as maturation threat, instrumentation threat, regression threat, and history threat [41,48].

XSL•FO
RenderX

Practical and Research Implications

Practical Implications

Although this review found that chatbots may improve depression, distress, stress, and acrophobia, definitive conclusions regarding those results could not be drawn due to the high risk of bias in the included studies, low quality of evidence, lack of studies assessing each outcome, small sample size in the included studies, and contradictions in results of some included studies. For this reason, results should be viewed with caution by users, health care providers, caregivers, policymakers, and chatbot developers.

Given the weak and conflicting evidence found in this review, users should not use chatbots as a replacement for mental health professionals. Instead, health professionals should consider offering chatbots as an adjunct to already available interventions to encourage individuals to seek medical advice where appropriate and as a signpost to available support and treatment.

Most chatbots in this review were implemented in developed counties. People in developing countries may be more in need of chatbots than those in developed countries given that developing countries have a greater shortage of mental health professionals than developed countries (0.1 per 1,000,000 people vs 9 per 100,000 people) [7,8]. System developers should consider implementing more chatbots in developing countries.

Two-thirds of the chatbots in this review used predefined rules and decision trees to generate their responses, while the remaining chatbots used artificial intelligence. In contrast to rule-based chatbots, artificial intelligence chatbots can generate responses to complicated queries and enable users to control the conversation [13]. Artificial intelligence chatbots can exhibit more empathetic behaviors and humanlike filler language than rule-based chatbots [19]. This may make artificial intelligence chatbots more effective in building rapport with users, thereby improving their mental health [42]. It could be argued that artificial intelligence chatbots are more prone to errors than rule-based chatbots, but these errors can be minimized and diminished by extensive training and greater use [49]. Accordingly, we recommend developers concentrate efforts around artificial intelligence chatbots to improve the effectiveness.

Research Implications

This review showed that there is a lack of evidence assessing the effectiveness and safety of chatbots. Accordingly, we encourage researchers to conduct more studies in this area. Further, they should undertake more studies in developing countries and recruit large, clinical samples given the lack of such evidence, as found in the current review.

The overall risk of bias was high in most included studies mainly due to issues in the measurement of the outcomes, selection of the reported result, and confounding. Future studies should follow recommended guidelines or tools (eg, RoB 2 and ROBINS-I) when conducting and reporting their studies in order to avoid such biases.

Due to poor reporting practices, we were unable to include many studies in the meta-analysis. As well as encouraging more high-level studies (ie, RCTs), there is a need for authors to be more consistent in their reporting of trial outcomes. For example, in our review, many studies failed to report basic descriptive statistics such as mean, SD, and sample size. Ensuring studies adhere to accepted guidelines for reporting RCTs (eg, CONSORT-EHEALTH [50]) would be of considerable benefit to the field.

In the current review, the comparators in all two-group trials were either no intervention or education. For those outcomes that hold promise (eg, depression, distress, and acrophobia), we encourage researchers to compare chatbots with other active interventions such as asynchronous electronic interventions or other types of chatbots (eg, rule-based chatbots versus artificial intelligence chatbots or embodied chatbots versus non-embodied chatbots).

According to a scoping review conducted by Abd-alrazaq et al [13], chatbots are used for many mental disorders, such as autism, post-traumatic stress disorder, substance use disorders, schizophrenia, and dementia. The current review did not find any study assessing the effectiveness or safety of chatbots used for these disorders. This highlights a pressing need to examine the effectiveness and safety of chatbots targeting patients with autism, post-traumatic stress disorder, substance use disorders, schizophrenia, and dementia.

As this review focused on the effectiveness and safety of chatbots, we excluded many studies that assessed the usability and acceptance of chatbots in mental health. Given that usability and acceptance of technology are considered important factors for their successful implementation, the evidence regarding those outcomes should be summarized through systematic reviews.

The current review identified heterogeneity in the tools used to measure the same outcomes and in the research design. For instance, severity of depression was measured using PHQ-9, Beck Depression Inventory II, or Hospital Anxiety and Depression Scale. Further, while some studies assessed outcomes before and after interventions, other studies examined them only after interventions. The field would benefit from future studies using a common set of outcome measures to ease comparison and interpretation of results between studies. Only one study assessed the long-term effectiveness and safety of chatbots, where participants were followed for 12 weeks. The effectiveness and safety outcomes of chatbots may be different when considering long-term, relative to short-term, findings; it is essential to assess long-term outcomes.

Conclusion

Although the included studies showed that chatbots may be safe and improve depression, distress, stress, and acrophobia, definitive conclusions regarding the effectiveness and safety of chatbots could not be drawn in this review for several reasons. First, the statistically significant difference between chatbots and other interventions on the severity of depression was not clinically important. Second, the risk of bias was high in most included studies, and the quality of the meta-analyzed evidence ranged from very low to low. Third, the evidence for each outcome came from only a few studies that also had small

XSL•FO
RenderX

sample sizes. Fourth, studies showed conflicting results for some outcomes (ie, anxiety and positive and negative affect). Researchers should avoid shortcomings in the study designs reported in this review. Health care providers should consider offering chatbots as an adjunct to already available interventions.

Acknowledgments

The publication of this article was funded by the Qatar National Library. This study was a part of a project funded by the Qatar National Research Fund (NPRP12S-0303-190204). The project title is "A Personalized and Intelligent Digital Mental Health Platform for Qatar and the Arab world".

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOC File , 66 KB - jmir_v22i7e16021_app1.doc]

Multimedia Appendix 2
Search strings utilized for searching each bibliographic database.
[DOCX File , 28 KB - jmir_v22i7e16021_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 18 KB - jmir_v22i7e16021_app3.docx]

Multimedia Appendix 4
Characteristics of each included study.
[DOCX File , 19 KB - jmir_v22i7e16021_app4.docx]

Multimedia Appendix 5
Characteristics of the intervention in each study.
[DOCX File , 18 KB - jmir_v22i7e16021_app5.docx]

Multimedia Appendix 6
Characteristics of comparators and measured outcomes in each included study.
[DOCX File , 19 KB - jmir_v22i7e16021_app6.docx]

Multimedia Appendix 7
Reviewers' judgements about each "risk of bias" domain for each included RCT.
[DOCX File , 24 KB - jmir_v22i7e16021_app7.docx]

Multimedia Appendix 8
Reviewers' judgements about each "risk of bias" domain for each included quasi-experiment.
[DOCX File , 69 KB - jmir_v22i7e16021_app8.docx]

Multimedia Appendix 9
GRADE evidence profile.
[DOCX File , 15 KB - jmir_v22i7e16021_app9.docx]

References

1. Mental Health Foundation. Fundamental Facts About Mental Health 2016. London: Mental Health Foundation; 2016. URL: https://www.mentalhealth.org.uk/sites/default/files/fundamental-facts-about-mental-health-2016.pdf [accessed 2020-06-08]
2. Mental Health Foundation. Fundamental Facts About Mental Health 2015. London: Mental Health Foundation; 2015. URL: https://www.mentalhealth.org.uk/sites/default/files/fundamental-facts-15.pdf [accessed 2020-06-08]

XSL•FO
RenderX

3. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS One 2015 Feb;10(2):e0116820 [FREE Full text] [doi: 10.1371/journal.pone.0116820] [Medline: 25658103]
4. Jones SP, Patel V, Saxena S, Radcliffe N, Ali Al-Marri S, Darzi A. How Google's 'ten Things We Know To Be True' could guide the development of mental health mobile apps. Health Aff (Millwood) 2014 Sep;33(9):1603-1611. [doi: 10.1377/hlthaff.2014.0380] [Medline: 25201665]
5. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
6. Wainberg ML, Lu FG, Riba MB. Global Mental Health. Acad Psychiatry 2016 Aug 3;40(4):647-649 [FREE Full text] [doi: 10.1007/s40596-016-0577-0] [Medline: 27259490]
7. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. BJPsych Int 2016 Aug 02;13(3):61-63 [FREE Full text] [doi: 10.1192/s2056474000001240] [Medline: 29093905]
8. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 2012 Dec;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4]
9. Hester RD. Lack of access to mental health services contributing to the high suicide rates among veterans. Int J Ment Health Syst 2017 Aug 18;11(1):47 [FREE Full text] [doi: 10.1186/s13033-017-0154-2] [Medline: 28828036]
10. Anthes E. Mental health: There's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
11. Schueller SM, Glover AC, Rufa AK, Dowdle CL, Gross GD, Karnik NS, et al. A Mobile Phone-Based Intervention to Improve Mental Health Among Homeless Young Adults: Pilot Feasibility Trial. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e12347 [FREE Full text] [doi: 10.2196/12347] [Medline: 31267980]
12. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
13. Abd-Alrazaq AA, Alajlani M, Alalwan A, Bewick B, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform 2019 Dec;132:103978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
14. Kumar V, Keerthana A. Sanative Chatbot For Health Seekers. IJECS 2016 Mar 29:16022-16025. [doi: 10.18535/ijecs/v5i3.28]
15. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
16. Radziwill N, Benton M. Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv 2017:170404579.
17. Schunemann H, Oxman A, Vist G, Higgins J, Deeks J, Glasziou P. Chapter 12: Interpreting results and drawing conclusions. In: Higgins J, Green S. editors. Cochrane handbook for systematic reviews of interventions Sussex, UK: John Wiley & Sons; 2008:359-387.
18. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
19. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700-b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
21. Higgins J, Deeks J. Chapter 7: selecting studies and collecting data. In: J. Higgins, Green S. editors. Cochrane Handbook for Systematic Reviews of Interventions. ed. Sussex, UK: John Wiley & Sons; 2008:151-185.
22. Altman D. Practical statistics for medical research. 1 ed: CRC press; 1990:0412276305.
23. Higgins JP, Sterne JA, Savovic J, Page MJ, Hróbjartsson A, Boutron I, et al. Appraising the risk of bias in randomized trials using the Cochrane Risk of Bias Tool. Cochrane Database of Systematic Reviews 2016;10(Suppl 1):1.
24. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12:i4919. [doi: 10.1136/bmj.i4919]
25. Copay A, Subach B, Glassman S, Polly D, Schuler T. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7(5):541-546 [FREE Full text] [doi: 10.1016/j.spinee.2007.01.008] [Medline: 17448732]
26. Deeks J, Higgins J, Altman D. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins J, Green S. editors. Cochrane handbook for systematic reviews of interventions. Sussex, UK: John Wiley & Sons; 2008:A-96.

XSL•FO
RenderX

27. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
28. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
29. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
30. Pinto MD, Greenblatt AM, Hickman RL, Rice HM, Thomas TL, Clochesy JM. Assessing the Critical Parameters of eSMART-MH: A Promising Avatar-Based Digital Therapeutic Intervention to Reduce Depressive Symptoms. Perspect Psychiatr Care 2016 Jul;52(3):157-168. [doi: 10.1111/ppc.12112] [Medline: 25800698]
31. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
32. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. Canada: ACM; 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21–26, 2018; Montreal, QC, Canada p. A. [doi: 10.1145/3173574.3173972]
33. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
34. Demirci H. User experience over time with conversational agents case study of woebot on supporting subjective well-being. Ankara, Turkey: Middle East Technical University; 2018.
35. Suganuma S, Sakamoto D, Shimoyama H. An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]
36. Luerssen M, Hawke T. Virtual Agents as a Service: Applications in Healthcare. Australia: ACM; 2018 Presented at: Proceedings of the 18th International Conference on Intelligent Virtual Agents; November 5-8, 2018; Sydney, NSW, Australia p. 1-12. [doi: 10.1145/3267851.3267858]
37. Huang J, Li Q, Xue Y, Cheng T, Xu S, Jia J. Teenchat: a chatterbot system for sensing and releasing adolescents' stress. In: editors. 2015 Presented at: International Conference on Health Information Science; May 28-30, 2015; Melbourne, Australia p. 133-145.
38. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. The Lancet Psychiatry 2018 Aug;5(8):625-632. [doi: 10.1016/S2215-0366(18)30226-8] [Medline: 2018]
39. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
40. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
41. Bhattacherjee A. Social science research: Principles, methods, and practices. University of South Florida: Creative Commons Attribution; 2012.
42. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. Front. Robot. AI 2017 Oct 12;4:1. [doi: 10.3389/frobt.2017.00051]
43. Martínez-Miranda J, Bresó A, García-Gómez J. Look on the bright side: a model of cognitive change in virtual agents. In: editors. 2014 Presented at: International Conference on Intelligent Virtual Agents; August 27-29, 2014; Boston, MA, USA p. 285-294. [doi: 10.1007/978-3-319-09767-1_37]
44. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
45. Dowling M, Rickwood D. A naturalistic study of the effects of synchronous online chat counselling on young people's psychological distress, life satisfaction and hope. Couns. Psychother. Res 2015 Jul 14;15(4):274-283. [doi: 10.1002/capr.12037]
46. Wolitzky-Taylor KB, Horowitz JD, Powers MB, Telch MJ. Psychological approaches in the treatment of specific phobias: a meta-analysis. Clin Psychol Rev 2008 Jul;28(6):1021-1037. [doi: 10.1016/j.cpr.2008.02.007] [Medline: 18410984]
47. Littell J, Corcoran J, Pillai V. Systematic reviews and meta-analysis. London, UK: Oxford University Press; 2008.
48. Shadish W, Cook T, Campbell D. Experimental and quasi-experimental designs for generalized causal inference. Boston, United States: Cengage Learning; 2002:9780395615560.

XSL•FO
RenderX

49. López-Cózar R, Callejas Z, Espejo G, Griol D. Enhancement of conversational agents by means of multimodal interaction. In: Diana P, Ismael P, editors. Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. Hershey, PA, USA: IGI Global; 2011:223-252.
50. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

AQ: Acrophobia Questionnaire.
BDI-2: Beck Depression Inventory II.
FS: Flourishing Scale.
GAD-7: Generalized Anxiety Disorder scale.
HAD-S: Hospital Anxiety and Depression Scale.
HIQ: Heights Interpretation Questionnaire.
K10: Kessler Psychological Distress Scale.
MD: mean difference.
MCID: Minimal clinically important difference
OASIS: Overall Anxiety Severity and Impairment Scale.
PANAS: Positive and Negative Affect Schedule.
PHQ-9: Patient Health Questionnaire.
PSS-10: Perceived Stress Scale.
RCT: randomized controlled trial.
RoB 2: Risk-of-Bias 2.
ROBINS-I: Risk Of Bias In Non-randomized Studies – of Interventions.
SMD: standardized mean difference.
WHO-5-J: World Health Organization-5 Well-Being Index.

Edited by G Eysenbach; submitted 27.08.19; peer-reviewed by K Denecke, B McMillan, H Cho; comments to author 22.10.19; revised version received 30.10.19; accepted 15.12.19; published 13.07.20.

Please cite as:
Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M
Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(7):e16021
URL: http://www.jmir.org/2020/7/e16021/
doi:10.2196/16021
PMID:32673216

©Alaa Ali Abd-Alrazaq, Asma Rababeh, Mohannad Alajlani, Bridgette M Bewick, Mowafa Househ. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptance and Commitment Therapy Delivered via a Mobile Phone Messaging Robot to Decrease Postoperative Opioid Use in Patients With Orthopedic Trauma: Randomized Controlled Trial

Chris A Anthony[1,2], MD; Edward Octavio Rojas[3], MD; Valerie Keffala[3], PhD; Natalie Ann Glass[3], PhD, MHCDS; Apurva S Shah[4], MD, MBA; Benjamin J Miller[3], MD, MS; Matthew Hogue[3], MD; Michael C Willey[3], MD; Matthew Karam[3], MD; John Lawrence Marsh[3], MD

[1]Department of Orthopaedics, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Orthopedic Surgery, Washington University School of Medicine, St Louis, MO, United States

[3]Department of Orthopaedics and Rehabilitation, University of Iowa Hospitals and Clinics, Iowa City, IA, United States

[4]Children's Hospital of Philadelphia Main Campus Division of Orthopaedics, Philadelphia, PA, United States

Corresponding Author:
Edward Octavio Rojas, MD
Department of Orthopaedics and Rehabilitation
University of Iowa Hospitals and Clinics
200 Hawkins Drive
Iowa City, IA, 52242
United States
Phone: 1 319 356 8690
Email: edward-rojas@uiowa.edu

Abstract

Background: Acceptance and commitment therapy (ACT) is a pragmatic approach to help individuals decrease avoidable pain.

Objective: This study aims to evaluate the effects of ACT delivered via an automated mobile messaging robot on postoperative opioid use and patient-reported outcomes (PROs) in patients with orthopedic trauma who underwent operative intervention for their injuries.

Methods: Adult patients presenting to a level 1 trauma center who underwent operative fixation of a traumatic upper or lower extremity fracture and who used mobile phone text messaging were eligible for the study. Patients were randomized in a 1:1 ratio to either the intervention group, who received twice-daily mobile phone messages communicating an ACT-based intervention for the first 2 weeks after surgery, or the control group, who received no messages. Baseline PROs were completed. Two weeks after the operative intervention, follow-up was performed in the form of an opioid medication pill count and postoperative administration of PROs. The mean number of opioid tablets used by patients was calculated and compared between groups. The mean PRO scores were also compared between the groups.

Results: A total of 82 subjects were enrolled in the study. Of the 82 participants, 76 (38 ACT and 38 controls) completed the study. No differences between groups in demographic factors were identified. The intervention group used an average of 26.1 (SD 21.4) opioid tablets, whereas the control group used 41.1 (SD 22.0) tablets, resulting in 36.5% ([41.1-26.1]/41.1) less tablets used by subjects receiving the mobile phone–based ACT intervention (P=.004). The intervention group subjects reported a lower postoperative Patient-Reported Outcome Measure Information System Pain Intensity score (mean 45.9, SD 7.2) than control group subjects (mean 49.7, SD 8.8; P=.04).

Conclusions: In this study, the delivery of an ACT-based intervention via an automated mobile messaging robot in the acute postoperative period decreased opioid use in selected patients with orthopedic trauma. Participants receiving the ACT-based intervention also reported lower pain intensity after 2 weeks, although this may not represent a clinically important difference.

Trial Registration: ClinicalTrials.gov NCT03991546; https://clinicaltrials.gov/ct2/show/NCT03991546

(J Med Internet Res 2020;22(7):e17750) doi:10.2196/17750

XSL•FO
RenderX

KEYWORDS

acceptance and commitment therapy; opioid crisis; patient-reported outcome measures; postoperative pain; orthopedics; text messaging; chatbot; conversational agents; mHealth

Introduction

Opioid Medication Issues

Public health concerns regarding opioid medications persist, and health care systems are currently seeking solutions to the ongoing epidemic [1]. In 2017, the rate of drug overdose deaths involving opioids in the United States increased by 12%, totaling 47,600 cases, and prescription opioid medications accounted for over 17,000 of these [2]. Even small amounts of additional daily opioid utilization (10 morphine milliequivalents [MME]) by patients can lead to an increased risk of long-term misuse [3]. In addition, every week of continued opioid utilization represents an increased risk of eventual misuse by patients [3]. Previous studies have found that orthopedic trauma patients use a decreasing amount of opioid medication in the first 2 postoperative weeks, with 6-15 days being the optimal opioid use period [4,5]. In line with these findings, previous studies have used a 2-week postoperative period to assess opioid medication consumption in surgical patients [4].

Patient-Reported Outcomes

Patient-reported outcomes (PROs) allow patients to quantify aspects of their orthopedic condition in a standardized fashion [6,7]. These are important tools for determining the efficacy of health care treatments and assessment of clinical research and can be used in determining compensation for health care services provided [6-8]. The National Institutes of Health developed the Patient-Reported Outcome Measure Information System (PROMIS) tools to advance PROs by creating question banks that could be used for many major health issues [9]. The PROMIS Pain Intensity 1A Short form, PROMIS Pain Intensity 3A Short form, PROMIS Pain Interference 8A Short form, and PROMIS Emotional Distress-Anxiety 8A Short form all employ a fixed low number of questions that are highly reliable when compared with their respective domain's full item bank, making them excellent tools for both patients and clinicians [9,10].

Acceptance and Commitment Therapy

Acceptance and commitment therapy (ACT) is a cognitive contextual behavioral therapy that employs a pragmatic approach to help individuals decrease pain and live according to self-identified personal values [11,12]. The goal of ACT is to augment an individual's psychological flexibility, thus improving their life according to 6 core cognitive processes: acceptance, defusion, contact with the present moment, self-as-context, values, and committed action (Table 1) [11]. Employing these cognitive processes to increase psychological flexibility allows people to choose their actions based on what they value most, resulting in decreased avoidance behaviors and negative cognitive associations [11]. ACT has proven to be effective across multiple studies and patient populations in the treatment of pain [13]. Several studies report a high value for ACT in the management of chronic pain when compared with standard pharmacological treatment alone [13-16]. Moreover, earlier cessation of pain and opioid utilization in at-risk orthopedic surgery patients receiving office-based ACT interventions has also been reported [17]. However, traditional ACT interventions require a clinic-based, interdisciplinary team approach, which is not always feasible for both patients and health care systems [17,18].

XSL•FO
RenderX

Table 1. Acceptance and commitment therapy core principles with associated messages.

Core principle	Example mobile phone message
Values: know what matters most	Stop for a moment and remember the 3 values you identified earlier today. Remind yourself how important these values are in your life. As your day comes to an end, remember that YOU are in control of the thoughts that exist in your mind. We encourage you to spend time thinking about your 3 core values identified earlier today.
Acceptance: setting expectation that pain is a part of surgery	Feelings of pain and feelings about your experience of pain are normal after surgery. Acknowledge and accept these feelings as part of the recovery process. Remember how you feel now is temporary and your healing process will continue. Call to mind pleasant feelings or thoughts that you experienced today.
Present moment awareness: mindfulness and awareness for our thoughts in the present moment	Awareness of the present moment and your breathing may change with pain-related emotions or thoughts. Remember you can always count on your breathing to bring you back to the present moment and help you move through your current experience of pain.
Self-as-context: awareness of what is being observed and noticed by ourselves	We cannot change that a feeling or thought may arise, but we can choose how we respond to our feelings and thoughts. Remember that dwelling on pain, discouraging feelings, and thoughts after surgery are NOT consistent with your life goals and values. Observe things that try to move you away from your values and only act on things that are compatible with who you want to be and what matters to you.
Committed action: doing what it takes to live according to our values	Healing after surgery requires you to act. We previously discussed your life goals, meaning, and purpose. Take action today and move closer toward what you want in life. Recognize that pain may be present but make the choice that it will not impede your progress toward what you really want in life. Be present in the moment and ensure your actions remain true to what you want most. All actions you make no matter how small, are an important steppingstone on your road to recovery.
Defusion: watch your thinking and interact with thoughts in a way consistent with your values	If you ever feel pain after surgery know that the feeling is real but what it actually represents is not what you might think. Our mind is capable of making us feel pain, even though there is no damage going on in our body. Pause, become more aware in the moment and chose a skillful response that will help you move toward your overall goals and values.

Mobile Phone Messaging Communication

Evolving communication methods, such as automated mobile phone messaging [4,19-21], for health care purposes are increasingly important, as patients prefer these communication methods for delivering and receiving medical information [22]. Software-driven, automated mobile phone messaging robots (also called chatbots or conversational agents) are low-cost tools that can deliver predefined text-based information and receive incoming responses with high reliability when patients either prefer it or it is necessary to communicate at distance [19,23]. This technology demonstrates high efficacy as part of the treatment of conditions ranging from hypertension to substance abuse [24-26], and it has also proved effective in increasing perioperative communication after hip and knee arthroplasty [27] and collecting pain and opioid medication data from patients following orthopedic trauma and hand procedures [4,19]. In addition, automated mobile phone messaging robots have been validated in the collection of PROs in patients undergoing orthopedic hand [20] and hip preservation procedures [23]. Although mobile phone messaging robots (*Chat bots*) provide the benefit of communicating with patients at distance with no need for human intervention, they do introduce delivery of health care that lacks human interaction with unknown effects [28].

Health care teams caring for patients with traumatic orthopedic injuries have traditionally used opioid medication in the postoperative setting, and these patients are at risk for prolonged opioid utilization in the postoperative period. We theorized that the combination of ACT delivered via automated mobile phone messaging may help to decrease pain and opioid utilization in the acute postoperative setting. The aim of this prospective randomized controlled trial was to evaluate the effectiveness of ACT delivered via an automated mobile messaging robot on (1) decreasing early postoperative opioid utilization and (2) pain-related PROs in the first 2 weeks following surgery for acute traumatic orthopedic injuries.

Methods

Study Approval

This randomized controlled trial was registered with ClinicalTrials.gov (NCT03991546) and reporting is consistent with the Consolidated Standards of Reporting Trials guidelines (Multimedia Appendix 1) [29]. The study was performed at a single center university hospital in Iowa City, Iowa, United States. Ethical approval of this study was provided by the University of Iowa institutional review board, and the study was determined to be Health Insurance Portability and Accountability Act compliant.

Recruitment and Randomization

Adults presenting to a university hospital level 1 trauma center indicated for operative fixation of a traumatic upper or lower fracture were considered for the study (Table 2). The exclusion criteria are listed in Textbox 1. Eligible patients were approached before surgery by a research assistant in a private room. Individuals not excluded by screening questions and interested in participating underwent the informed consent process (Textbox 1). During consent, all subjects were informed of the outcomes of interests, different study arms, and that no changes would be made to their care in terms of postoperative medication, regardless of study participation.

Participants were randomized to either the control or intervention group using a standard web-based random number generator with a range set from 1 to 10 and a 1:1 ratio by a

XSL•FO
RenderX

research assistant. Owing to the nature of this study, subjects and the enrolling research assistant were not blinded to the participant's study group following randomization.

At the time of consent, subjects were required to complete paper forms comprising a basic demographics questionnaire and baseline PROs consisting of the PROMIS Pain Intensity 1A Short form, PROMIS Pain Intensity 3A Short form, PROMIS Pain Interference 8A Short form, and PROMIS Emotional Distress-Anxiety 8A Short form (Multimedia Appendix 2). Following completion of all PROs, participants were randomized to their study group. Subjects who received an odd number from the 1 to 10 range set on the random number generator were placed in the intervention group, whereas subjects given an even number were placed in the control group.

Table 2. Injury by final study group (N=76).

Injury	Acceptance and commitment therapy group participants, n	Control group participants, n
Acetabular fracture	1	1
Ankle fracture	15	14
Calcaneus fracture	0	1
Clavicle fracture	0	1
Distal femur fracture	0	2
Distal humerus fracture	1	0
Elbow fracture	2	5
Femoral neck fracture	2	2
Femoral shaft stress fracture	0	1
Intertrochanteric hip fracture	1	0
Navicular fracture	1	0
Patella fracture	1	0
Polytrauma[a]	2	1
Proximal humerus fracture	2	1
Subtrochanteric femur fracture	0	2
Tibial plateau fracture	4	3
Tibial plafond fracture	6	4

[a]Polytrauma was defined as a patient with a fracture to more than one upper or lower extremity.

Textbox 1. Exclusion criteria.

Screening questions

- No personal mobile phone with text messaging capabilities
- Poor familiarity reading or sending mobile messages

Patient factors

- Open fracture
- Infection at the fracture site
- Prior fracture temporization with an external fixator
- Revision surgery for nonunion or hardware failure
- Bilateral upper extremity injuries impeding their ability to use a mobile phone
- Fractures of the distal hand or distal foot only
- Admission to an intensive care unit
- Current cancer diagnosis or dementia
- Inpatient for more than 7 days of the 2-week study period
- Discharged without an opioid pain medication prescription
- Initial plan for operative fixation changed to treatment with joint arthroplasty

XSL•FO
RenderX

Study Interventions

The intervention group received twice-daily, text-based mobile messages communicating an ACT-based intervention for the first 2 weeks following surgery (Multimedia Appendix 3). Control group subjects did not receive the ACT intervention or any other form of mobile message communication. The mobile messaging ACT protocol consisted of twice per day mobile messages, morning and evening, starting on postoperative day (POD) 1 and ending on POD 14. These mobile phone messages provided participants with an ACT-based intervention that was developed in collaboration with a pain psychologist (VK) specializing in ACT for chronic pain. These messages used all the principles presented in Table 1 with the objective of helping recipients understand and develop better coping skills in relation to their postoperative pain. An example message from day 1 is as follows:

> *Maintaining focus on what you value most in life is sometimes difficult after surgery. Do not let the momentary discomforts due to surgery take away from what you want most in life. Pick 3 things that matter most to you in life. Remind yourself of these 3 things you value most during your recovery process.*

Outside of the mobile messaging intervention, both groups received the same standard postoperative care, health care team communications, and instructions for completing the study follow-up.

A chart review was performed to collect demographic information such as subject age, comorbid conditions, and preoperative outpatient opioid medication prescriptions for treatment of their current traumatic orthopedic injury. All subjects, regardless of group, were seen by a research team member after surgery to review which of their discharge medications was the medication of interest for the study and to confirm that the intervention group subjects received their first mobile phone ACT message. Participants in both groups were instructed to have their opioid medication bottle available at follow-up to confirm their opioid tablet consumption. Owing to the changes in health care teams, staff preferences, and allergies, the opioid pain medications administered at discharge were not standardized between study groups. Following discharge on POD 14, subjects were contacted by phone or seen in the clinic by the research team for follow-up. At this time, the subjects' opioid pain medication consumption was assessed, and they completed a second set of PROs.

Outcome Measures

The primary outcome of this study was the amount of opioid pain medication consumed by subjects, and the secondary outcomes analyzed were net changes from baseline PRO scores at the 2-week follow-up.

The method that participants employed to report their opioid medication consumption and how PROs were captured during follow-up were recorded (Table 3). Subjects using their pill bottle to confirm the remaining number of opioid pain medication tablets from their discharge prescription on POD 14 were denoted as reporting a pill count. Cases where subjects or their care facility kept a log of tablet consumption were classified as reporting a daily log. Subjects reporting the number of tablets they used without the use of a log or pill count were designated as providing an estimate. The percentage of opioid pain medication used, total MME, and percentage of available MME consumed were calculated. The mean number of opioid pain medication tablets and MME used by the subjects were compared between groups. The raw scores for PROs were converted to corresponding *t*-scores using the appropriate PROMIS scoring manual [30]. The changes in PROs from baseline to POD 14 were also calculated by subtracting POD 14 scores from baseline scores, as higher *t*-scores signify a poorer outcome; thus, lower scores on POD 14 indicate an improvement from baseline PROs. The mean PRO scores and changes were compared between the groups.

XSL•FO
RenderX

Table 3. Comparison of subject demographics by enrolled study group.

Subject characteristic	Acceptance and commitment therapy group (n=42)	Control group (n=40)	*P* value
Age (years), mean (SD)	45.5 (15.9)	48.7 (14.6)	.41
BMI (kg/m^2), mean (SD)	30.5 (7.3)	31.1 (8.3)	.94
Sex, n (%)			.65
Female	22 (52)	19 (48)	
Male	20 (48)	21 (52)	
Subjects removed or lost to follow-up, n (%)	4 (10)	2 (5)	N/A[a]
Preoperative PROMIS[b] Pain Intensity 1A Score, mean (SD)	5.4 (2.9)	6.2 (2.6)	.20
Preoperative PROMIS Pain Intensity 3A Score, mean (SD)	54.9 (7.3)	57.1 (8.2)	.23
Mean Preoperative PROMIS Pain Interference 8A Score, mean (SD)	63.6 (11.4)	66.1 (8.4)	.30
Mean Preoperative PROMIS Emotional Distress-Anxiety 8A Score, mean (SD)	56.5 (11.4)	56.5 (9.2)	.99
Days between injury and surgery, mean (range)	4 (1-33)	3 (1-50)	.26
Disposition[c], n (%)			.68
Home	36 (95)	34 (90)	
Skilled nursing facility or acute rehabilitation	2 (5)	4 (10)	
Ethnicity/race, n (%)			.86
White	37 (88)	35 (88)	
African American	4 (10)	4 (10)	
Asian	1 (2)	0 (0)	
Hispanic	0 (0)	1 (2)	
Preoperative outpatient opioid prescription, n (%)	23 (55)	17 (43)	.17
Current psychiatric diagnosis, n (%)	15 (36)	9 (23)	.14
History/current substance abuse diagnosis, n (%)	8 (19)	3 (8)	.10
Diabetes diagnosis, n (%)	2 (5)	7 (18)	.15
Current smoker, n (%)	7 (17)	9 (23)	.57
Current lumbago diagnosis, n (%)	1 (2)	2 (5)	>.99
History of/current chronic pain diagnosis, n (%)	10 (24)	8 (20)	.59
Number of opioid tablets prescribed[c], mean (SD)	58.8 (27.3)	61.6 (22.0)	.62
Opioid utilization reporting method[c], n (%)			.47
Pill count	34 (90)	30 (79)	
Daily log	3 (8)	6 (16)	
Estimate	1 (2)	2 (5)	
Patients filling only one postoperative opioid prescription[c], n (%)	34 (90)	34 (90)	>.99

[a]N/A: not applicable.

[b]PROMIS: Patient-Reported Outcome Measures Information System.

[c]Data calculated using final study population only (n=38).

XSL•FO
RenderX

Statistical Analysis

Participant characteristics were described using mean (SD) or median (minimum to maximum) for continuous variables and frequencies (percentages) for categorical variables. Visual review of histograms and the results of the Shapiro-Wilk test of continuous variables revealed that only age and BMI were not normally distributed. Between-group differences were evaluated using *t* tests or Wilcoxon rank-sum tests (age and BMI) for continuous variables and chi-square or exact tests for categorical variables, as appropriate.

To evaluate whether the intervention versus control group had a lower opioid use on average, we determined the number of tablets and MME taken in each group and compared means using *t* tests. Using a previous study of opioid medication usage in orthopedic trauma patients [4], the sample size estimated to observe a 30% decrease in opioid utilization among 2 groups required a total of 74 subjects to achieve 80% power at an alpha of .05. The percent decrease is calculated using the formula . A separate power analysis was calculated for the PRO portion of the study, and it was determined that a total of 36 subjects would provide 80% power to detect a 10-point difference (1 standard deviation) in *t*-scores for the PROMIS instruments at an alpha level of .05. Statistical analyses were performed using SAS software version 9.4 (SAS Institute, Inc).

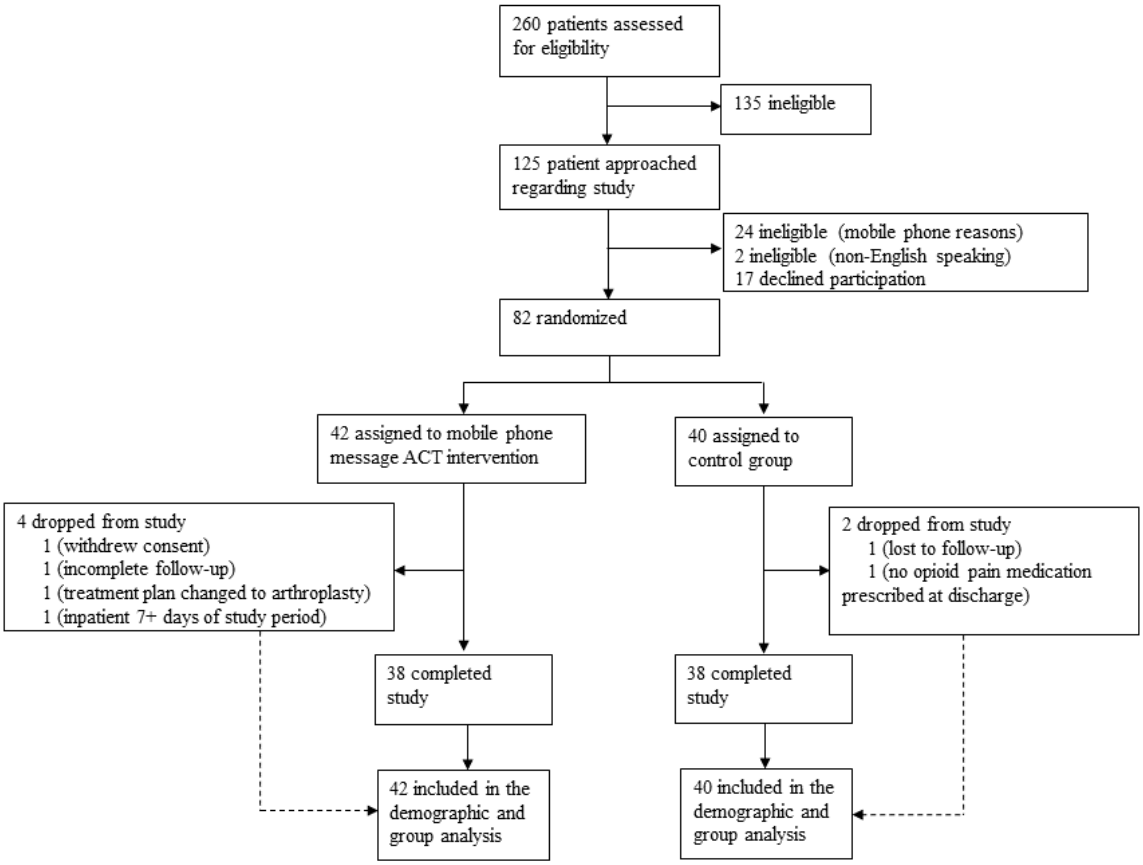
Results

Study Participants

A total of 125 individuals were approached regarding the study over the 5-month enrollment period between February 2019 and June 2019. Of the 125 individuals, 2 patients were excluded at this time, as they were non–English-speaking, and an additional 24 patients were excluded because they did not use mobile phone messaging or did not have a personal mobile phone. This resulted in a total of 99 eligible people who were presented the study, 17 of whom declined participation (Figure 1). Overall, 82 subjects were enrolled, and 6 dropped from the study after providing consent because of various issues: one patient lost to follow-up, one patient withdrew at follow-up, one patient had incomplete follow-up, one patient's operative plan changed to arthroplasty, and 2 subjects remained inpatient for over 7 days of the study period (Figure 1). This resulted in a final population of 76 subjects (38 per study group). The enrollment period concluded once a powered sample for the primary aim was obtained. A breakdown of the subjects enrolled, reasons for excluding subjects, and subjects removed from the study after consent are presented in Figure 1. Participant demographics for the intervention and control groups are presented in Table 3. The analyses of all collected demographic factors showed no differences between the intervention and control groups in all factors such as subject age (*P*=.42), current psychiatric diagnosis (including depression, anxiety, bipolar type 1, obsessive-compulsive disorder, posttraumatic stress disorder, panic disorder, and attention-deficit disorder; *P*=.14), or substance abuse history (*P*=.10; Table 3). Furthermore, no differences between groups were found for injury type, disposition following discharge, method for reporting opioid medication consumption, preoperative opioid medication prescriptions, or preoperative PROs (Table 3).

XSL•FO
RenderX

Figure 1. Consolidated Standards of Reporting Trials flowchart detailing the selection of eligible patients for study enrollment and their status through study completion. ACT: acceptance and commitment therapy.



Opioid Pain Medication Use

No differences between groups were observed in the amount of opioid medication tablets or MME prescribed at discharge (tablets for the ACT group: mean 58.8, SD 27.3 vs tablets for the control group: mean 61.6, SD 22.0). A further breakdown of the medications prescribed to subjects within the study period is presented in Table 4. The mean opioid tablet use for subjects in the ACT-based intervention group was 26.1 (SD 21.4) tablets, whereas the control group used a mean of 41.1 (SD 22.0) tablets, resulting in 36.5% less tablets used by subjects receiving the ACT-based intervention (*P*=.004; Table 5). Similarly, subjects in the intervention group consumed a mean of 199.9 (SD 163.2) MME, on average, whereas the control group subjects consumed a mean of 307.0 (SD 166.0) MME, indicating 34.9% less MME used by subjects in the intervention group (*P*=.006; Table 5).

Table 4. Frequency of outpatient opioid pain medications prescribed by enrolled study group (N=82).

Medication	Morphine milliequivalents per tablet	Frequency	
		Acceptance and commitment therapy	Control
Hydrocodone-acetaminophen 5-325 mg	5	1	2
Hydrocodone-acetaminophen 10-325 mg	10	2	0
Hydromorphone 2 mg	8	6	6
Oxycodone 5 mg	7.5	27	27
Oxycodone-acetaminophen 5-325 mg	7.5	10	8
Total opioid prescriptions provided	N/A[a]	46	43

[a]N/A: not applicable.

XSL•FO
RenderX

Table 5. Opioid pain medication utilization by group during the 2-week study period.

Attribute	Opioid tablets dispensed			Opioid tablets consumed				Morphine milliequivalents consumed			
	ACT[a] (n=38)	Control (n=38)	*P* value	ACT (n=38)	Control (n=38)	Decrease[b] (%)	*P* value	ACT (n=38)	Control (n=38)	Decrease (%)	*P* value
Mean (SD)	58.8 (27.3)	61.6 (22.0)	.62	26.2 (21.4)	41.1 (22.0)	37	.004	199.9 (163.2)	307.0 (166.0)	35	.006
Median (minimum-maximum)	60.0 (10-146)	60.0 (15-120)	.62	21.0 (0-80)	43.5 (0-80)	37	.004	157.5 (0-600)	307.5 (0-600)	35	.006

[a]ACT: acceptance and commitment therapy.

[b]Calculated by the formula .

Patient-Reported Outcomes

PROMIS instrument *t*-score values for both ACT and control group subjects are presented in Table 6. At 2-week follow-up, the intervention group subjects reported lower postoperative PROMIS Pain Intensity 3A (mean 45.9, SD 7.2) and Pain Interference 8A (mean 56.6, SD 9.4) scores compared with the control group's postoperative Pain Intensity 3A (mean 49.7, SD 8.8; *P*=.04) and Pain Interference 8A scores (mean 60.6, SD 8.2; *P*=.05; Table 6). No differences were observed between groups at 2-week follow-up in the PROMIS Pain Intensity 1A or PROMIS Emotional Distress-Anxiety 8A forms (Table 6).

Table 6. Mean Patient-Reported Outcome Measures Information System score and change within the 2-week study period by study group.

PROMIS[a] instrument	Preoperative score				Postoperative score					Net score change				
	ACT[b]		Control		ACT		Control		*P* value	ACT		Control		*P* value
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range		Mean (SD)	Range	Mean (SD)	Range	
Pain Intensity 1A[c]	5.4 (2.9)	0 to 10	6.2 (2.6)	1 to 10	3.4 (2.2)	0 to 9	4.1 (2.4)	1 to 9	.22	−2.0 (2.9)	−10 to 7	−2.1 (2.3)	−9 to 2	.79
Pain Intensity 3A	54.9 (7.3)	36.3 to 71.8	57.1 (8.2)	40.2 to 71.8	45.9 (7.2)	30.7 to 64.1	49.7 (8.8)	30.7 to 67.4	.04	−9.0 (8.5)	−25.5 to 10	−7.4 (7.7)	−23.9 to 6.1	.38
Pain Interference 8A	63.6 (11.4)	40.7 to 77	66.1 (8.4)	40.7 to 77.0	56.6 (9.4)	40.7 to 72.0	60.6 (8.2)	40.7 to 77.0	.048	−7.1 (13.7)	−36.3 to 24.8	−5.4 (10.4)	−26.2 to 19.5	.55
Emotional Distress-Anxiety 8A	56.5 (11.4)	37.1 to 80	56.5 (9.2)	37.1 to 76.7	51.5 (10.4)	37.1 to 75.4	52.3 (10.6)	37.1 to 76.7	.76	−4.9 (10.1)	−33.7 to 12.3	−4.2 (9.4)	−20.3 to 16.5	.74

[a]PROMIS: Patient-Reported Outcome Measures Information System.

[b]ACT: acceptance and commitment therapy.

[c]Scores presented are raw numerical scores, as no *t*-score conversion is available for the selected instrument.

Discussion

Principal Findings

This randomized trial delivered an ACT-based intervention via an automated mobile messaging robot to postoperative orthopedic patients. The subjects who received the ACT-based mobile phone intervention used a lower number of opioid tablets and consumed less MME in the first 2 weeks after their injury. We also found that the intervention group reported less pain intensity and pain interference at the 2-week follow-up. These data demonstrate that ACT-based automated mobile messaging protocols may be effective in reducing the amount of opioid medication used and may positively affect postoperative PROs in patients undergoing operative fixation of their acute fractures.

Effects on Opioid Use

Improved mood symptoms, less pain interference, and faster cessation of opioid pain medication are some of the recognized benefits of using ACT in clinic-based, interdisciplinary approaches to pain management after surgery [17,18]. Previous investigations have used automated mobile phone messaging robots to deliver PROs [20,23], improve communication with patients [27,31], deliver postoperative orthopedic care [32], and inquire about pain and opioid utilization [4,19,33]. In this study, we used ACT and a mobile phone messaging robot to assess whether these tools in combination could decrease opioid utilization and improve individuals' perception of their early recovery from injury. Prior work has demonstrated a quicker time to opioid cessation and a decrease in postoperative opioid utilization (14% less in the ACT group) when used in office ACT-based treatments [17,18]. Subjects receiving the ACT

XSL•FO
RenderX

intervention via automated mobile phone messages reported over 36% less opioid tablets and more than 34% less MME consumed than corresponding control subjects who did not receive ACT. Our findings suggest that software-based communication using ACT through a mobile phone has the potential to have a large impact on the utilization of postoperative pain medication by patients in the first weeks after surgery for fractures. Further study is required to determine if these effects are long lasting and to determine which injuries and patients receive the greatest benefit. In addition, future investigations and trials should consider the effect of software delivery of ACT and other behavioral therapies on different cohorts of patients.

Effects on PROs

PROs, such as PROMIS, allow patients to quantify aspects of their orthopedic condition in a standardized fashion [6,7]. These are important tools for determining the efficacy of health care treatments and assessment of clinical research and can be used in determining compensation for health care services provided [6-8]. The National Institutes of Health developed the PROMIS tools to advance PROs by creating question banks that could be used for many major health issues [9]. We found that despite less utilization of opioids, subjects in the ACT-based intervention group reported less pain intensity and pain interference at 2 weeks. This most likely does not represent a clinically important difference based on the SD methodology used in prior works with PROMIS tools (Table 6), but it at least suggests that the intervention group did not experience greater pain [34]. There were no other differences between study groups in the other domains at the 2-week follow-up. Previous studies have reported that patients who consume more opioid medication report higher pain at both short- and long-term time points, which is reflected in our findings for both PROMIS pain intensity 3A and the employed pain interference measure [35,36]. Future research efforts may benefit from employing alternative PRO measures to identify the effects of ACT-based interventions, including assessment of psychologic flexibility. Future research may also consider possible modifications of our study protocol to include a longer intervention period and more than one follow-up data point. Future work may also consider designing an ACT-based tool that is more focused on demonstrating an effect on PROs.

Limitations

Several limitations were present in this study. First, we were limited to a single level 1 trauma center, which may affect the reproducibility of our results across other health care settings. Next, the exclusion criteria for this study were extensive, and thus, the results may not be generalizable to the entire scope of orthopedic trauma patients. We attempted to include a diverse set of injuries and yet excluded patients with a high likelihood of confounding problems from open fractures or prolonged initial hospitalization. Future studies assessing the effects of ACT-based interventions similar to ours should aim for less restrictive exclusion criteria to apply this intervention to a larger, more diverse population. The research assistants were not blinded to the patients' study group. In addition, patients understood the outcomes of interest in this study, which could be susceptible to reporting bias. In addition, participants were not blinded to their treatment group. The lack of blinding could potentially introduce response or reporting bias, making this a potential area of improvement for studies seeking to follow the present methodology. This could be accomplished through the implementation of a control messaging protocol. Moreover, a retrospective chart review was used to obtain several patient factors, including comorbid conditions and dispensing of preoperative outpatient opioid medication prescriptions. The collection of information in this manner relies on accurate charting and transfer of documents from outside institutions, which may have been incomplete.

Conclusions

In this study, delivering an ACT-based intervention via an automated mobile messaging robot in the acute postoperative period decreased opioid utilization in orthopedic trauma patients in the first 2 weeks after their injury. Subjects in the ACT-based intervention group also reported lower pain intensity and pain interference after 2 weeks, although this likely did not represent a clinically important difference. Future studies may apply this intervention in other patient populations to assess its efficacy on a larger scale and may include assessment of pain and opioid use in a longer time frame after injury.

Acknowledgments

This study was possible thanks to a generous grant from the Orthopaedic Trauma Association.

Conflicts of Interest

CA reports personal fees from McKinsey & Company, outside the submitted work. MW reports nonfinancial support from Zimmer Biomet, outside the submitted work. MK reports stock or stock options from Iowa Simulation Solutions LLC and stock or stock options from Mortise Medical LLC, outside the submitted work. JM reports stock or stock options from Zimmer Biomet, stock or stock options from FxRedux, nonfinancial support and stock or stock options from Oxford Press, and stock or stock options from Tornier, outside the submitted work.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1617 KB - jmir_v22i7e17750_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Patient-Reported Outcome Measure Information System tools completed by the study subjects.
[PDF File (Adobe PDF File), 97 KB - jmir_v22i7e17750_app2.pdf]

Multimedia Appendix 3
Acceptance and commitment therapy–based automated mobile phone messaging protocol.
[PDF File (Adobe PDF File), 44 KB - jmir_v22i7e17750_app3.pdf]

References

1. Azar A. Renewal of Determination That A Public Health Emergency Exists. US Department of Health and Human Services. 2019. URL: https://www.phe.gov/emergency/news/healthactions/phe/Pages/covid19-21apr2020.aspx [accessed 2000-07-09]
2. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths - United States, 2013-2017. MMWR Morb Mortal Wkly Rep 2018 Jan 4;67(5152):1419-1427 [FREE Full text] [doi: 10.15585/mmwr.mm675152e1] [Medline: 30605448]
3. Brat GA, Agniel D, Beam A, Yorkgitis B, Bicket M, Homer M, et al. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. Br Med J 2018 Jan 17;360:j5790 [FREE Full text] [doi: 10.1136/bmj.j5790] [Medline: 29343479]
4. Anthony CA, Volkmar A, Shah AS, Willey M, Karam M, Marsh JL. Communication with orthopedic trauma patients via an automated mobile phone messaging robot. Telemed J E Health 2018 Jul;24(7):504-509. [doi: 10.1089/tmj.2017.0188] [Medline: 29261036]
5. Scully RE, Schoenfeld AJ, Jiang W, Lipsitz S, Chaudhary MA, Learn PA, et al. Defining optimal length of opioid pain medication prescription after common surgical procedures. JAMA Surg 2018 Jan 1;153(1):37-43 [FREE Full text] [doi: 10.1001/jamasurg.2017.3132] [Medline: 28973092]
6. Deshpande PR, Rajan S, Sudeepthi BL, Nazir CP. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
7. Ahmed S, Berzon RA, Revicki DA, Lenderking WR, Moinpour CM, Basch E, International Society for Quality of Life Research. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Med Care 2012 Dec;50(12):1060-1070. [doi: 10.1097/MLR.0b013e318268aaff] [Medline: 22922434]
8. Review of Physician and Advanced Practitioner Recruiting Incentives. Merritt Hawkins. 2019. URL: https://www.merritthawkins.com/trends-and-insights/article/reports/2019-review-of-physician-and-advanced-practitioner-recruiting-incentives/ [accessed 2020-07-09]
9. Ader DN. Developing the patient-reported outcomes measurement information system (PROMIS). Med Care 2007 May;45(Suppl 1):S1-S2 [FREE Full text] [doi: 10.1097/01.mlr.0000260537.45076.74]
10. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS adult health profiles: efficient short-form measures of seven health domains. Value Health 2019 May;22(5):537-544 [FREE Full text] [doi: 10.1016/j.jval.2019.02.004] [Medline: 31104731]
11. Batten S. Essentials of Acceptance and Commitment Therapy. London, UK: Sage Publications; 2011.
12. Hayes S. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies - republished article. Behav Ther 2016 Nov;47(6):869-885. [doi: 10.1016/j.beth.2016.11.006] [Medline: 27993338]
13. Hughes LS, Clark J, Colclough JA, Dale E, McMillan D. Acceptance and commitment therapy (ACT) for chronic pain: a Systematic review and meta-analyses. Clin J Pain 2017 Jun;33(6):552-568. [doi: 10.1097/AJP.0000000000000425] [Medline: 27479642]
14. Dahl J, Wilson K, Nilsson A. Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and pain symptoms: a preliminary randomized trial. Behav Ther 2004;35(4):785-801 [FREE Full text] [doi: 10.1016/S0005-7894(04)80020-0]
15. Powers MB, Vording MB, Emmelkamp PM. Acceptance and commitment therapy: a meta-analytic review. Psychother Psychosom 2009;78(2):73-80. [doi: 10.1159/000190790] [Medline: 19142046]
16. McCracken LM, Vowles KE. Acceptance and commitment therapy and mindfulness for chronic pain: model, process, and progress. Am Psychol 2014;69(2):178-187. [doi: 10.1037/a0035623] [Medline: 24547803]
17. Dindo L, Zimmerman MB, Hadlandsmyth K, StMarie B, Embree J, Marchman J, et al. Acceptance and commitment therapy for prevention of chronic postsurgical pain and opioid use in at-risk veterans: a pilot randomized controlled study. J Pain 2018 Oct;19(10):1211-1221 [FREE Full text] [doi: 10.1016/j.jpain.2018.04.016] [Medline: 29777950]
18. Abid Azam M, Weinrib AZ, Montbriand J, Burns LC, McMillan K, Clarke H, et al. Acceptance and commitment therapy to manage pain and opioid use after major surgery: preliminary outcomes from the Toronto general hospital transitional pain service. Can J Pain 2017 Jun 28;1(1):37-49. [doi: 10.1080/24740527.2017.1325317]
19. Anthony CA, Lawler EA, Ward CM, Lin IC, Shah AS. Use of an automated mobile phone messaging robot in postoperative patient monitoring. Telemed J E Health 2018 Jan;24(1):61-66. [doi: 10.1089/tmj.2017.0055] [Medline: 28622079]

XSL•FO
RenderX

20. Anthony CA, Lawler EA, Glass NA, McDonald K, Shah AS. Delivery of patient-reported outcome instruments by automated mobile phone text messaging. Hand (N Y) 2017 Nov;12(6):614-621 [FREE Full text] [doi: 10.1177/1558944716672201] [Medline: 29091492]
21. Anthony CA, Peterson AR. Utilization of a text-messaging robot to assess intraday variation in concussion symptom severity scores. Clin J Sport Med 2015 Mar;25(2):149-152. [doi: 10.1097/JSM.0000000000000115] [Medline: 24905538]
22. Liu TC, Ohueri CW, Schryver EM, Bozic KJ, Koenig KM. Patient-identified barriers and facilitators to pre-visit patient-reported outcomes measures completion in patients with hip and knee pain. J Arthroplasty 2018 Mar;33(3):643-9.e1. [doi: 10.1016/j.arth.2017.10.022] [Medline: 29169689]
23. Scott EJ, Anthony CA, Rooney P, Lynch TS, Willey MC, Westermann RW. Mobile phone administration of hip-specific patient-reported outcome instruments correlates highly with in-office administration. J Am Acad Orthop Surg 2020 Jan 1;28(1):e41-e46. [doi: 10.5435/JAAOS-D-18-00708] [Medline: 31860543]
24. Anthony C, Polgreen L, Chounramany J, Foster E, Goerdt C, Miller M, et al. Outpatient blood pressure monitoring using bi-directional text messaging. J Am Soc Hypertens 2015 May;9(5):375-381 [FREE Full text] [doi: 10.1016/j.jash.2015.01.008] [Medline: 25771023]
25. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mhealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
26. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
27. Day MA, Anthony CA, Bedard NA, Glass NA, Clark CR, Callaghan JJ, et al. Increasing perioperative communication with automated mobile phone messaging in total joint arthroplasty. J Arthroplasty 2018 Jan;33(1):19-24. [doi: 10.1016/j.arth.2017.08.046] [Medline: 29017803]
28. Jones KR, Lekhak N, Kaewluang N. Using mobile phones and short message service to deliver self-management interventions for chronic conditions: a meta-review. Worldviews Evid Based Nurs 2014 Apr;11(2):81-88. [doi: 10.1111/wvn.12030] [Medline: 24597522]
29. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
30. PROMIS® Scoring Manuals. Health Measures. URL: http://www.healthmeasures.net/index.php?option=com_content&view=article&id=180&Itemid=994) [accessed 2020-07-09]
31. Goz V, Anthony C, Pugely A, Lawrence B, Spina N, Brodke D, et al. Software-based postoperative communication with patients undergoing spine surgery. Global Spine J 2019 Feb;9(1):14-17 [FREE Full text] [doi: 10.1177/2192568217728047] [Medline: 30775203]
32. Tofte JN, Anthony CA, Polgreen PM, Buckwalter JA, Caldwell LS, Fowler TP, et al. Postoperative care via smartphone following carpal tunnel release. J Telemed Telecare 2020 May;26(4):223-231. [doi: 10.1177/1357633X18807606] [Medline: 30428766]
33. Nelson SE, Adams AJ, Buczek MJ, Anthony CA, Shah AS. Postoperative pain and opioid use in children with supracondylar humeral fractures: balancing analgesia and opioid stewardship. J Bone Joint Surg Am 2019 Jan 16;101(2):119-126. [doi: 10.2106/JBJS.18.00657] [Medline: 30653041]
34. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. Arthritis Care Res (Hoboken) 2013 Oct;65(10):1625-1633 [FREE Full text] [doi: 10.1002/acr.22025] [Medline: 23592494]
35. Bot AG, Bekkers S, Arnstein PM, Smith RM, Ring D. Opioid use after fracture surgery correlates with pain intensity and satisfaction with pain relief. Clin Orthop Relat Res 2014 Aug;472(8):2542-2549 [FREE Full text] [doi: 10.1007/s11999-014-3660-4] [Medline: 24777731]
36. Koehler RM, Okoroafor UC, Cannada LK. A systematic review of opioid use after extremity trauma in orthopedic surgery. Injury 2018 Jun;49(6):1003-1007. [doi: 10.1016/j.injury.2018.04.003] [Medline: 29704954]

Abbreviations

ACT: acceptance and commitment therapy
MME: morphine milliequivalents
POD: postoperative day
PRO: patient-reported outcome
PROMIS: Patient-Reported Outcome Measure Information System

XSL•FO
RenderX

Edited by G Eysenbach; submitted 09.01.20; peer-reviewed by L Dindo, H Zhao; comments to author 05.03.20; revised version received 29.04.20; accepted 20.05.20; published 29.07.20.

Please cite as:
Anthony CA, Rojas EO, Keffala V, Glass NA, Shah AS, Miller BJ, Hogue M, Willey MC, Karam M, Marsh JL
Acceptance and Commitment Therapy Delivered via a Mobile Phone Messaging Robot to Decrease Postoperative Opioid Use in Patients With Orthopedic Trauma: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e17750
URL: https://www.jmir.org/2020/7/e17750
doi:10.2196/17750
PMID:32723723

©Chris A Anthony, Edward Octavio Rojas, Valerie Keffala, Natalie Ann Glass, Apurva S Shah, Benjamin J Miller, Matthew Hogue, Michael C Willey, Matthew Karam, John Lawrence Marsh. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reactions to an Online Demonstration of the Effect of Increased Fruit and Vegetable Consumption on Appearance: Survey Study

Patrick Cairns[1], MA, MSc; Gozde Ozakinci[2], PhD, CPsychol; David Ian Perrett[1], BSc, DPhil

[1]Perception Lab, School of Psychology & Neuroscience, University of St Andrews, St Andrews, United Kingdom

[2]School of Medicine, University of St Andrews, St Andrews, United Kingdom

Corresponding Author:
Patrick Cairns, MA, MSc
Perception Lab
School of Psychology & Neuroscience
University of St Andrews
St Mary's Quad
South Street
St Andrews, KY16 9JP
United Kingdom
Phone: 44 1334 463044
Email: pc74@st-andrews.ac.uk

Abstract

Background: Inadequate fruit and vegetable consumption causes a considerable disease burden and premature mortality. Despite public health promotion of a healthy diet, the average consumption is still below recommended levels. Fruit and vegetable consumption influences human skin color, increasing red/yellow/orange pigment in the skin. Given that this color is deemed attractive and healthy-looking, the appearance benefit may motivate to eat more fruit and vegetables. Such appearance motivation could be particularly useful in young individuals who currently eat the least fruit and vegetables.

Objective: Our objectives were to assess how widely the impact of diet on skin color is known within the UK and to compare the strength of motivation to eat fruit and vegetables based on health and appearance benefits among different demographic groups.

Methods: Four groups of UK residents (N=200 per group) were recruited through the Prolific online platform. Groups comprised younger (aged 18-24) and older adults (aged 40-60) of low and high self-reported socioeconomic status (1-5 and 6-10 on a 10-point rating scale). Facial images simulating the skin color associated with low and high fruit and vegetable diets were shown to participants. Questionnaires were used to assess (1) background knowledge of the health and skin color effects of dietary fruit and vegetables, (2) the specific motivational impact of the skin color illustration, and (3) the relative importance of motivation to consume fruit and vegetables arising from health and skin color appearance benefits.

Results: We found that 61% (n=487) of all participants were unaware of the dietary–skin color association. We also found that 57% (n=457) of participants found the simple demonstration of the dietary impact on skin color positively motivating to eat more fruit and vegetables. The visual demonstration was equally motivating for participants of high and low self-reported socioeconomic status (P=.63) and different ethnic backgrounds (White N=453, Black N=182, Asian N=87, P=.22). Health benefits from a diet high in fruit and vegetables were regarded as more motivating than skin color appearance benefits. The appearance-changing benefits of a high fruit and vegetable diet (compared to the health benefits) were relatively more important for the younger participants (Mann-Whitney U=96,263, P<.001) and for women (N=489) than for men (N=310, U=83,763, P=.01).

Conclusions: These findings indicate that the promotion of the skin color effects of diets high in fruit and vegetables could provide additional motivation for a healthier diet. Our study indicates the broad appeal of appearance benefits from dietary fruit and vegetable (across ethnicity and socioeconomic status) and particularly amongst young adults where an inadequate diet is most prevalent.

(J Med Internet Res 2020;22(7):e15726) doi:10.2196/15726

KEYWORDS

diet; skin appearance; motivations; fruit and vegetables; carotenoid

XSL•FO
RenderX

Introduction

Background

Inadequate fruit and vegetable consumption is estimated to lead to between 5.6 to 7.8 million premature deaths per year worldwide [1], chiefly through incidences of cardiovascular disease including coronary heart disease and stroke [1-3], diabetes and its complications [4,5] and several cancers [1,2,6]. Globally inadequate intake of fruit and vegetables is also responsible for the loss of up to 103 million years of healthy life due to disability [7]. The negative ramifications of these lifestyle-attributable diseases are widely felt. In addition to the consequences for personal wellbeing, poor population health contributes to an overburdening of healthcare systems and fiscal strain due to lost productivity [8].

Only 29% of adults in the UK [9] report eating the recommended 5 portions of fruit and vegetables per day. Fewer men (26%) than women (32%) meet the 5 a day guideline. Young people aged 16-24 were also less likely than other adults aged 45-60 to get their 5 a day (average 3.3 portions per day compared to 3.8). Figures are worse for children, with only 18% of children aged 5-15 eating 5 portions per day. The National Diet and Nutrition Survey showed that the proportion of adults (~30%) consuming the 5 a day recommendation has changed little over the last decade [10]. Residential area deprivation and lower socioeconomic status independently predict decreased fruit and vegetable consumption [11,12].

Health promotion efforts to increase fruit and vegetable consumption vary from individual approaches such as personal (or parental) advice and counseling to public health campaigns. Although much has been achieved in terms of understanding what psychological techniques may work or not to facilitate healthy eating [13], given the current state of inadequate consumption of fruit and vegetables, we are still in need of novel and innovative methods in addition to existing ones to encourage healthy eating.

Skin Color Effects

Fruit and vegetable consumption influences human skin color, increasing the presence of red/yellow carotenoid pigments in the skin [14-18]. Indeed, this skin color is a reliable biomarker of fruit and vegetable and carotenoid intake [19-22]. When asked to choose the skin color that looks healthiest in photographs of themselves or others, participants consistently choose a skin color that represents a higher fruit and vegetable consumption [14,23-25]. Likewise, when asked to choose the most attractive facial photograph, participants choose the one with increased carotenoid skin coloration over baseline coloration [26] and even over increased suntan coloration [27,28]. Some reports point to a limitation of the effect of carotenoid skin color. Appleton et al [29] found that carotenoid color did not affect attractiveness when pose and expression were unconstrained, and Tan et al [24] found that while a subtle level of carotenoid color increased attractiveness a large amount was deemed unattractive in a Malaysian Chinese population.

In terms of using the effect of fruit and vegetable consumption on skin color to motivate increased fruit and vegetable consumption, limited trials have been conducted. Whitehead et al [30] found that those shown a personalized demonstration of the potential skin color improvements arising from increased fruit and vegetable consumption reported an increase in fruit and vegetable consumption at a 10-week follow-up.

The present study aimed to assess the extent to which the effects of fruit and vegetable consumption on skin color are known within the UK. After extensive media coverage, we expected that the effect of fruit and vegetable consumption on skin color would be known by a proportion of the population, but the demographics of those familiar with the effect are unclear.

Secondly, we aimed to assess the extent to which UK residents are motivated to eat fruit and vegetables by a simple demonstration of the effects of fruit and vegetable consumption on skin color. Given the previous studies, we expected that a majority of individuals would express positive motivation following the online demonstration.

Specific subsections of society may be more receptive to appearance-based incentives than others. Within the US, Hayes and Ross [31] found younger participants to be more concerned about their appearance than older participants. Women were also more concerned about their appearance than men [31]. Across all demographic groups, the appearance was a powerful motivation for healthy eating [31]. By analyzing questionnaire data from 236 college-age women, Chung et al [32] found appearance (particularly weight maintenance) to be an essential factor in women's decisions to eat fruit and vegetables. It was, therefore, predicted that younger and female participants would show the greatest motivation to increase fruit and vegetable consumption following the online demonstration of appearance changes related to diet. While it is clear that low socioeconomic status is predictive of reduced fruit and vegetable consumption [11,12], it is unclear whether socioeconomic status relates to the dietary motivation from skin color change.

Methods

Participant Enrollment

Ethical approval was obtained from the University of St Andrews School of Psychology Ethics Committee (PS13092). The study was completed with 802 participants recruited via Prolific, an online UK-based recruitment platform for surveys and behavioral experiments [33].

The Prolific company supplies sociodemographic information for their panelists. Prolific asks recruits to answer the following question: "Think of a ladder (image of a ladder with 10 rungs) as representing where people stand in society. At the top of the ladder are the people who are best off—those who have the most money, most education, and the best jobs. At the bottom are the people who are worst off—who have the least money, least education and the worst jobs or no job. The higher up you are on this ladder, the closer you are to people at the very top, and the lower you are, the closer you are to the bottom. Where would you put yourself on the ladder? Choose the number (1-10) whose position best represents where you would be on this ladder."

XSL•FO
RenderX

Participants were recruited in four waves: (1) 202 18-26 year-olds who rated themselves as belonging to the lower half of self-perceived socioeconomic status (rating of 1-5 on the 10-point visual analog ladder scale), (2) 199 18-26 year-olds who rated themselves as belonging to the upper half of self-perceived socioeconomic status (6-10 rating on the 10 point scale), (3) 200 40-60 year-olds who rated themselves as belonging to the lower half of self-perceived socioeconomic status, and (4) 201 40-60 year-olds who rated themselves as belonging to the upper half of self-perceived socioeconomic status.

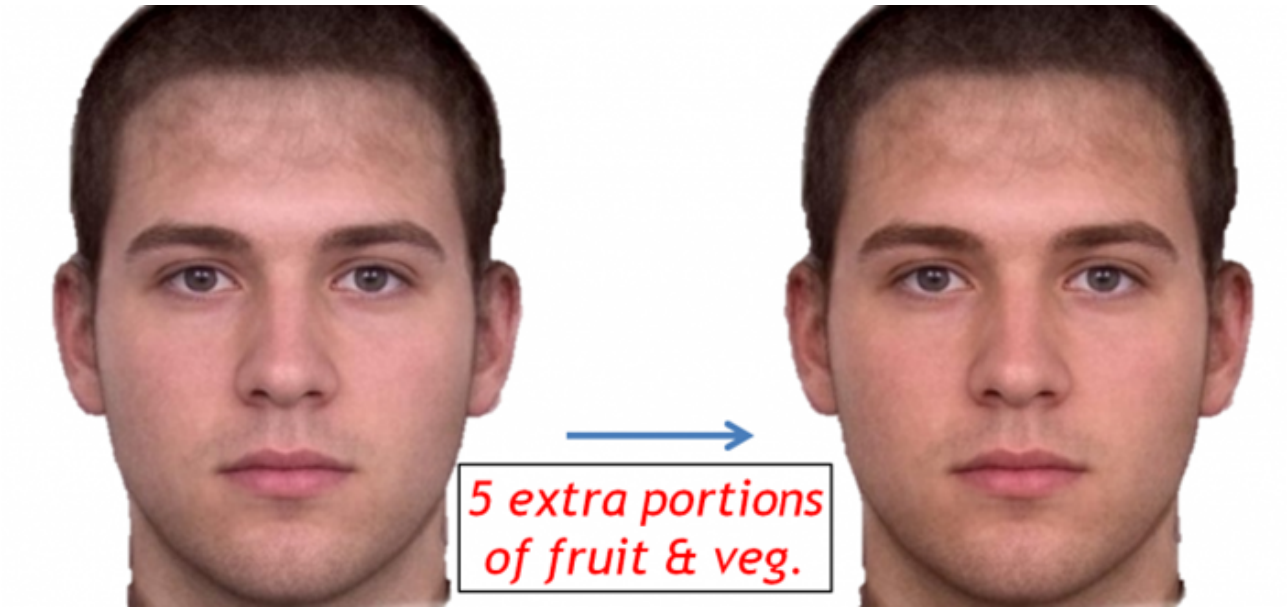
Participants were paid £0.6 for taking part (at an average rate of £9 per hour, 80% above stipulated minimum rate), with the questionnaire taking approximately 4 minutes to complete.

Image Transforms

Participants were shown before and after images of a male face (Figure 1). The left-hand image was a composite image made by combining 13 images of men who had consented to their image being published. This image was then graphically changed to represent the skin color changes that occur following the consumption of 5 portions of fruit and vegetables for 6 weeks. Image transformation to simulate the effect of increasing dietary fruit and vegetables followed methods outlined previously [14,23,24,27]. This single image pair was used to make the demonstration as simple as possible.

The color change in the presented image pair was measured by selecting a large patch of the image of facial skin between the eyes and neck and analyzing the average color values of the pixels within the image patch. The right-hand image in Figure 1 was darker, redder, and yellower than the left image. The color difference between right and left images in CIELab color space was L*a*b*=-1.5, 2.0, 4.1. The displayed color difference is equivalent to the skin color change (L*a*b*=-1.8, 1.2, 3.8) after 4 weeks of a 500 mL/day dietary smoothie supplement of 6 extra daily portions of fruit and vegetables [15]. The color difference we use is one-quarter of the difference used in other studies [27,34].

Figure 1. The image shown to participants before answering questions about the skin color benefits of fruit and vegetable consumption.



Procedure

Participants were first shown an image (Figure 1) showing the effect of fruit and vegetable consumption on skin color benefits. Participants were then given a questionnaire asking demographics (gender, age, ethnicity) and how many portions of fruit and vegetables they eat per day (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or more). They were also asked whether they were aware that diet affected skin color appearance and health. “Before seeing this display, were you aware that eating fruit and vegetables (a) can impart a golden glow to your skin color? (b) may help reduce the risk of the two main killer diseases in the UK—heart disease and some cancers?”. Answer options were: no, somewhat, yes (rescored as 0, 1, 2).

One question assessed the motivational effect of the display: “Has seeing the effect of diet on skin color in this exhibit made you want to eat less or more fruit and vegetables?”. Answer options were: a lot less, a little less, no change, a little more, a lot more.

Two further questions assessed the relative motivation of participants: “How much would the following make you want to eat more fruit and vegetables: (a) getting a skin color that looks healthy and attractive (within 20 days), (b) reducing your chance of heart disease and some forms of cancer?” Answer options were: none at all, a little, a moderate amount, a lot, a great deal.

Analysis

Non-parametric statistics were used for ordinal data (awareness and motivation), and fruit and vegetable consumption as skew departed from normality. Mann-Whitney U tests were used to compare 2 groups, Kruskal-Wallis, to compare more than 2 groups and Wilcoxon Signed Rank test T to explore choice across a single group. Three of the 802 (<1%) participants reported consuming 11 or more portions of fruit and vegetables

XSL•FO
RenderX

per day. Therefore, consumption data was approximated as an interval scale ranging from 0 to 11.

Missing Data

For each variable, the analysis was restricted to participants with data either through their registration with Prolific (younger 18-26 vs older 40-60 age bracket, socioeconomic status) or through our questionnaire (age at the time of testing, gender, ethnicity). Data were missing for 36 participants (5%) for age at the time of testing, 20 (3%) for ethnicity, and 12 (1%) for socioeconomic status. Descriptions of the study population in terms of gender, age, and ethnicity are based on answers to our questionnaire. No attempt was made to compensate for missing data.

Results

Participant Demographics

Gender

Participants included 489 females (61%) and 311 males (39%). Two individuals reported their gender as "Other."

Age

In our questionnaire, 381 of the 401 younger participants (in the age bracket 18-26) reported age at the time of testing age (mean 21.57, SD 2.09, range 18-30 years). Of the 401 older participants in the 40-60 age bracket, 385 reported age at the time of testing (mean 48.7, SD 5.83, range 38-61 years).

Socioeconomic Status

We observed 790 participants (1% missing data) who had Prolific categories of self-perceived socioeconomic status scale ranging from 1 (the lowest) to 10 (the highest), mean rating=5.32 (SD 1.67). The frequency of self-report rating was 12 (1%), 32 (4%), 72 (9%), 123 (20%), 157 (24%), 190 (19%), 148 (6%), 47 (1%), 4 (1%), and 5 (1%) across the 10 status levels (a distribution with acceptable skew –0.27 and kurtosis=0.17).

Ethnicity

We observed that 782 participants (2% missing data) chose an image that best represented their ethnic background from four images representing Black (African), White (Caucasian), East Asian, and West Asian (Indian/Pakistani). Of these, 182 participants (23%) identified themselves as having African ethnicity, 87 participants (11%) as Asian, 454 participants (57%) as Caucasian, 23 participants (3%) as West Asian, 36 participants (5%) as having more than one ethnicity, and 18 participants (2%) reported an ethnicity not shown by our questionnaire stimuli.

Fruit and Vegetable Consumption

The mean level of fruit and vegetable consumption across the entire sample was mean 3.38 (SD 1.76) (skew 0.40 and kurtosis=0.10) portions per day. Young adults reported eating less than older adults (3.13, SD 1.72 vs 3.63, SD 1.76 portions per day; younger adults mean rank=363.9, N=399; older adults mean rank=436.9, N=401; Mann-Whitney U=94,601, P<.001, r=.161). Low socioeconomic status participants reported eating less than those of high socioeconomic status (3.15, SD 1.70 vs 3.62, SD 1.78 portions per day; low status mean rank=363.45, N=395; high status mean rank=425.71, N=393; U=89,884, P<.001, r=.139). Men reported eating less than women (3.11, SD 1.60 vs 3.55, SD 1.83 portions per day; men mean rank=367.25, N=310; women mean rank=419.99, N=488; U=85,638, P<.001, r=.114). Consumption of fruit and vegetables did not differ across three ethnic groups, Black, White, and Asian (N=721, Kruskal-Wallis Test=1.884, df=2, P=.39).

Awareness of Appearance Benefits

We had 487 participants (61%) report they were unaware of the association between diet and skin color; 219 (27%) were somewhat aware, and 95 (12%) were aware. In contrast, 66 (8%) participants were unaware of the health risks (eg, cancer, heart disease) associated with a diet low in fruit and vegetables. Younger participants were more aware of the fruit and vegetable effect on skin color than older participants (younger adults mean rank=417.67, N=400; older adults mean rank=384.37, N=401; U=73,531, P=.02, r=–.081). Participants who reported low socioeconomic status were equally aware of the effect as participants reporting high socioeconomic status (low socioeconomic status mean rank=397.13, N=396; high socioeconomic status mean rank=392.86, N=393; U=76,972, P=.76, r=–.01). Women were more aware of the fruit and vegetable effect on skin color than men (women mean rank=416.67, N=489, men mean rank=373.71, N=310; U=83,946, P=.003, r=.104). Awareness of the fruit and vegetable effect on skin color differed across three ethnic groups, Black, White, and Asian (N=722, Kruskal-Wallis Test=7.915, df=2, P=.02). White participants showed the least awareness (Black mean rank=382.71, Asian mean rank=392.55, White mean rank=347.02).

Motivation Change from Seeing an Effect on Skin Color

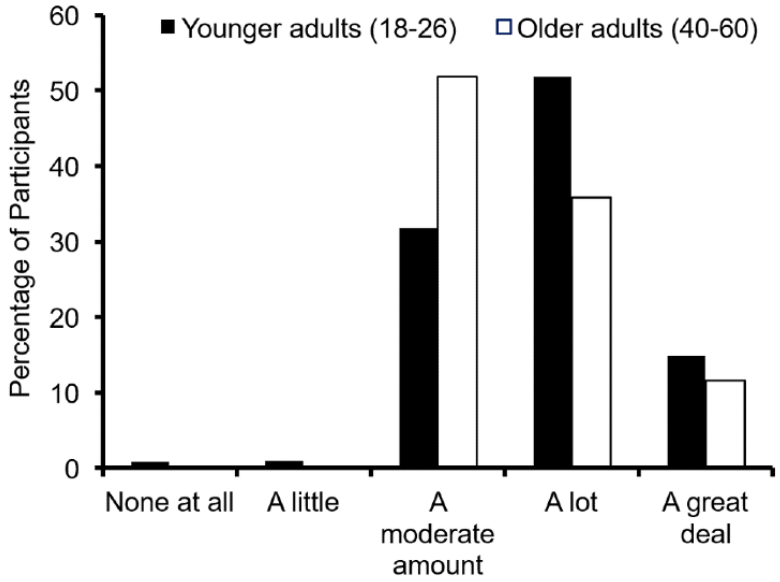
We observed 457 participants (57%) reporting that the simple demonstration of the dietary impact on skin color positively motivated them to eat more fruit and vegetables.

Age

Comparing younger (18-26 years old) and older participants (40-60 years old) showed that the younger group was more likely to endorse the view that the appearance demonstration motivated diet change than the older group (Figure 2, younger participants mean rank=435.79, N=400; older participants mean rank=366.30, N=401; U=66,284, P<.001, r=–.164).

XSL•FO
RenderX

Figure 2. Effect of age on appearance motivation. Change in motivation to eat fruit and vegetables as a result of seeing the demonstration images for two age groups. Younger adults were more positively motivated than older participants.



Gender

Participant gender had no significant effect on the motivation for diet change from the demonstration of the effects of fruit and vegetables on skin color (women mean rank=408.37, N=489, men mean rank=386.80, N=310; U=79,886, *P*=.16, r=.05). The appearance demonstration was thus equally motivating to women and men.

Socioeconomic Status

Comparing low (self-reporting levels 1-5) and high socioeconomic status (self-reporting levels 6-10) participants showed that the two groups were equally likely to endorse the view that the appearance demonstration was motivating to change diet (mean rank low=396.12, N=396; mean rank high=393.88, N=393; U=77,372, *P*=.88, r=–.005). Hence, the appearance motivation was equal across socioeconomic status.

Ethnicity

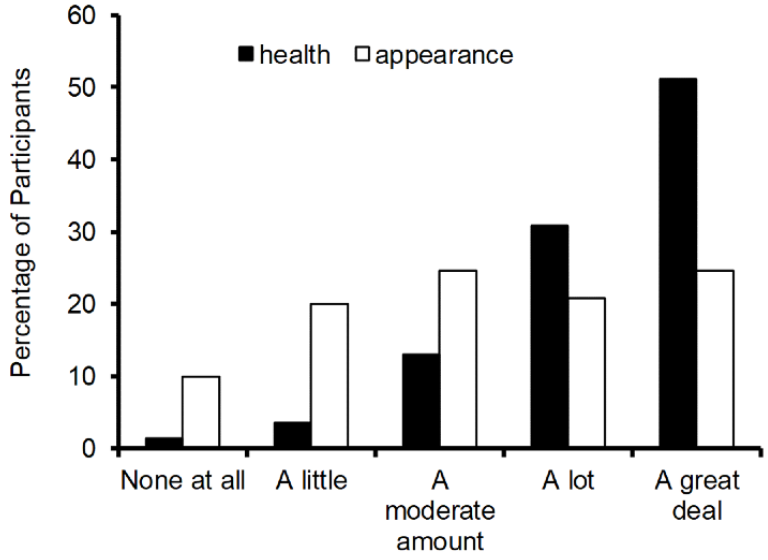
The distribution of answers (to the question of whether the demonstration of skin color change motivated dietary change) across three ethnic groups, Black, White and Asian, was not significantly different (N=721, Kruskal-Wallis Test=2.999, *df*=2, *P*=.22). Comparing the median score of participants answering this question to a median of zero, the score expected by random choice, each ethnic group showed a significant bias to answer the question affirmatively (Black participants: One-Sample Wilcoxon Signed Rank test T=5,899.5, N=182, asymptotic two-tailed significance *P*<.001; Asian participants: T=1,573.5, N=87, *P*<.001; White participants: T=31,311, N=452, *P*<.001). Hence Black, White and Asian groups found the color demonstration motivating (Figure 1).

Appearance vs Health Motivation for Dietary Fruit and Vegetables

The vast majority (n=720, 90%) of participants answered that skin color appearance would motivate them to eat more fruit and vegetables. By contrast, virtually all participants (n=792, 99%) were motivated to a greater or lesser extent by the health benefits of fruit and vegetables in reducing the chance of heart disease and cancer (Figure 3). The response categories were recoded as a score (0-4) to compare the relative importance of health and appearance motivation. The mean numerical score for health motivation was 3.27, while the mean score for appearance was 2.30 (Wilcoxon Signed-Rank test T=118,103; N=801, *P*<.001, r=.618).

XSL•FO
RenderX

Figure 3. Relative motivation of health and appearance for the consumption of fruit and vegetables.



Gender

An appearance/health contrast score was computed for each participant (the 0-4 score of the importance of appearance minus the 0-4 score of the importance of health for consuming fruit and vegetables) to compare health and appearance motivation. Comparing contrast scores for women and men showed that the relative importance of appearance to health in motivating diet change was higher for women (mean rank=416.29, N=489) than men (mean rank=374.3, N=310; U=83,763, *P*=.01, r=.09).

Age

Younger adults (mean rank=442.96, N=400) found the appearance benefits more motivating relative to the health benefits (mean rank=359.15, N=401, U=63,416, *P*<.001, r=.19).

Socioeconomic Status

There was no effect of socioeconomic status on the relative motivation of health and appearance (mean rank high=392.52, N=393; mean rank low=397.46, N=396; U=78,839, *P*=.75, r=–.011).

Knowledge of the Effect of Fruit and Vegetables on Skin Color

The importance of appearance relative to health in motivating diet change was higher for those participants who already knew that diet alters skin color (mean rank=445.79, N=314) than for those not aware of the effect (mean rank=372.12, N=487, U=90,522, *P*<.001, r=.161).

Discussion

Principal Findings

The primary findings of the study were (a) the limited percentage of people (40%) with knowledge of the dietary–skin color association and (b) that many people (57%) found the simple demonstration of the dietary impact on skin color positively motivating to eat more fruit and vegetables, while very few (1%) reported the demonstration was demotivating. The combination of these two findings points to the benefits that can be accrued through publicizing and demonstrating the dietary effect on skin color. Media coverage in the UK has been extensive (reaching millions of TV viewers), but there is still a substantial proportion of the population that reports not knowing the dietary effects. Of the 801 participants, 487 (61%) were unaware of the effects of fruit and vegetable consumption on skin color. Of these, 248 (51%) stated they were motivated to eat more fruit and vegetables as a result of seeing the demonstration. Hence increasing public awareness of the dietary effects on the skin could provide additional motivation for a healthy diet for 31% of the sample population (% of those unaware multiplied by % of those motivated). Extrapolating this to the UK adult population of approximately 50 million, this amounts to 16.5 million people.

Age and Socioeconomic Status

Low socioeconomic status and young adult age are established independent categories predicting low fruit and vegetable consumption [10,11]. In the current study, it was again evident that the younger adults and those reporting low socioeconomic status were reporting lower fruit and vegetable consumption. It is noteworthy that the motivational effects of appearance were most prominent in younger adults. By contrast, we were unable to detect any influence of socioeconomic status on appearance motivation. Hence, it is reasonable to expect that promotion of appearance benefit would not widen health inequalities, and drawing attention to the color benefits of a healthy diet would have an impact on a younger population independent of socioeconomic status, thereby increasing the chance of their life-long adoption of an improved diet.

Gender

We expected women to be more motivated by appearance benefits than men. Comparing reactions to the visual demonstration, we found no significant impact of gender. For both men and women in the present study, motivation for a diet high in fruit and vegetables was more significant for the benefits to health than the motivation for an improved skin color appearance. Nonetheless, motivation from benefits to appearance relative to benefits to health was higher in women than in men.

XSL•FO
RenderX

Fruit and Vegetable Consumption

The average consumption of fruit and vegetables reported by the entire sample here was 3 portions per day (significantly less than the 5 a day recommended by the British National Healthcare System). Hence publicity of the appearance benefits of a proper diet rich in fruit and vegetables could augment motivation for a healthy diet. It is relevant here that the majority of our participants (57%) reported that the appearance benefits were positively motivating. Indeed, appearance has proved valuable in encouraging other aspects of healthy behavior (eg, sun protection) [35].

Ethnicity Effects

The participants in this study reported a variety of ethnic backgrounds (23% Black, 11% Asian, and 57% White). The effect of carotenoids is likely to be less apparent for individuals with darkly pigmented skin. Nonetheless, carotenoid effects on skin color are apparent across different ethnicities. Coetzee & Perrett [36] found that carotenoid supplementation produced an increase in skin yellowness in sun-protected skin regions of African participants. Likewise, a diet rich in fruit and vegetables increases skin yellowness in Asian participants [15,24] as it does in White participants [14,16,26].

While the skin color effects of diet are detectable in a range of ethnic groups, different cultures may vary in their perception of whether the skin color change is desirable. Black South African participants were found to perceive increased skin yellowness positively [23]. However, a recent report indicates that Asian participants from Malaysia find only a slight increase in carotenoid skin color more attractive than the skin color baseline [24]. A further report concludes that mainland Chinese participants do not find increased yellow skin pigmentation attractive [37]. Nonetheless, whatever the influence of culture on color preferences, the current study found that Asian, Black, and White participants from the UK were equally impressed by the impact of diet on skin color.

Limitations

We detected several factors associated with attitudes to the appearance benefit arising from a diet high in fruit and vegetables. Socioeconomic status did not influence appearance motivation. Our measures of socioeconomic status and dietary intake were self-reported, and portions of fruit and vegetables were not defined (eg, as 80 g). More objective measures and greater sampling of those with low and high socioeconomic status individuals are needed to be confident in null effects. We have not assessed residential area deprivation [11] or sense of personal relative deprivation [38], both predictors of diet and self-reported health.

We illustrated the dietary effect with a single Caucasian male face. The demonstration may be more effective if displayed with the same gender, age, and ethnicity as the participant or better still with an image of the participant [30]. The demonstration was made early in the survey, which may have led to participants increasing their reported fruit and vegetable consumption. The dietary effect on skin color is not unique since other aspects of lifestyle (eg, fitness and reduced body fat) have similar impacts on skin color appearance [39].

We measured the effects of appearance on self-reported motivation to consume more fruit and vegetables. We acknowledge that motivation is necessary but not sufficient for actual dietary change [40]. The impact of appearance on real dietary change is likely to be smaller than the impact on motivation to change. Translating motivation into an actual increase in fruit and vegetable consumption remains an issue that is affected by a person's perception that they can control their behavior [14,40,41]. Nonetheless, any change in motivation to consume more fruit and vegetables is an essential step in the right direction of a healthy lifestyle.

Acknowledgments

This work was supported by grants from a British Academy Wolfson Research Professorship (grant number: WRP/2008/87) and University of St Andrews Knowledge Exchange and Impact Fund 2017/18. The authors gratefully acknowledge Lesley Ferrier, Xue Lei, and Martha Lucia Borras Guevara for assistance with study preparation, and Anne Perrett for proofreading.

Conflicts of Interest

None to declare.

References

1. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017 Jun 01;46(3):1029-1056 [FREE Full text] [doi: 10.1093/ije/dyw319] [Medline: 28338764]
2. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014 Jul 29;349:g4490 [FREE Full text] [doi: 10.1136/bmj.g4490] [Medline: 25073782]
3. Hu D, Huang J, Wang Y, Zhang D, Qu Y. Fruits and Vegetables Consumption and Risk of Stroke. Stroke 2014 Jun;45(6):1613-1619 [FREE Full text] [doi: 10.1161/strokeaha.114.004836]

XSL•FO
RenderX

4. Li M, Fan Y, Zhang X, Hou W, Tang Z. Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies. BMJ Open 2014 Nov 05;4(11):e005497 [FREE Full text] [doi: 10.1136/bmjopen-2014-005497] [Medline: 25377009]
5. Du H, Li L, Bennett D, Guo Y, Turnbull I, Yang L, China Kadoorie Biobank study. Fresh fruit consumption in relation to incident diabetes and diabetic vascular complications: A 7-y prospective study of 0.5 million Chinese adults. PLoS Med 2017 Apr;14(4):e1002279 [FREE Full text] [doi: 10.1371/journal.pmed.1002279] [Medline: 28399126]
6. Yip CSC, Chan W, Fielding R. The Associations of Fruit and Vegetable Intakes with Burden of Diseases: A Systematic Review of Meta-Analyses. J Acad Nutr Diet 2019 Mar 01;119(3):464-481 [FREE Full text] [doi: 10.1016/j.jand.2018.11.007] [Medline: 30639206]
7. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
8. The EIU. Financing the future: Choices and challenges in global health. 2015. URL: http://globalhealth.eiu.com/briefing-paper/ [accessed 2019-04-09] [WebCite Cache ID 77VMfqe1N]
9. Health Survey for England Summary of key findings. 2017. URL: http://healthsurvey.hscic.gov.uk [accessed 2019-03-22] [WebCite Cache ID 774SfD8PI]
10. The National Diet and Nutrition Survey, Public Health England. 2019. URL: https://www.gov.uk/government/collections/national-diet-and-nutrition-survey#current-ndns-results [accessed 2019-03-22]
11. Shohaimi S, Welch A, Bingham S, Luben R, Day N, Wareham N, et al. Residential area deprivation predicts fruit and vegetable consumption independently of individual educational level and occupational social class: a cross sectional population study in the Norfolk cohort of the European Prospective Investigation into Cancer (EPIC-Norfolk). J Epidemiol Community Health 2004 Aug;58(8):686-691 [FREE Full text] [doi: 10.1136/jech.2003.008490] [Medline: 15252072]
12. Lakshman R, McConville A, How S, Flowers J, Wareham N, Cosford P. Association between area-level socioeconomic deprivation and a cluster of behavioural risk factors: cross-sectional, population-based study. J Public Health (Oxf) 2011 Jun;33(2):234-245 [FREE Full text] [doi: 10.1093/pubmed/fdq072] [Medline: 20884643]
13. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
14. Whitehead RD, Re D, Xiao D, Ozakinci G, Perrett DI. You are what you eat: within-subject increases in fruit and vegetable consumption confer beneficial skin-color changes. PLoS One 2012;7(3):e32988 [FREE Full text] [doi: 10.1371/journal.pone.0032988] [Medline: 22412966]
15. Tan KW, Graf BA, Mitra SR, Stephen ID. Daily Consumption of a Fruit and Vegetable Smoothie Alters Facial Skin Color. PLoS One 2015;10(7):e0133445 [FREE Full text] [doi: 10.1371/journal.pone.0133445] [Medline: 26186449]
16. Pezdirc K, Hutchesson MJ, Whitehead R, Ozakinci G, Perrett D, Collins CE. Fruit, Vegetable and Dietary Carotenoid Intakes Explain Variation in Skin-Color in Young Caucasian Women: A Cross-Sectional Study. Nutrients 2015 Jul 15;7(7):5800-5815 [FREE Full text] [doi: 10.3390/nu7075251] [Medline: 26184306]
17. Pezdirc K, Hutchesson MJ, Williams RL, Rollo ME, Burrows TL, Wood LG, et al. Consuming High-Carotenoid Fruit and Vegetables Influences Skin Yellowness and Plasma Carotenoids in Young Women: A Single-Blind Randomized Crossover Trial. J Acad Nutr Diet 2016 Aug;116(8):1257-1265. [doi: 10.1016/j.jand.2016.03.012] [Medline: 27160341]
18. Bixley G, Clark K, James A. Skin colour predicts fruit and vegetable intake in young Caucasian men: A cross-sectional study. Journal of Nutrition & Intermediary Metabolism 2018 Jun;12:20-27. [doi: 10.1016/j.jnim.2018.06.001]
19. Mayne ST, Cartmel B, Scarmo S, Lin H, Leffell DJ, Welch E, et al. Noninvasive assessment of dermal carotenoids as a biomarker of fruit and vegetable intake. Am J Clin Nutr 2010 Oct;92(4):794-800 [FREE Full text] [doi: 10.3945/ajcn.2010.29707] [Medline: 20685953]
20. Scarmo S, Henebery K, Peracchio H, Cartmel B, Lin H, Ermakov IV, et al. Skin carotenoid status measured by resonance Raman spectroscopy as a biomarker of fruit and vegetable intake in preschool children. Eur J Clin Nutr 2012 May;66(5):555-560 [FREE Full text] [doi: 10.1038/ejcn.2012.31] [Medline: 22434053]
21. Darvin M, Patzelt A, Knorr F, Blume-Peytavi U, Sterry W, Lademann J. One-year study on the variation of carotenoid antioxidant substances in living human skin: influence of dietary supplementation and stress factors. J Biomed Opt 2008;13(4):044028 [FREE Full text] [doi: 10.1117/1.2952076] [Medline: 19021355]
22. Ermakov IV, Ermakova M, Sharifzadeh M, Gorusupudi A, Farnsworth K, Bernstein PS, et al. Optical assessment of skin carotenoid status as a biomarker of vegetable and fruit intake. Arch Biochem Biophys 2018 May 15;646:46-54 [FREE Full text] [doi: 10.1016/j.abb.2018.03.033] [Medline: 29601824]
23. Stephen ID, Coetzee V, Perrett DI. Carotenoid and melanin pigment coloration affect perceived human health. Evolution and Human Behavior 2011 May;32(3):216-227. [doi: 10.1016/j.evolhumbehav.2010.09.003]
24. Tan KW, Graf BA, Mitra SR, Stephen ID. Impact of fresh fruit smoothie consumption on apparent health of Asian faces. Evolution and Human Behavior 2017 Jul;38(4):522-529 [FREE Full text] [doi: 10.1016/j.evolhumbehav.2017.02.004] [Medline: 26186449]
25. Pezdirc K, Rollo ME, Whitehead R, Hutchesson MJ, Ozakinci G, Perrett D, et al. Perceptions of carotenoid and melanin colouration in faces among young Australian adults. Aust J Psychol 2017 May 19;70(1):85-90. [doi: 10.1111/ajpy.12163]

XSL•FO
RenderX

26. Foo Y, Rhodes G, Simmons L. The carotenoid beta-carotene enhances facial color, attractiveness and perceived health, but not actual health, in humans. Behav Ecol 2017;28(2):570-578. [doi: 10.1093/beheco/arw188]
27. Lefevre CE, Perrett DI. Fruit over sunbed: carotenoid skin colouration is found more attractive than melanin colouration. Q J Exp Psychol (Hove) 2015;68(2):284-293. [doi: 10.1080/17470218.2014.944194] [Medline: 25014019]
28. Whitehead RD, Ozakinci G, Perrett DI. Attractive skin coloration: harnessing sexual selection to improve diet and health. Evol Psychol 2012 Dec 20;10(5):842-854. [Medline: 23253790]
29. Appleton KM, McGrath AJ, McKinley MC, Draffin CR, Hamill LL, Young IS, et al. The value of facial attractiveness for encouraging fruit and vegetable consumption: analyses from a randomized controlled trial. BMC Public Health 2018 Mar 01;18(1):298 [FREE Full text] [doi: 10.1186/s12889-018-5202-6] [Medline: 29490640]
30. Whitehead RD, Ozakinci G, Perrett DI. A randomized controlled trial of an appearance-based dietary intervention. Health Psychol 2014 Jan;33(1):99-102. [doi: 10.1037/a0032322] [Medline: 23527517]
31. Hayes D, Ross CE. Concern with appearance, health beliefs, and eating habits. J Health Soc Behav 1987 Jun;28(2):120-130. [Medline: 3611701]
32. Chung S, Hoerr S, Levine R, Coleman G. Processes underlying young women's decisions to eat fruits and vegetables. J Hum Nutr Diet 2006 Aug;19(4):287-298. [doi: 10.1111/j.1365-277X.2006.00704.x] [Medline: 16911241]
33. Prolific. 2020. URL: https://www.prolific.co [accessed 2019-03-22]
34. Lefevre CE, Ewbank MP, Calder AJ, von dem Hagen E, Perrett DI. It is all in the face: carotenoid skin coloration loses attractiveness outside the face. Biol Lett 2013;9(6):20130633 [FREE Full text] [doi: 10.1098/rsbl.2013.0633] [Medline: 24307526]
35. Williams AL, Grogan S, Clark-Carter D, Buckley E. Appearance-based interventions to reduce ultraviolet exposure and/or increase sun protection intentions and behaviours: a systematic review and meta-analyses. Br J Health Psychol 2013 Feb;18(1):182-217. [doi: 10.1111/j.2044-8287.2012.02089.x] [Medline: 22989352]
36. Coetzee V, Perrett DI. Effect of beta-carotene supplementation on African skin. J Biomed Opt 2014 Feb;19(2):025004 [FREE Full text] [doi: 10.1117/1.JBO.19.2.025004] [Medline: 24525826]
37. Han C, Wang H, Hahn AC, Fisher CI, Kandrik M, Fasolt V, et al. Cultural differences in preferences for facial coloration. Evolution and Human Behavior 2018 Mar;39(2):154-159. [doi: 10.1016/j.evolhumbehav.2017.11.005]
38. Callan MJ, Kim H, Matthews WJ. Predicting self-rated mental and physical health: the contributions of subjective socioeconomic status and personal relative deprivation. Front Psychol 2015;6:1415 [FREE Full text] [doi: 10.3389/fpsyg.2015.01415] [Medline: 26441786]
39. Perrett DI, Talamas SN, Cairns P, Henderson AJ. Skin Color Cues to Human Health: Carotenoids, Aerobic Fitness, and Body Fat. Front Psychol 2020;11:392 [FREE Full text] [doi: 10.3389/fpsyg.2020.00392] [Medline: 32218759]
40. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
41. Menozzi D, Sogari G, Mora C. Explaining Vegetable Consumption among Young Adults: An Application of the Theory of Planned Behaviour. Nutrients 2015 Sep 10;7(9):7633-7650 [FREE Full text] [doi: 10.3390/nu7095357] [Medline: 26378570]

Edited by G Eysenbach, K Fortuna; submitted 02.08.19; peer-reviewed by J Woodside, C Collins, K Appleton; comments to author 02.03.20; revised version received 27.04.20; accepted 14.05.20; published 14.07.20.

Please cite as:
Cairns P, Ozakinci G, Perrett DI
Reactions to an Online Demonstration of the Effect of Increased Fruit and Vegetable Consumption on Appearance: Survey Study
J Med Internet Res 2020;22(7):e15726
URL: https://www.jmir.org/2020/7/e15726
doi:10.2196/15726
PMID:32459626

©Patrick Cairns, Gozde Ozakinci, David Ian Perrett. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Current State of Research, Challenges, and Future Research Directions of Blockchain Technology in Patient Care: Systematic Review

Polina Durneva[1*], BA; Karlene Cousins[1*], BSc, MSci, PhD, JD; Min Chen[1*], BA, MA, PhD

Department of Information Systems & Business Analytics, Florida International University, Miami, FL, United States

[*]all authors contributed equally

Corresponding Author:
Karlene Cousins, BSc, MSci, PhD, JD
Department of Information Systems & Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 305 348 2160
Email: kcousins@fiu.edu

Abstract

Background: Blockchain offers a promising new distributed technology to address the challenges of data standardization, system interoperability, security, privacy, and accessibility of medical records.

Objective: The purpose of this review is to assess the research on the use of blockchain technology for patient care and the associated challenges and to provide a research agenda for future research.

Methods: This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines. We queried the Cumulative Index of Nursing and Allied Health Literature (CINAHL), PubMed, Excerpta Medica dataBASE (EMBASE), and Web of Science databases for peer-reviewed research articles published up to December 2019 that examined the implementation of blockchain technology in health care settings. We identified 800 articles from which we selected 70 empirical research articles for a detailed review.

Results: Blockchain-based patient care applications include medical information systems, personal health records, mobile health and telemedicine, data preservation systems and social networks, health information exchanges and remote monitoring systems, and medical research systems. These blockchain-based health care applications may improve patient engagement and empowerment, improve health care provider access to information, and enhance the use of health care information for medical research.

Conclusions: Blockchain health information technology (HIT) provides benefits such as ensuring data privacy and security of health data, facilitating interoperability of heterogeneous HIT systems, and improving the quality of health care outcomes. However, barriers to using blockchain technology to build HIT include security and privacy vulnerabilities, user resistance, high computing power requirements and implementation costs, inefficient consensus algorithms, and challenges of integrating blockchain with existing HIT. With 51% of the research focused on medical information systems such as electronic health record and electronic medical record, and 53% of the research focused on data security and privacy issues, this review shows that HIT research is primarily focused on the use of blockchain technologies to address the current challenges HIT faces. Although Blockchain presents significant potential for disrupting health care, most ideas are in their infancy.

(J Med Internet Res 2020;22(7):e18619) doi:10.2196/18619

KEYWORDS

blockchain; health information technology; systematic review; security; privacy; interoperability; health outcomes

XSL•FO
RenderX

Introduction

Background

There is a growing need to integrate health care information across a range of uses and stakeholders and to secure such data from unauthorized breaches, while making it easier for patients to access patient data. Blockchain is a distributed technology that has the potential to address data standardization challenges, system interoperability, and accessibility of medical records to support a more secure, patient-centric approach to health care information systems.

Blockchain is a secure and immutable transaction ledger [1,2], which is distributed in a decentralized manner across all computing devices that are part of the blockchain infrastructure [3]. Blockchain's decentralized design facilitates peer-to-peer–based network transactions between users without the need for a trusted third party. Although more commonly used in decentralized financial applications such as cryptocurrencies and initial coin offerings (ICOs), blockchain's advanced features (eg, consensus mechanisms, digital signatures, and hash chains) promise to address the unique challenges health information systems commonly face, such as poor security, privacy, efficiency, and interoperability.

Blockchains can be classified into 2 types: permissionless and permissioned blockchains [4]. The permissionless blockchain is open to the public, allows anyone to join the blockchain without approval, and usually provides an economic incentive for participating in the blockchain. Most known cryptocurrency blockchains are public, such as Bitcoin and Litecoin. A permissioned blockchain incorporates access control mechanisms to restrict user access. Permissioned blockchains are further classified as private or consortium blockchains based on their governance structure. A private blockchain is managed by a single organization and is usually used in enterprise solutions. The consortium blockchain is semiprivate, has a controlled user group, and works across different organizations. Compared with the public blockchain, permissioned blockchains are restrictive, and a central authority grants access to the blockchain. Therefore, permissioned blockchains lose some of the advantages of decentralization, but are more effective in securely sharing and managing real-time data among participating health care stakeholders. Embleema and the Synaptic Health Alliance are examples of health care blockchain consortia that allow selected health care stakeholders, including patients, advocacy groups, life sciences companies, payers, authorities, and care centers, to form nodes on the networks and manage health data securely, without a central authority.

Researchers are recognizing blockchain's potential as a disruptive technology in health care and have begun to conduct research on how to leverage blockchain. We seek to understand what progress has been made in blockchain health care research and which problems researchers are addressing and encountering in use case implementations. We answer these questions by conducting a literature review and synthesizing the empirical blockchain health care research. The format of this systematic review adhered to the literature review standards of the review methodology, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Multimedia Appendix 1) [5].

This analysis shows that health care researchers are focusing on interoperability and platform issues and are exploring opportunities to use blockchain to enhance privacy and security and improve data integrity and transparency. We also identify the different types of blockchain health information technology (HIT) and discuss their implications for patient engagement and empowerment, provider access to personal health information (PHI), and medical or clinical research. We discuss the barriers and challenges of using and implementing blockchain HIT and propose new research directions.

The rest of the paper is organized as follows. First, in the *Methods* section, we describe the search process and selection criteria. In the *Results* section, we summarize our findings by highlighting the HIT challenges that blockchain addresses, the different types of blockchain-based HIT, how blockchain HIT research has evolved over time, and barriers and challenges to implementation. In the *Discussion* section, we suggest potential areas of research. In the *Conclusion* section, we end our review by summarizing and highlighting the implications of our key findings.

Objective

The purpose of this review is to evaluate the current literature on the application of blockchain technology in health care with a focus on patient care, to assess the current state of research, the associated challenges, and potential areas for future research. In accordance with our objectives, the research questions are as follows:

1. What current HIT challenges does blockchain address?
2. What are the predominant applications for patient care in blockchain HIT research?
3. How has blockchain HIT research evolved over time?
4. What health care activities are impacted by current blockchain HIT research?
5. What are the challenges associated with blockchain-based HIT implementations?

Methods

Study Identification and Selection

Eligible papers were published in academic peer-reviewed journals and conference proceedings in English.

We searched bibliographic databases such as Web of Science, PubMed, Cumulative Index of Nursing and Allied Health Literature (CINAHL), and Excerpta Medica dataBASE (EMBASE). As PubMed, CINAHL, and EMBASE focus mostly on medical content, we used the search term "blockchain" to query these databases to execute a broad search so as not to miss important results. As the Web of Science provides content related to a broader range of topics, we used the search terms "blockchain" AND "health" OR "blockchain" and "medical" to query this database. We searched for all papers published up to December 12, 2019. The searches identified 800 potential articles: Web of Science, 298; PubMed, 184; CINAHL, 129; and EMBASE, 189. To validate our results, we conducted an

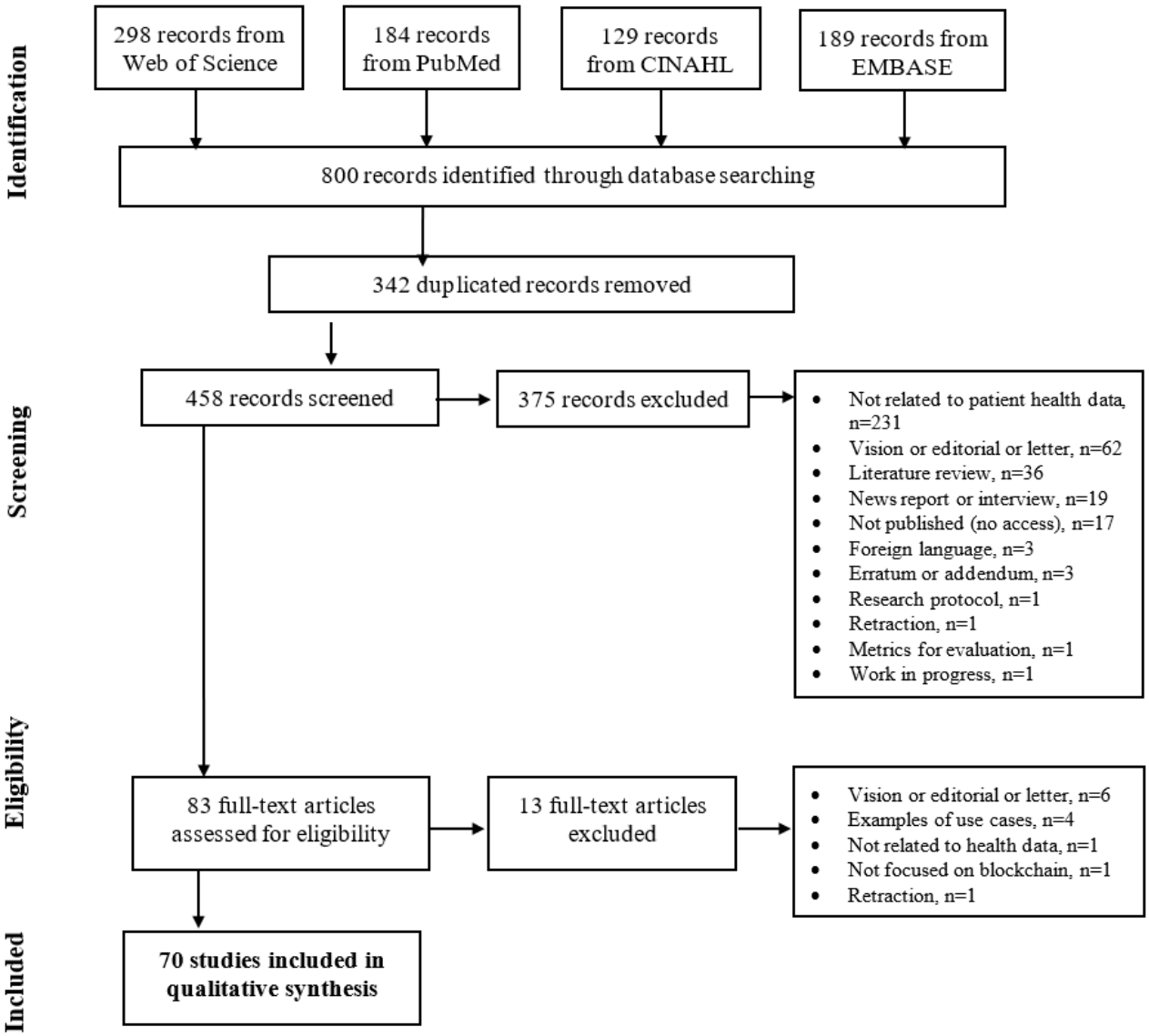
XSL•FO
RenderX

alternative search using the search terms “blockchain” AND (health OR medic OR biomedic OR clinic). The alternative search terms came from prior research [6]. This alternative search identified only 652 potential articles and no new results, compared with the 800 potential articles initially identified: Web of Science, 264; PubMed, 146; CINAHL, 9; and EMBASE, 151.

Data Extraction and Analysis of Types of Blockchain Health Care Applications and Benefits

Two reviewers used Rayyan, a web app for systematic reviews, to independently review all titles and abstracts. A total of 342 duplicates were identified and removed, resulting in 458 papers being selected for full review. The reviewers disagreed on 29 papers, resulting in an inter-rater reliability of 93.7% (429/458). All disagreements were resolved during a consensus meeting. After this assessment, 83 out of the 458 studies remained for analysis. Most of the excluded papers (n=231) were not related to patients’ health data. For example, some studies focused on the application of blockchain in supply chains or cryptocurrency. We labeled 62 excluded papers pertaining to speculations about future blockchain technology projects for the health industry as visionary/editorial/commentary papers. Focusing primarily on peer-reviewed publications, we also excluded news reports, literature reviews, working papers, and research protocols. After screening for full-text eligibility, we identified 70 studies for the final review. Figure 1 presents our study identification and selection process.

Figure 1. Study identification and selection process. CINAHL: Cumulative Index of Nursing and Allied Health Literature; EMBASE: Excerpta Medica dataBASE.



We evaluated each paper to determine the following: (1) the types of blockchain-based HIT applications that researchers focused on including electronic health records (EHRs), personal health records (PHR), health information exchange (HIE), and telemedicine; (2) benefits of using blockchain for HIT; (3) health care activities that would benefit from their use, including patient engagement and empowerment, medical/clinical research, and provider access and use; and (4) the barriers and challenges associated with the implementation and maintenance of blockchain-based HIT.

The next stage of analysis was to further characterize the studies based on 4 categories that represent the issues that blockchain

XSL•FO
RenderX

technology addressed in each paper: security, privacy, interoperability, and health care outcomes. These categories were determined a priori based on prior research that identified these as major challenges experienced during HIT implementation and use [7-9]. As the analysis proceeded, additional subcategories emerged. We also merged data access, security, and privacy into a single category. At each stage of the review, the reviewers resolved all disagreements by discussion and reached a consensus. Further details can be found in Multimedia Appendix 2.

In the next section, we report on the results of this analysis.

Results

How Blockchain Addresses HIT's Challenges

In this section, we examine the current HIT challenges that blockchain addresses.

Health care data have special properties. The content must be true and complete, and the data must be traceable and resistant to alteration, forgery, or deletion. Due to legal requirements to preserve patient privacy and secure protected health information, health care data must also be safe and anonymous when stored. Health care data contain sensitive personal information, and thus, prevention mechanisms must be established to prevent unauthorized staff from obtaining and extracting information. To facilitate medical or clinical research, health care data must be made accessible while hiding the identity of the patients to whom it belongs. In addition, the data should be encrypted so that once the data are stolen, they cannot be understood without decryption.

HIT faces challenges with respect to ensuring data security, privacy, and integrity. Sharing health care data is challenging because patient data are often stored in disparate systems and achieving interoperability presents challenges such as connecting heterogeneous systems securely, restricting access by unauthorized parties, and maintaining data integrity. Poor health care data quality adversely affects the quality of health care outcomes.

Prior research shows that blockchain can help to overcome the aforementioned challenges. We categorize this research into 3 groups to represent the main HIT challenges that blockchain can address: data security and privacy, interoperability, and health care quality outcomes (Table 1).

Most of the research reviewed focused on blockchain's use to strengthen HIT security or patients' privacy during health data exchange or access; 53% (37/70) papers focused on addressing patients' lack of control over the privacy and security of their data [10-46], and 40% (28/70) papers addressed blockchain's ability to prevent data tampering [10,12,21,25,33,34,37,38,40-44,46-60]. Data breaches were addressed in 37% (26/70) papers [15,18,19,22,24,29,31,32,34,36-38,43,44,54,56-66], 9% (6/70) papers mentioned malicious attacks (eg, impersonation) that blockchain could potentially resolve [17,34,44,56,60,67], and 4% (3/70) papers focused on how blockchain can preserve patients' anonymity while third parties accessed their health and medical records for activities such as medical research [41,42,68].

The research studies also investigated how blockchain addresses HIT interoperability issues. Interoperability is the ability of different information systems, devices, and applications (*systems*) to access, exchange, integrate, and cooperatively use data in a coordinated manner, within and across organizational, regional, and national boundaries, to provide timely and seamless portability of information [69]. Given that HIT infrastructure might vary by hospital, department, and other structural divisions, HIT incompatibility may arise when data transfer is attempted across these systems, as in the case of HIE; 9% (6/70) papers evaluated blockchain's use for resolving the poor incompatibility of existing HIT [21,22,26,27,30,55], and 4% (3/70) papers focused on how blockchain can address the poor integration of large volumes of data from different sources [26,37,61]. One challenge to interoperability is the waiting time for data to be updated, affecting the temporal facet of interoperability and real-time access. The lack of real-time access in interoperable systems was discussed in 6% (4/70) papers [13,40,70,71]. The inconsistency of data structures across heterogenous systems makes it difficult to ensure data integrity when data transfer is attempted across interoperable systems; 23% (16/70) papers indicated that blockchain could enhance data integrity across interoperable systems [16-18,33,35-37,40,41,44,49,64,66,72-74]. Furthermore, data transparency (which refers to the accessibility to data despite its location, data credibility, and data accuracy) across interoperable systems was discussed in 14% (10/70) papers [14,33,40,47,52,59,70,72,75,76].

The third category covers HIT's alignment with the health organizations' goals to achieve specific health care outcomes such as health care quality and system satisfaction; 10% (7/70) papers focused on addressing the inefficiency of current health systems, such as lengthy processing times and the inability to simultaneously process large data volumes [16,24,28,37,55,67,77]. With regard to the quality of patient care, quite often, missing or incorrect health data lead to repetitive lab tests or diagnostic errors, which can be detrimental to patients' health; 9% (6/70) papers discussed how blockchain can reduce misdiagnosis and overtreatment [27,32,35,40,78,79]. Data dredging may occur in clinical trials where researchers may alter or omit data to achieve a statistically significant result in their experiments; 4% (3/70) papers addressed data dredging [32,52,70]. Another issue is the lack of trust between stakeholders (patients, providers, and researchers) because of stakeholders' capability to modify and edit health data, which may lead to errors. A total of 4% (3/70) papers examined how blockchain could facilitate the immutability and traceability of health records updates [32,61,80]. The high cost of maintenance of the current health systems was mentioned in 3% (2/70) papers [37,73]. Blockchain technology could potentially reduce maintenance costs by increasing the efficiency and speed of health care data management.

XSL•FO
RenderX

Table 1. Health information technology challenges that blockchain addresses.

Challenge	Frequency (N=70), n
Data access, security, and privacy	
Patient lack of control over data	37
Data tampering	28
Data breaches	26
Other malicious attacks (eg, impersonation)	6
Nonanonymous access to records	3
Interoperability	
Lack of data integrity	16
Lack of data transparency	10
Poor compatibility with existing electronic health structures	6
Lack of real-time access to results	4
Poor integration of large volumes of data	3
Health care outcomes	
Low efficiency	7
Misdiagnosis/overtreatment	6
Data dredging	3
Lack of trust between stakeholders	4
High cost of maintenance and data management	2

In the next section, we discuss how blockchain addresses the current challenges surrounding data security and privacy, interoperability, and health care outcomes across a broad spectrum of health care applications and activities.

Blockchain HIT Applications and Health Care Activities Impacted

Table 2 illustrates the different types of blockchain HITs that researchers focused on. Blockchain HIT research focuses predominantly on improving medical information management systems (MIMS) such as EHRs, electronic medical records (EMRs), and PHRs.

Table 2. Blockchain health information technology applications.

Blockchain health information technology	Frequency (N=70), n
Medical information management system	36
Personal health records	7
Clinical trial or health research platform	6
Health information exchange	6
Remote patient monitoring	5
Mobile health	3
Medical image sharing platform	2
Telemedicine platform	2
Predictive or classification modelling	2
Pervasive social network	1
Data preservation systems	1

We analyzed the blockchain HIT research by the year of first mention (Figure 2). In 2016-2017, researchers worked on providing comprehensive and secure real-time access to patient data in HIT applications such as EHR, EMR, PHR, mobile health (mHealth), remote patient monitoring (RPM), pervasive social networks (PSNs), and clinical trial systems. Research during this period focused on increasing data security and privacy and leveraging blockchain's unique properties of decentralization, immutability, anonymization, and transaction synchronization to provide a single view of patient data. In

XSL•FO
RenderX

2018, blockchain HIT research expanded to include improving health information management, exchange, and synchronization across decentralized nodes in telemedicine, medical image sharing, data preservation systems (DPSs), and HIE. In 2019, blockchain HIT research expanded to leveraging blockchain's decentralization properties to eliminate a single point of control while carrying out predictive analytics and classification modeling across multiple institutions.

Blockchain HIT improves patient engagement and empowerment, research quality and processes, and provider's information access and use (Figure 3). We describe how blockchain can improve each of these areas in further detail in the next section.

Figure 2. Blockchain health information technology research evolution over time by date of first mention.

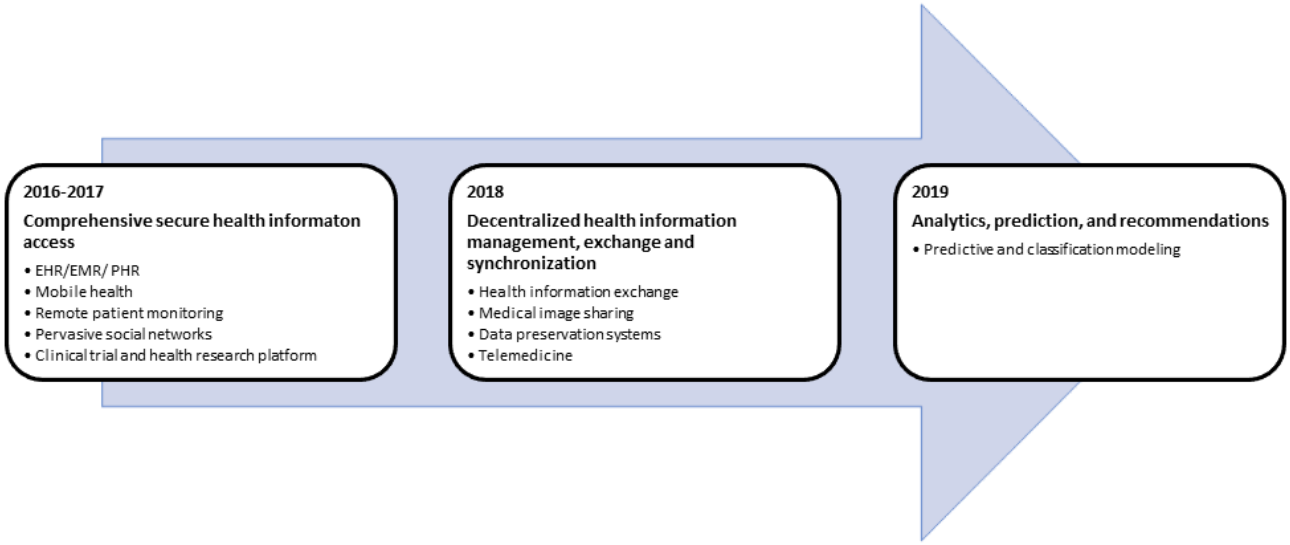
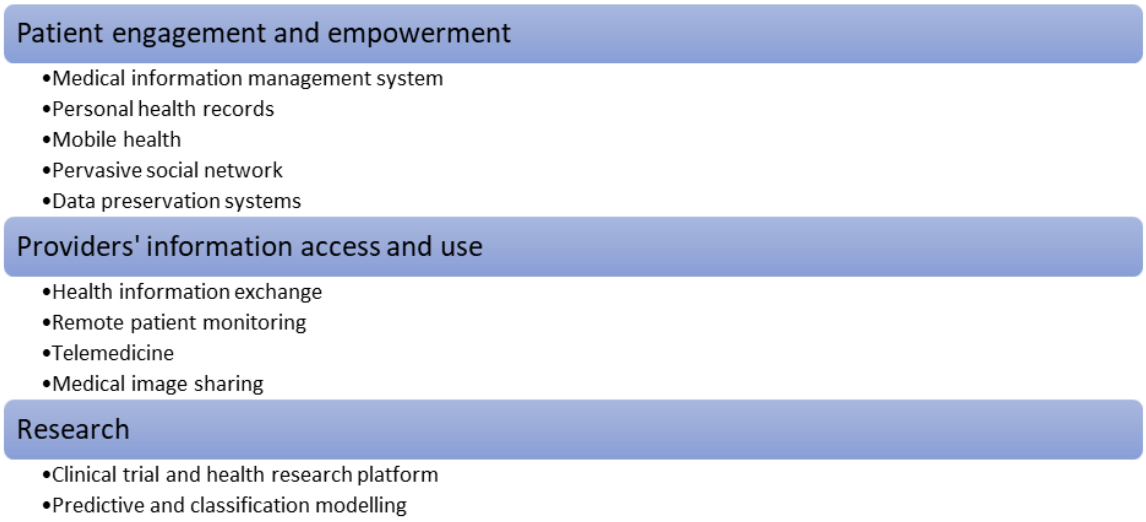


Figure 3. Health care activities that blockchain health information technology impacts.



Patient Engagement and Empowerment

We define patient engagement as the act of patient and provider working together to improve the patient's health [81]. Patient engagement may result in patient empowerment. Patient empowerment refers to a process through which patients gain greater control over decisions and actions that affect their health [82]. Patient engagement and empowerment promote shared decision making about the patient's care by both the providers and the patient and patient-centered care that is highly responsive to individual patients' preferences, needs, and values [83].

Patient engagement and empowerment may benefit from 6 types of blockchain HIT: MIMS, PHRs, mHealth, telemedicine, PSNs, and DPSs.

Medical Information Management Systems

MIMS refer to EHRs and EMRs, which store and manage patients' health information. The EHR is a computerized and standardized information model with information relevant to the health and wellness of an individual, enabling integration among multiple health care providers [30]. The EMR differs from the EHR as it focuses on the internal medical domain of health organizations and is not integrated between health care providers. Blockchain's use for EHRs, EMRs, and other record systems in which records are originated and controlled by

XSL•FO
RenderX

providers was discussed in 51% (36/70) papers [11,14-21,27,29,30,33,35-37,40,42-44,46,49,53,55,59,60,63,67,68,72-74,76-78,84]. Using blockchain to integrate various health record systems (EHR, EMR, PHR, etc), mHealth, and telemedicine platforms can contribute to patient empowerment and patient engagement by providing a synchronized view of patient information to patients and health care providers, to support patient decision making and control.

It is challenging to keep track of all data and ensure its immutability in the context of complex medical cases that require multiple diverse activities, patients' visits, and providers' treatments [17]. Blockchain-based MIMS facilitate the creation of immutable records with traceable transactions that cannot be changed over time. Blockchain can provide patients and providers with more secure and private access to the data and allow them to better collaborate in complex medical situations.

Blockchain technology facilitates smart contracts, which is a computer protocol intended to digitally facilitate, verify, or enforce the negotiation or performance of a contract between two parties without a third party's involvement. Smart contracts can facilitate formalized contracts to streamline patient consent for treatment, patient authorization for access to and release of PHI, and the authentication and verification of those seeking access to patient data. Patient privacy is also enhanced as blockchain can allow patients to authorize access to their records while remaining anonymous, for example, for research purposes. Patients can also publish medical information to the blockchain anonymously, which is also useful for sharing information with third parties and for research purposes. Blockchain's properties, including its cryptographic techniques and immutability and autonomy, enhance user authentication and verification, access control, and HIT security and privacy [18].

In the health care sector, the resilience of health records systems in medical information systems is crucial, given the increasing number of data breaches in health care [85]. Blockchain's distributed design is more resistant to malicious attacks, as there is no single point of control or failure. Overall, such increased security and privacy may motivate patients to engage in full disclosure of their condition and contribute to patient empowerment by facilitating better provider feedback, counseling, and patient decision making.

PHRs

PHR refer to the representation of information regarding an individual's health, including wellness, development, and welfare, which the patient owns and can share with third parties [30]. PHR is oriented to the patient but can be integrated with the EHR. Blockchain's use for PHR was discussed in 10% (7/70) papers [13,23,25,26,30,64,78]. PHR's main challenges include poor interoperability with other systems, large data volumes, outdated data, duplication, lack of standardization, and fragmentation. Due to its distributed design, a permissioned blockchain is well positioned to address these challenges by providing a distributed ledger where transactions can be recorded by multiple, authenticated parties. Due to its decentralized design, blockchain-based PHR can provide patients with a unified view of their scattered health records by interconnecting scattered patient data across several health care organizations. Furthermore, as the PHR is updated with the patient's medical conditions and services received, health providers can become aware of changes in their patients' conditions and alter and modify treatment plans, even though updates may occur elsewhere. These features can improve the communication effectiveness between a provider and a patient, providing the patient with the information needed to take charge of their health care decision making. In terms of patient engagement and empowerment, blockchain-based PHR can provide patients with control over their health data and facilitate providers' real-time updates of patient data.

mHealth and Telemedicine

mHealth, telemedicine platforms, DPSs, and PSNs can also contribute to patient engagement and empowerment. A total of 4% (3/70) papers focusing on mHealth indicated that blockchain's distributed design can prevent data tampering and provide a single point of control [22,31,51]. Data tampering may occur during a malicious attack and lead to a loss of data reliability within the mHealth application [51]. Tamperproof systems are critical in scenarios where treatment is automatically administered to patients without human intervention, based on the data the mHealth application collected. Tamperproof systems are also essential to ensure that medical decision making is based on accurate information to prevent harm to patients. Blockchain enables tamper-resistant systems by maintaining a constantly growing list of transactional records divided into blocks and uses consensus algorithms to allow multiple parties to agree on a common state. Due to its distributed design, blockchain is well suited to mHealth applications as it allows for frequent updates.

A total of 3% (2/70) papers [48,71] mentioned blockchain-based tele-dermatology and other telecare platforms. Blockchain enhances telemedicine platforms by (1) anonymizing patients' information and access; (2) decentralizing control and allowing patients to manage their data and consultations independently; (3) providing a secure global EHR to share and exchange data for research, teaching, or clinical work; (4) limiting health care fraud by using smart contracts to automate invoice processing; (5) eliminating mediators and reducing administrative costs; and (6) fundraising through ICOs (eg, DocCoin, PointNurse, and Medical Chain). In the case of DermaNet, the tele-dermatology platform, blockchain shortens the virtual distance between patients and providers by efficiently sharing information about changes in dermatological disease, improving quality of care, and allowing actors to trust each other [44]. Building trust between health care providers and patients supports patient engagement and empowerment.

DPSs and PSNs

DPSs notarize *data to provide legal evidence for medical disputes and medical negligence* [58]; 1% (1/70) of papers addressed how blockchain can enhance DPSs [58].

Current DPSs may be unreliable as persons can tamper with the data, reducing the accuracy of patient information. If a third-party notarized company is contracted to provide preservation services, personal information may be leaked, and it is difficult to guarantee the reliability and availability of DPSs [66].

XSL•FO
RenderX

Blockchain provides an immutable, distributed transaction ledger that reduces the abovementioned challenges associated with DPSs. Blockchain-based DPSs provide a reliable and secure data storage solution that prevents data tampering and allows transactions to be synchronized. Decentralization makes the DPS more secure, as attackers are not able to attack all nodes simultaneously. Information security and data integrity are strengthened as blockchain's synchronization and consensus algorithms ensure that all copies of the data are the same at every node. The DPS is capable of storing patient data anonymously, thus preserving patient privacy in the event of tampering or breach.

A total of 1% (1/70) of papers focused on the creation of a secure system for PSN-based health care systems [54]. PSN-based health care systems enable users to share health data that medical sensors collect, for disease monitoring and control and remote medical care. Due to its distributed design, blockchain reinforces the synchronization of transactions and operations in such systems, increasing health information's security, reliability, and accuracy. These features contribute to patients' confidence in the security and privacy of their data and promote health care engagement and empowerment.

Health Care Provider's Health Information Access and Use

HIEs, RPM Systems, and Medical Image Sharing Platforms

Blockchain-based HITs such as HIE, RPM systems that leverage Internet of Things (IoT) technologies, and medical image sharing systems provide benefits to health care providers by improving health information access and use.

HIE refers to the process of electronic transfer of patient health information and medical data among health care providers and institutions so as to provide health care providers with better information for diagnosis and patient treatment. The potential benefits of HIE include decreasing rates of patient readmission, avoiding medication errors, improving diagnoses, and decreasing duplicate testing [20]. However, HIE systems suffer from challenges such as threats to patient privacy, information security threats and vulnerabilities, poor integration of disparate data sources, and dependency on centralized data storage [16]. Furthermore, most HIE are designed for health care providers, and patients cannot access their data in the HIE when they visit hospitals outside their home systems; 9% (6/70) papers discussed how blockchain could improve HIE systems to facilitate the sharing of health records between providers, patients, and hospitals [16,32,39,50,57,79]. These papers suggest that blockchain's features of decentralized transaction validation, data provenance insurance, data sharing, and data integration may resolve HIE's challenges. For example, as blockchain technology utilizes distributed databases to store all transactions, it is possible to design a permissioned blockchain system to connect multiple EHR databases from different clinical sites to perform information exchange [20]. The private chain allows only authorized users to join the system, and smart contracts regulate all transactions. These blockchain-enabled smart contracts ensure data provenance and immutability, decentralization, restricted access to and anonymity of patients' records, and other benefits.

RPM enables health care providers to use sensor data to monitor patients outside of the clinical setting [12,41,62,66,75]; 7% (5/70) papers explored the use of blockchain in RPM systems. As the RPM realm expands, concerns about efficient and secure transmission of medical data arise as data come from different sensors, are a lucrative target for hackers, and must conform to health care data security and privacy regulations. Furthermore, patient treatment events initiated by IoT must be securely logged to show the patient's treatment and a record of who permitted it, to protect the integrity of the patient's care and maintain an accurate timeline of events.

These papers proposed the design of blockchain-based IoT technology to build a network architecture to better manage data from remote sensors to address these challenges. For example, smart contracts could support real-time patient monitoring and medical interventions by sending notifications to patients and medical practitioners while maintaining a secure record of who initiated these activities [77]. Blockchain-based IoT systems could also allow patients to maintain anonymity, create a permanent digital trace of patients' health records, and prevent a single point of control or failure through a highly decentralized system structure. These features would allow providers to securely access their patients' data and better control their patients' conditions.

A total of 3% (2/70) of papers focused on the development of blockchain-based cross-domain medical image sharing platforms [24,65]. In the past, digital images between providers were predominantly shared through a physical copy (eg, a CD) [24]. To address issues associated with digital image transfer, the Radiological Society of North America developed the Image Sharing Network that allows digital image transfer through a third-party clearinghouse. This raised new concerns regarding data storage centralization and intermediaries' involvement in the medical image exchange process [24]. Blockchain solves these concerns by decentralizing the entire system by removing the need for a central intermediary, creating immutable records that ease communication between providers, and eliminating third-party access to protected health information. This approach satisfies many criteria of an interoperable health system and is generalizable to other contexts [56].

Health Care Research

Clinical Trial Management and Health Care Research

When conducting medical research, data records are usually widely accessible, but the patients to whom they refer are anonymous [84]. Blockchain improves health care research practices and supports patient data anonymization during the research process through sophisticated cryptographic techniques. A total of 9% (6/70) of papers focused on creating blockchain-based systems capable of improving clinical trial management and enhancing patients' trust in health care research, impacting activities such as collecting, storing, and tracking patients' informed consent; improving data integrity; and sharing clinical data between providers [10,16,52,61,70,80]. This is particularly relevant to clinical investigator–related deficiencies, as the US Food and Drug Administration reports that about 10% of clinical trials suffer from consent collection issues such as unapproved forms and outdated consent

XSL•FO
RenderX

documents [10]. In BlockTrial [52], consent algorithms in the blockchain system can enable clinical research stakeholders to share and update patient consent forms and retrieve relevant data. In addition, the system can empower patients to become more active and fully informed research partners [52]. Blockchain can also ensure data accuracy in clinical trials without confirmation by a third-party contract research agency, thus reducing the cost of clinical trials [63].

In the health care analytics domain, blockchain was used in privacy-preserving predictive modeling [47] and a cloud-based health resource–sharing model used for breast tumor diagnosis [79]. In the case of blockchain-based privacy-preserving modeling, the removal of the *server* role eliminated a single point of control [47]. This single point of control poses various security risks, as multiple institutions that want to create a generalizable model on health or genomic data would have to rely on a single party.

Blockchain HIT Implementation Challenges

As shown in Table 3, despite blockchain's multiple advantages for health care use, studies indicated various downsides. These limitations point to open areas for research; 16% (11/70) papers revealed that blockchain technologies have the capability of enhancing security and privacy, yet have security and privacy vulnerabilities and weak access control mechanisms [23,24,32-34,41,57,58,70,75,79]. For example, blockchain-based structural health monitoring systems do not have mechanisms to guarantee the security of the data placed on the blockchain, although it is expected that the blockchain would enhance the security of the overall system [75]. Furthermore, research has shown that although it is almost impossible to alter and modify blockchain records in an EMR system, it is possible to tamper with and hack smart contracts [40].

Researchers have discussed high computing power and implementation costs in 16% (11/70) papers [12,23,25,29,47,48,52,57,61,76,79]. Prior research shows that blockchain-based EHRs consume significant computational power and take a large amount of time to execute tasks [29]. Blockchain-based clinical trial platforms that utilize Ethereum blockchain technology are also costly [52].

Researchers raised concerns about lengthy response and transaction processing times in 7% (5/70) papers [12,32,36,41,57]. For example, blockchain RPM systems rely on sensor-based data that must be accumulated, acted upon, and added to the blockchain, potentially introducing delays [41].

The potential resistance to blockchain adoption by patients and providers was raised in 7% (5/70) papers [11,24,32,48,70]. Public perceptions of blockchain technology might stand in the way of the successful implementation of blockchain-based EHRs because of blockchain's nascency and association with the negative use of blockchain-based technology such as cryptocurrencies within black markets [11].

Although researchers have indicated that blockchain-based electronic health (eHealth) structures may resolve the incompatibility of existing eHealth structures [21,22,26,27,30,55], new blockchain health systems may be incompatible with health care legacy systems [13,48,52,61,76]. Therefore, more research is needed to investigate approaches to address the interoperability between legacy and blockchain systems.

A total of 6% (4/70) of papers pointed out issues associated with blockchain's consensus algorithms [33,36,51,75]. For example, in small networks with a limited number of peers, the Practical Byzantine Fault Tolerance algorithm, which is designed to prevent catastrophic system failures due to malicious nodes, can be disabled if more than a third of the peers are offline at the same time. Therefore, it is important to increase the number of peers to prevent malicious peers from occupying the entire system.

Linking participants to their digital identity is a predominant concern [10,13,25]. Guaranteeing the blockchain informant's identity and authenticity is not foolproof, whether the informant is a patient, physician, or a sensor connected to a patient. Although blockchain technology helps prevent data block fraud, it is challenging to ensure that only authentic informants can access the health records and to prevent attacks on the blockchain [55]. Although blockchain demonstrates the potential to preserve the privacy of patients, further testing for security, privacy, and user authentication is needed.

Another constraint is data storage limitations, as discussed in 4% (3/70) studies [23,25,60]. Stakeholders are constrained by the amount of data that they can store in the blockchain. Furthermore, legislation such as Article 17 Right to Erasure, in Europe's General Data Protection Regulation (GDPR), gives citizens the right to request the modification and deletion of personal data, which is difficult, given the permanency of data recorded on the blockchain [19].

XSL•FO
RenderX

Table 3. Blockchain health information technology barriers and challenges.

Barrier or challenge	Frequency (N=70), n
No guarantee of security or privacy	11
Computing power and implementation costs	11
Users' resistance to a new technology	5
Long response time	5
Integration of existing electronic health structures into blockchain	5
Issues (eg, lack of efficiency) associated with consensus algorithms	4
Linking participants to their digital identity	3
Data storage limitation	3
Not mentioned	44

Discussion

Principal Findings

We presented the current state of research on blockchain technologies in patient care. We identified 3 main categories of research: data privacy and security, interoperability, and health care outcomes. We highlighted the health care applications that leverage blockchain technology. These health care applications include MIMS (EHR and EMR), PHRs, mHealth and telemedicine, DPSs and social networks, HIEs, remote monitoring systems, and medical research systems. These applications may improve patient engagement and empowerment, improve health care providers' access to information, and enhance health care information use for medical research. However, several challenges and implementation barriers exist, such as security and privacy vulnerabilities, user resistance, high computational power and implementation costs, inefficient consensus algorithms, and challenges integrating blockchain with existing HIT. Despite blockchain's disruptive potential, it is important that blockchain's limitations are further examined and understood and available alternatives considered before embarking upon any large-scale blockchain implementation.

Future Directions

Following are some of the challenges that have received the least attention in the literature. Therefore, researchers should focus on investigating the following research questions.

1. How can blockchain's interoperability and compatibility with existing HIT infrastructures be improved?
2. How can blockchains storage limitations be addressed?
3. How can blockchain adoption be promoted?
4. What blockchain HIT research should be pursued to disrupt health care?

Improving Blockchain's Interoperability and Compatibility With Existing HIT Infrastructure

Organizations need to understand how to connect their HIT blockchain to compatible blockchains, noncompatible blockchains, or nonblockchain platforms. Research should focus on ascertaining effective integration governance models and new interoperability and data standards and exploring performance optimization approaches. Researchers should also explore the feasibility of integration approaches for health care systems, such as cross-authentication (for compatible blockchains), oracles (which transfer external data to the blockchain for on-chain use), or application programing interfaces (for incompatible blockchains). The feasibility of translators that use open standards such as Health Level 7 and Fast Healthcare Interoperability Resources and *open* EHR or an equivalent ontology to connect proprietary systems to blockchain HIT is an open area of research [13]. Blockchain silos that different stakeholders create present complexity and interoperability challenges that should be investigated. An understanding of the legal and regulatory implications of blockchain interoperability is needed, as security and privacy concerns are pertinent issues.

Addressing Blockchain Storage Limitations

HIT researchers are becoming aware of the technical limitations regarding data storage on the blockchain, and researchers are beginning to explore alternative approaches to conform with the GDPR. Recommendations include recording metadata such as addresses, hash values, and time stamps on the blockchain while storing PHI off-chain elsewhere, such as in the cloud or on hospital servers [23]. However, splitting data storage may degrade system performance. Future research should determine feasible ways to optimize the cost, performance, and efficiency of implementations that split data storage on and off the blockchain.

Promoting Adoption of Blockchain

The public's negative perception of blockchain applications, such as cryptocurrencies, contributes to user resistance [11]. Lessons learned from the financial technology domain show that users are motivated to adopt cryptocurrencies when these applications are aligned with users' value systems [86]. To support this perspective, nascent HIT research shows that patients hold a favorable attitude toward the implementation of blockchain-based HIE mechanisms for privacy protection, coordination, and information exchange purposes [32]. User resistance may also arise because stakeholders do not understand blockchain technology. Researchers can explore how industry consortia of influential health care players working together to educate stakeholders and illustrate proof of concept may increase network effects and spur adoption. Further research could also

XSL•FO
RenderX

explore what incentivizes stakeholders to work together to adopt blockchain and to find common solutions. The role of the blockchain HIT startup ecosystem in stimulating user adoption is also a relevant area of research. Researchers can also explore how to lift adoption barriers presented by legal and regulatory constraints caused by legislation such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the GDPR, which did not contemplate the unique ways in which blockchain technology handles data privacy and security. Considerations of what should be stored on and off chains are important for HIPAA and GDPR compliance and adoption.

Pursuing Research Opportunities for Disruption of Health Care

HIT researchers are currently focused on improving health care applications, but there is room for researchers to explore blockchain HIT's disruptive potential. Most HIT research has focused on the use of smart contracts in HIT applications. It is worth noting that smart contracts are the building blocks of decentralized autonomous organizations (DAOs) and decentralized applications and services (D-APPs). DAOs and D-APPs have the potential to facilitate novel DAOs to disrupt patient care in areas such as organ and blood donations, electronic prescriptions, laboratories, and other diagnostic services.

Limitations

This review did not focus on the use of blockchain in the health industry outside of patient care, such as in the health insurance marketplace, pharmaceutical industry supply chain, or the health care provider credentialing process where blockchain is being proposed to reinvent these industries. The COVID-19 pandemic has exposed many high-value use cases where blockchain-enabled technology could address evolving societal needs and support accelerated responses to disruptions, such as supply chain management to connect health care organizations to trusted sources for necessary medical supplies, tracking and monitoring drugs' origin and journey from pharmacy to patient administration, and verifying credentials of clinicians and health care professionals volunteering at hospitals to alleviate workforce demands. A further review could explore blockchain research in these areas as part of the pandemic response.

Comparison With Prior Work

Prior reviews included conceptual papers, industry reports, and empirical research and primarily focused on identifying use cases and associated challenges [1-3,6,87-89]. To ascertain the current state of research, this review only focused on empirical research on the use of blockchain for patient care. Prior reviews included studies up to 2018 [1-3,6,87-89]. In this review, 58 of the 70 included empirical research studies were published in 2018 and 2019. We further distinguish this review by describing how blockchain HIT research focused on patient care evolved over time, from 2016 to 2019.

We build upon a systematic review of the blockchain HIT literature up to 2018 conducted by Agbo et al [6], which differs in its focus, inclusion criteria, results, and research agenda. Specifically, Agbo et al [6] identified 4 blockchain applications used for patient care: EMRs (which we expanded to include EHRs and PHRs); research; RPM; and health analytics, which we further broke down into predictive analytics and classification modeling. This review updates the blockchain HIT applications reported in the study by Agbo et al [6] by identifying 5 additional discrete categories of blockchain HIT research focused on patient care: HIEs, medical image sharing, mHealth and telemedicine, PSNs, and DPSs. Agbo et al [6] identified 5 challenges when implementing blockchain HIT applications: interoperability, security and privacy, scalability, speed, and patient engagement. Challenges found in our review but not included in the study by Agbo et al [6] included computing power and implementation costs, user resistance during implementation, consensus algorithm inefficiencies, difficulties in linking patients to their digital identity, and data storage limitations. To validate our search terms, we performed an alternative search of the literature using the search terms provided by Agbo et al [6]. However, the search terms we developed provided more relevant results for our purposes than the search terms by Agbo et al [6].

Conclusions

We have presented the current state of research on the use of blockchain technologies with a focus on patient care. Although blockchain presents significant potential for disrupting health care, most ideas are in their infancy. With 51% of the research focused on medical information systems such as EHR and EMR and 53% of the research focused on data security and privacy issues, this review shows that HIT research is primarily focused on the use of blockchain technologies to address the current challenges HIT faces. Future research can focus on how blockchain can disrupt patient care and help overcome the challenges in health care delivery post-COVID-19 by creating new decentralized organizations, applications, and services.

Acknowledgments

The authors greatly appreciate the help of Dr Xuan Tan for screening the papers. The content is solely the responsibility of the authors.

Authors' Contributions

KC and MC conceived the idea for this study. PD, KC, and MC designed the study. PD conducted the literature review and drafted the analysis. PD, KC, and MC interpreted the results and drafted the manuscript. All authors read, revised, and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[DOCX File , 19 KB - jmir_v22i7e18619_app1.docx]

Multimedia Appendix 2
Summary of the included studies.
[DOCX File , 30 KB - jmir_v22i7e18619_app2.docx]

References

1. McGhin T, Choo K, Liu C, He D. Blockchain in healthcare applications: research challenges and opportunities. J Netw Comput App 2019 Jun;135:62-75 [FREE Full text] [doi: 10.1016/j.jnca.2019.02.027]
2. Hölbl M, Kompara M, Kamišalić A, Zlatolas LN. A systematic review of the use of blockchain in healthcare. Symmetry 2018 Oct 10;10(10):470 [FREE Full text] [doi: 10.3390/sym10100470]
3. Zheng Z, Xie S, Dai H, Chen X, Wang H. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In: Proceedings of the IEEE International Congress on Big Data. 2017 Presented at: BigData Congress'17; June 25-30, 2017; Honolulu, HI, USA. [doi: 10.1109/bigdatacongress.2017.85]
4. Alhadhrami Z, Alghfeli S, Alghfeli M, Abedlla J, Shuaib K. Introducing Blockchains for Healthcare. In: International Conference on Electrical and Computing Technologies and Applications. 2017 Presented at: ICECTA; November 21-23, 2017; Ras Al Khaimah, United Arab Emirates. [doi: 10.1109/icecta.2017.8252043]
5. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
6. Agbo CC, Mahmoud QH, Eklund JM. Blockchain technology in healthcare: a systematic review. Healthcare (Basel) 2019 May 4;7(2):- [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
7. 2019 Cost of A Data Breach Study Reveals Increase in US Healthcare Data Breach Costs. HIPAA Journal. 2019. URL: https://www.hipaajournal.com/2019-cost-of-a-data-breach-study-healthcare-data-breach-costs/ [accessed 2020-02-01]
8. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Oct;42(9):572-575 [FREE Full text] [Medline: 28890644]
9. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 5;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
10. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]
11. Cunningham J, Ainsworth J. Enabling patient control of personal electronic health records through distributed ledger technology. Stud Health Technol Inform 2017;245:45-48. [Medline: 29295049]
12. Dey T, Sunderkrishnan S, Jaiswal S, Katre N. HealthSense: A Medical Use Case of Internet of Things and Blockchain. In: International Conference on Intelligent Sustainable Systems. 2017 Presented at: ICISS; December 7-8, 2017; Palladam, India URL: https://doi.org/10.1109/ISS1.2017.8389459 [doi: 10.1109/iss1.2017.8389459]
13. Roehrs A, da Costa CA, da Rosa RR. OmniPHR: a distributed architecture model to integrate personal health records. J Biomed Inform 2017 Jul;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
14. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MEDshare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 2017;5:14757-14767. [doi: 10.1109/access.2017.2730843]
15. Xia Q, Sifah E, Smahi A, Amofa S, Zhang X. BBDS: blockchain-based data sharing for electronic medical records in cloud environments. Information 2017 Apr 17;8(2):44 [FREE Full text] [doi: 10.3390/info8020044]
16. Zhuang Y, Sheets L, Shae Z, Tsai JJ, Shyu C. Applying blockchain technology for health information exchange and persistent monitoring for clinical trials. AMIA Annu Symp Proc 2018;2018:1167-1175 [FREE Full text] [Medline: 30815159]
17. Alexaki S, Alexandris G, Katos V, Petroulakis N. Blockchain-based Electronic Patient Records for Regulated Circular Healthcare Jurisdictions. In: 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks. 2018 Presented at: CAMAD'18; September 17-19, 2018; Barcelona, Spain URL: https://doi.org/10.1109/CAMAD.2018.8514954 [doi: 10.1109/camad.2018.8514954]
18. Du Y, Liu J, Guan Z, Feng H. A Medical Information Service Platform Based on Distributed Cloud and Blockchain. In: IEEE International Conference on Smart Cloud. 2018 Presented at: SmartCloud'18; September 21-23, 2018; New York, NY, USA URL: https://doi.org/10.1109/SmartCloud.2018.00014 [doi: 10.1109/smartcloud.2018.00014]
19. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: efficient and secure medical data sharing via blockchain. J Med Syst 2018 Jul 21;42(8):136. [doi: 10.1007/s10916-018-0993-7] [Medline: 29931655]

XSL•FO
RenderX

20. Guo R, Shi H, Zhao Q, Zheng D. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. IEEE Access 2018;6:11676-11686 [FREE Full text] [doi: 10.1109/access.2018.2801266]
21. Kim M, Lee A, Kwon H, Kim J, Kim I. Sharing Medical Questionnaries Based on Blockchain. In: Proceedings of the International Conference on Bioinformatics and Biomedicine. 2018 Presented at: BIBM'18; December 3-6, 2018; Madrid, Spain URL: https://doi.org/10.1109/BIBM.2018.8621154 [doi: 10.1109/bibm.2018.8621154]
22. Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications. In: 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications. 2017 Presented at: PIMRC'17; October 8-13, 2017; Montreal, QC, Canada URL: https://doi.org/10.1109/PIMRC.2017.8292361 [doi: 10.1109/pimrc.2017.8292361]
23. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 8;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
24. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Informatics J 2019 Dec;25(4):1398-1411. [doi: 10.1177/1460458218769699] [Medline: 29692204]
25. Rahmadika S, Rhee K. Blockchain technology for providing an architecture model of decentralized personal health information. Int J Eng Bus Manag 2018 Aug;10:184797901879058 [FREE Full text] [doi: 10.1177/1847979018790589]
26. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. J Biomed Inform 2019 Apr;92:103140 [FREE Full text] [doi: 10.1016/j.jbi.2019.103140] [Medline: 30844481]
27. Silva C, Aquino G, Melo S, Egídio D. A fog computing-based architecture for medical records management. Wirel Commun Mob Com 2019 Feb 27;2019:1-16 [FREE Full text] [doi: 10.1155/2019/1968960]
28. Tang F, Ma S, Xiang Y, Lin C. An efficient authentication scheme for blockchain-based electronic health records. IEEE Access 2019;7:41678-41689 [FREE Full text] [doi: 10.1109/access.2019.2904300]
29. Vora J, Nayyar A, Tanwar S, Tyagi S, Kumar N, Obaidat M, et al. BHEEM: A Blockchain-Based Framework for Securing Electronic Health Records. In: IEEE Globecom Workshops. 2018 Presented at: GC Wkshps; December 9-13, 2018; Abu Dhabi, United Arab Emirates URL: https://doi.org/10.1109/GLOCOMW.2018.8644088 [doi: 10.1109/glocomw.2018.8644088]
30. Wang S, Wang J, Wang X, Qiu T, Yuan Y, Ouyang L, et al. Blockchain-powered parallel healthcare systems based on the ACP approach. IEEE Trans Comput Soc Syst 2018 Dec;5(4):942-950 [FREE Full text] [doi: 10.1109/tcss.2018.2865526]
31. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
32. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
33. Hang L, Choi E. A novel EMT integrity management based on a medical blockchain platform in hospital. Electronics 2019 Apr 25;8(4):467 [FREE Full text] [doi: 10.3390/electronics8040467]
34. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and scalable mhealth data management using blockchain combined with client hashchain: system design and validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]
35. Tian H, He J, Ding Y. Medical data management on blockchain with privacy. J Med Syst 2019 Jan 3;43(2):26. [doi: 10.1007/s10916-018-1144-x] [Medline: 30603816]
36. Yang J, Onik M, Lee N, Ahmed M, Kim C. Proof-of-familiarity: a privacy-preserved blockchain scheme for collaborative medical decision-making. Appl Sci 2019 Apr 1;9(7):1370 [FREE Full text] [doi: 10.3390/app9071370]
37. Zhou T, Li X, Zhao H. Med-PPPHIS: blockchain-based personal healthcare information system for national physique monitoring and scientific exercise guiding. J Med Syst 2019 Aug 14;43(9):305. [doi: 10.1007/s10916-019-1430-2] [Medline: 31410583]
38. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2018 Dec 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]
39. Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. J Diabetes Sci Technol 2019 Mar;13(2):248-253 [FREE Full text] [doi: 10.1177/1932296818790281] [Medline: 30047789]
40. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
41. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. J Med Syst 2018 Jul 6;42(7):130. [doi: 10.1007/s10916-018-0982-x] [Medline: 29876661]
42. Kurdi H, Alsalamah S, Alatawi A, Alfaraj S, Altoaimy L, Ahmed S. HealthyBroker: a trustworthy blockchain-based multi-cloud broker for patient-centered ehealth services. Electronics 2019 May 29;8(6):602 [FREE Full text] [doi: 10.3390/electronics8060602]

XSL•FO
RenderX

43. Hussein A, ArunKumar N, Ramirez-Gonzalez G, Abdulhay E, Tavares J, de Albuquerque V. A medical records managing and securing blockchain based system supported by a genetic algorithm and discrete wavelet transform. Cogn Sys Res 2018 Dec;52:1-11 [FREE Full text] [doi: 10.1016/j.cogsys.2018.05.004]
44. Ramani V, Kumar T, Bracken A, Liyanage M, Ylianttila M. Secure and Efficient Data Accessibility in Blockchain Based Healthcare Systems. In: IEEE Global Communications Conference. 2018 Presented at: GLOBECOM'18; December 9-13, 2018; Abu Dhabi, United Arab Emirates URL: https://doi.org/10.1109/GLOCOM.2018.8647221 [doi: 10.1109/glocom.2018.8647221]
45. Lo Y, Yang C, Chien H, Chang S, Lu C, Chen R. Blockchain-enabled iWellChain framework integration with the national medical referral system: development and usability study. J Med Internet Res 2019 Dec 4;21(12):e13563 [FREE Full text] [doi: 10.2196/13563] [Medline: 31799935]
46. Beinke JH, Fitte C, Teuteberg F. Towards a stakeholder-oriented blockchain-based architecture for electronic health records: design science research study. J Med Internet Res 2019 Oct 7;21(10):e13585 [FREE Full text] [doi: 10.2196/13585] [Medline: 31593548]
47. Kuo T, Gabriel RA, Ohno-Machado L. Fair compute loads enabled by blockchain: sharing models by alternating client and server roles. J Am Med Inform Assoc 2019 May 1;26(5):392-403. [doi: 10.1093/jamia/ocy180] [Medline: 30892656]
48. Mannaro K, Baralla G, Pinna A, Ibba S. A blockchain approach applied to a teledermatology platform in the Sardinian region (Italy). Information 2018 Feb 23;9(2):44 [FREE Full text] [doi: 10.3390/info9020044]
49. Cao S, Zhang G, Liu P, Zhang X, Neri F. Cloud-assisted secure ehealth systems for tamper-proofing EHR via blockchain. Inf Sci 2019 Jun;485:427-440 [FREE Full text] [doi: 10.1016/j.ins.2019.02.038]
50. Castaldo L, Cinque V. Blockchain-based logging for the cross-border exchange of ehealth data in Europe. In: Security in Computer and Information Sciences. Euro-CYBERSEC 2018. Communications in Computer and Information Science. London, UK: Springer; 2018.
51. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
52. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
53. Wang H, Song Y. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain. J Med Syst 2018 Jul 5;42(8):152. [doi: 10.1007/s10916-018-0994-6] [Medline: 29974270]
54. Zhang J, Xue N, Huang X. A secure system for pervasive social network-based healthcare. IEEE Access 2016;4:9239-9250 [FREE Full text] [doi: 10.1109/access.2016.2645904]
55. Chen L, Lee W, Chang C, Choo K, Zhang N. Blockchain based searchable encryption for electronic health record sharing. Future Gener Comp Syt 2019 Jun;95:420-429 [FREE Full text] [doi: 10.1016/j.future.2019.01.018]
56. Thwin T, Vasupongayya S. Blockchain-based access control model to preserve privacy for personal health record systems. Secur Commun Netw 2019 Jun 25;2019:1-15 [FREE Full text] [doi: 10.1155/2019/8315614]
57. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating health data sharing: a solution based on the internet of things and distributed ledger technologies. J Med Internet Res 2019 Jun 6;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
58. Li H, Zhu L, Shen M, Gao F, Tao X, Liu S. Blockchain-based data preservation system for medical data. J Med Syst 2018 Jul 28;42(8):141. [doi: 10.1007/s10916-018-0997-3] [Medline: 29956058]
59. Nguyen D, Pathirana P, Ding M, Seneviratne A. Blockchain for secure EHRs sharing of mobile cloud based e-health systems. IEEE Access 2019;7:66792-66806 [FREE Full text] [doi: 10.1109/access.2019.2917555]
60. Radhakrishnan B, Joseph A, Sudhakar S. Securing Blockchain based Electronic Health Record using Multilevel Authentication. In: 5th International Conference on Advanced Computing & Communication Systems. 2019 Presented at: ICACCS'19; March 15-16, 2019; Coimbatore, India URL: https://doi.org/10.1109/ICACCS.2019.8728483 [doi: 10.1109/icaccs.2019.8728483]
61. Zhang P, White J, Schmidt D, Lenz G, Rosenbloom S. FHIRChain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
62. Zhao H, Zhang Y, Peng Y. Lightweight Backup and Efficient Recovery Scheme for Health Blockchain Keys. In: 13th International Symposium on Autonomous Decentralized System. 2017 Presented at: ISADS'17; March 22-24, 2017; Bangkok, Thailand URL: https://doi.org/10.1109/ISADS.2017.22 [doi: 10.1109/isads.2017.22]
63. Dagher G, Mohler J, Milojkovic M. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities and Society 2018;39:283-297 [FREE Full text] [doi: 10.1016/j.scs.2018.02.014]
64. Pham H, Tran T, Nakashima Y. A Secure Remote Healthcare System for Hospital Using Blockchain Smart Contract. In: IEEE Globecom Workshops. 2018 Presented at: GC Wkshps'18; December 9-13, 2018; Abu Dhabi, United Arab Emirates URL: https://doi.org/10.1109/GLOCOMW.2018.8644164 [doi: 10.1109/glocomw.2018.8644164]
65. Tang H, Tong N, Ouyang J, Tang H, Tong N, Ouyang J. Medical Images Sharing System Based on Blockchain and Smart Contract of Credit Scores. In: 1st IEEE International Conference on Hot Information-Centric Networking. 2018 Presented

XSL•FO
RenderX

at: HotICN; August 15-17, 2018; Shenzhen, China URL: https://doi.org/10.1109/HOTICN.2018.8605956 [doi: 10.1109/hoticn.2018.8605956]
66. Uddin M, Stranieri A, Gondal I, Balasubramanian A. Continuous patient monitoring with a patient centric agent: a block architecture. IEEE Access 2018;6:32700-32726 [FREE Full text] [doi: 10.1109/access.2018.2846779]
67. Li X, Huang X, Li C, Yu R, Shu L. EdgeCare: leveraging edge computing for collaborative data management in mobile healthcare systems. IEEE Access 2019;7:22011-22025 [FREE Full text] [doi: 10.1109/access.2019.2898265]
68. Badr S, Gomaa I, Abd-Elrahman E. Multi-tier blockchain framework for IoT-EHRs systems. Procedia Comput Sci 2018;141:159-166 [FREE Full text] [doi: 10.1016/j.procs.2018.10.162]
69. What is Interoperability in Healthcare? Healthcare Information and Management Systems. 2020. URL: https://www.himss.org/what-interoperability [accessed 2020-02-01]
70. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
71. Ji Y, Zhang J, Ma J, Yang C, Yao X. BMPLS: blockchain-based multi-level privacy-preserving location sharing scheme for telecare medical information systems. J Med Syst 2018 Jul 30;42(8):147. [doi: 10.1007/s10916-018-0998-2] [Medline: 29961160]
72. Hyla T, Pejaś J. eHealth integrity model based on permissioned blockchain. Future Internet 2019 Mar 24;11(3):76 [FREE Full text] [doi: 10.3390/fi11030076]
73. Kaur H, Alam MA, Jameel R, Mourya AK, Chang V. A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment. J Med Syst 2018 Jul 10;42(8):156. [doi: 10.1007/s10916-018-1007-5] [Medline: 29987560]
74. Kubendiran M, Singh S, Sangaiah A. Enhanced security framework for e-health systems using blockchain. J Inf Process Syst 2019;15:239-250 [FREE Full text] [doi: 10.3745/JIPS.04.0106]
75. Jo BW, Khan RM, Lee Y. Hybrid blockchain and internet-of-things network for underground structure health monitoring. Sensors (Basel) 2018 Dec 4;18(12):- [FREE Full text] [doi: 10.3390/s18124268] [Medline: 30518124]
76. Liu W, Zhu S, Mundie T. Advanced Block-Chain Architecture for E-Health Systems. In: 19th International Conference on e-Health Networking, Applications and Services. 2017 Presented at: Healthcom'17; October 12-15, 2017; Dalian, China URL: https://doi.org/10.1109/HealthCom.2017.8210847 [doi: 10.1109/healthcom.2017.8210847]
77. Xiao Z, Li Z, Liu Y, Feng L, Zhang W, Lertwuthikarn T. EMRShare: A Cross-Organizational Medical Data Sharing and Management Framework Using Permissioned Blockchain. In: 24th International Conference on Parallel and Distributed Systems. 2018 Presented at: ICPADS'18; December 11-13, 2018; Singapore URL: https://doi.org/10.1109/PADSW.2018.8645049 [doi: 10.1109/padsw.2018.8645049]
78. Talukder A, Chaitanya M, Arnold D, Sakurai K. Proof of Disease: A Blockchain Consensus Protocol for Accurate Medical Decisions and Reducing the Disease Burden. In: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. 2018 Presented at: SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI'18; October 8-12, 2018; Guangzhou, China URL: https://doi.org/10.1109/SmartWorld.2018.00079 [doi: 10.1109/smartworld.2018.00079]
79. Zhu X, Shi J, Lu C. Cloud health resource sharing based on consensus-oriented blockchain technology: case study on a breast tumor diagnosis service. J Med Internet Res 2019 Jul 23;21(7):e13767 [FREE Full text] [doi: 10.2196/13767] [Medline: 31339106]
80. Sutton A, Samavi R, Doyle T, Koff D. Digitized Trust in Human-in-the-Loop Health Research. In: 16th Annual Conference on Privacy, Security and Trust. 2018 Presented at: PST'18; August 28-30, 2018; Belfast, UK URL: https://doi.org/10.1109/PST.2018.8514168 [doi: 10.1109/pst.2018.8514168]
81. What is Patient Engagement? Healthcare Information and Management Systems. 2016. URL: https://www.himss.org/library/patient-engagement-toolkit [accessed 2020-02-01]
82. World Health Organization. HO Guidelines on Hand Hygiene in Health Care. First Global Patient Safety Challenge Clean Care Is Safer Care. Geneva, Switzerland: World Health Organization; 2009.
83. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
84. Tang C, Li C, Yu X, Zheng Z, Chen Z. Cooperative mining in blockchain networks with zero-determinant strategies. IEEE Trans Cybern 2019:1-6 [FREE Full text] [doi: 10.1109/tcyb.2019.2915253]
85. Healthcare Data Breach Statistics. HIPPA Journal. 2018. URL: https://www.hipaajournal.com/healthcare-data-breach-statistics/ [accessed 2020-02-01]
86. Cousins K, Subramanian H, Esmaeilzadeh P. A value-sensitive design perspective of cryptocurrencies: a research agenda. Commun Assoc Inf Syst 2019:511-547 epub ahead of print [FREE Full text] [doi: 10.17705/1CAIS.04527]
87. Casino F, Dasaklis K, Patsakis C. A systematic literature review of blockchain-based applications: current status, classification and open issues. Telemat Inform 2019 Mar;36:55-81 [FREE Full text] [doi: 10.1016/j.tele.2018.11.006]
88. Drosatos G, Kaldoudi E. Blockchain applications in the biomedical domain: a scoping review. Comput Struct Biotechnol J 2019;17:229-240 [FREE Full text] [doi: 10.1016/j.csbj.2019.01.010] [Medline: 30847041]

XSL•FO
RenderX

89. Khezr S, Moniruzzaman M, Yassine A, Benlamri R. Blockchain technology in healthcare: a comprehensive review and directions for future research. Appl Sci 2019 Apr 26;9(9):1736 [FREE Full text] [doi: 10.3390/app9091736]

Abbreviations

CINAHL: Cumulative Index of Nursing and Allied Health Literature
DAO: decentralized autonomous organizations
D-APPs: decentralized applications and services
DPS: data preservation system
eHealth: electronic health
EHR: electronic health records
EMBASE: Excerpta Medica dataBASE
EMR: electronic medical records
GDPR: General Data Protection Regulation
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act of 1996
HIT: health information technology
ICO: initial coin offerings
IoT: Internet of Things
mHealth: mobile health
PHI: personal health information
PHR: personal health records
PSN: pervasive social networks
RPM: remote patient monitoring

Edited by G Eysenbach; submitted 14.03.20; peer-reviewed by P Martin, L Sheets; comments to author 30.03.20; revised version received 02.05.20; accepted 13.05.20; published 20.07.20.

Please cite as:
Durneva P, Cousins K, Chen M
The Current State of Research, Challenges, and Future Research Directions of Blockchain Technology in Patient Care: Systematic Review
J Med Internet Res 2020;22(7):e18619
URL: http://www.jmir.org/2020/7/e18619/
doi:10.2196/18619
PMID:32706668

©Polina Durneva, Karlene Cousins, Min Chen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Integrating Health Technologies in Health Services for Syrian Refugees in Lebanon: Qualitative Study

Reem Talhouk[1,2], PhD; Chaza Akik[3], DrPH; Vera Araujo-Soares[4], PhD; Balsam Ahmad[4], PhD; Sandra Mesmar[3], MPH; Patrick Olivier[5], PhD; Madeline Balaam[6], PhD; Kyle Montague[2,7], PhD; Andrew Garbett[2], PhD; Hala Ghattas[3], PhD

[1]School of Design, Northumbria University, Newcastle upon Tyne, United Kingdom

[2]Open Lab, Computing, Newcastle upon Tyne, United Kingdom

[3]Center for Research on Population and Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[4]Institute of Health and Society, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[5]Faculty of Information Technology, Monash University, Melbourne, Australia

[6]KTH Royal Institute of Technology, Stockholm, Sweden

[7]Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Reem Talhouk, PhD
School of Design
Northumbria University
City Campus East 2
Newcastle upon Tyne
United Kingdom
Phone: 44 191 208 4642
Email: reem.talhouk@northumbria.ac.uk

Abstract

Background: Lebanon currently hosts around one million Syrian refugees. There has been an increasing interest in integrating eHealth and mHealth technologies into the provision of primary health care to refugees and Lebanese citizens.

Objective: We aimed to gain a deeper understanding of the potential for technology integration in primary health care provision in the context of the protracted Syrian refugee crisis in Lebanon.

Methods: A total of 17 face-to-face semistructured interviews were conducted with key informants (n=8) and health care providers (n=9) involved in the provision of health care to the Syrian refugee population in Lebanon. Interviews were audio recorded and directly translated and transcribed from Arabic to English. Thematic analysis was conducted.

Results: Study participants indicated that varying resources, primarily time and the availability of technologies at primary health care centers, were the main challenges for integrating technologies for the provision of health care services for refugees. This challenge is compounded by refugees being viewed by participants as a mobile population thus making primary health care centers less willing to invest in refugee health technologies. Lastly, participant views regarding the health and technology literacies of refugees varied and that was considered to be a challenge that needs to be addressed for the successful integration of refugee health technologies.

Conclusions: Our findings indicate that in the context of integrating technology into the provision of health care for refugees in a low or middle income country such as Lebanon, some barriers for technology integration related to the availability of resources are similar to those found elsewhere. However, we identified participant views of refugees' health and technology literacies to be a challenge specific to the context of this refugee crisis. These challenges need to be addressed when considering refugee health technologies. This could be done by increasing the visibility of refugee capabilities and configuring refugee health technologies so that they may create spaces in which refugees are empowered within the health care system and can work toward debunking the views discovered in this study.

(J Med Internet Res 2020;22(7):e14283) doi:10.2196/14283

XSL•FO
RenderX

KEYWORDS

Syrian refugees; Lebanon; health technologies; eHealth; mHealth; primary health care

Introduction

Lebanon currently hosts around one million Syrian refugees [1]. This influx of refugees has resulted in a 30% increase in the country's total population size, placing a burden on the Lebanese health care system [2]. However, despite political and social turmoil and a shortage in international financing and support in addressing the refugee crisis, the Lebanese health care system has proven to be resilient through coordinating efforts of various stakeholders (international and local) and maintaining a diverse health system while strengthening infrastructure and health and human resources [3]. As a result, refugee health care provision has been integrated into the Lebanese health care system with primary health care centers (PHCs) as the first point of entry into the health care system [3].

Technology that facilitates the transmission of health-related information (eHealth) [4] has been considered by the World Health Organization to be an effective tool that improves equity and quality of health care provision while reducing costs [5]. As such, it is argued that eHealth is an enabler of achieving universal health coverage, which is part of the agenda of meeting Sustainable Development Goal 3 to "ensure healthy lives and promote well-being for all at all ages" [5]. More specific to the context of refugees, eHealth has been found to aid in the monitoring and evaluation of refugee health statuses during times of protracted crises [6].

The increasing ubiquity of technologies among refugee communities has created a space in which refugee communities may engage with aid services through technologies [7]. Interviews with over 100 Syrian refugees in Lebanon found that most interviewees have access to a cell phone, of which 40% had smartphones [8]. A study with Syrian refugees in Lebanon [9] identified that refugees in informal tented settlements in rural Lebanon have access to at least one smartphone per household and access the internet through a Wi-Fi network set up in the settlement. Furthermore, Talhouk et al [9], through a series of focus groups, found that Syrian refugees in rural Lebanon are highly motivated to use smartphones as a means of communicating with health care providers and accessing health services provided by PHCs.

Several projects in Lebanon have begun integrating eHealth and mHealth, a category of eHealth technologies that use mobile phones for health services [10], to strengthen PHCs and improve the delivery of health care services to Syrian and Palestinian refugees as well as to Lebanese citizens. One initiative aimed to improve quality of care delivered by providing health care providers with electronic learning materials allowing for peer-to-peer exchanges through online forums [11,12]. An mHealth app for refugees was also piloted with clinicians in PHCs in Lebanon as a means of improving the provision of services to refugees suffering from chronic diseases and was shown to improve a number of quality of care indicators [13]. It is important to note that the mHealth tool used patient's cell phone number as an identifier in order to ensure that clinicians can access patient medical records if patients change locations [13].

Saleh et al [14] conducted a randomized controlled trial (RCT) with PHCs in rural Lebanon and in Palestinian refugee camps in which short message service (SMS) messages communicating medical information, importance of compliance, and appointment reminders were sent to participants in the intervention group and were shown to significantly improve blood pressure control and hemoglobin A_{1C} levels used for the diagnosis and monitoring of diabetes. However, this RCT also indicated that SMS had no significant impact on patients' use of primary health care services [14]. Also in collaboration with PHCs in rural Lebanon and Palestinian refugee camps, a netbook application was piloted as an eHealth tool to support community health workers in screening community members for diabetes and hypertension and subsequently referring them to primary health care clinics [4]. The study highlighted the ability of eHealth community-based interventions to identify new cases of diabetes and hypertension and refer them to the appropriate health services [4].

Low levels of agency in the Syrian refugee population in Lebanon and their low level of access and use of antenatal care services has also been the subject of mHealth interventions. Talhouk et al [15] piloted community health radio shows through mobile calls that delivered health information and responded to refugees' reproductive health questions. These interactions around reproductive health led to an increase in refugee agency and trust within patient-provider relationships [15]. A quantitative study conducted with primary health care providers across 22 PHCs in Lebanon showed high health care provider readiness to adopt eHealth and encouraged the Lebanese Ministry of Public Health (MoPH) and policy makers to enhance and scale up eHealth initiatives within the primary health care system [18]. Additionally, based on their success in piloting eHealth and mHealth technologies in PHCs run by the MoPH and the United Nations Relief and Works Agency for Palestine Refugees in the Near East [4,14], the authors called for the scaling up of mHealth and eHealth interventions to the wider PHC network in Lebanon (ie, PHCs funded and managed by other health care organizations).

While research has been conducted on eHealth systems for refugees in Lebanon, no study provides a qualitative perspective that engages with multiple stakeholders on the potential of implementing eHealth for Syrian refugees in Lebanon with the aim of meeting SDG3. Indeed, the existing studies to date do not explore the added considerations needed to integrate eHealth technologies into primary health care services specifically for refugees. We present a qualitative study to gain a deeper understanding of the potential for technology integration in the provision of primary health care in the context of the protracted Syrian refugee crisis in Lebanon. We use reproductive and maternal health care as an entry point as this research is part of a wider project on technologies for refugee women's health.

XSL•FO
RenderX

The findings from interviews with participants highlighted factors that need to be considered when implementing eHealth projects, such as (1) varying resources available at PHCs, (2) mobility of refugee populations, and (3) varying views of refugee technology and health literacies. In our discussion, we provide practical and theoretical implications for integrating eHealth technologies for refugee health care that include countering views regarding refugee health and technology literacies.

Methods

Data Collection

We recruited key informants and health care providers involved in the provision of reproductive and maternal health services to refugees. Participants spanned the humanitarian health response system both vertically (from policy makers to clinical staff) and horizontally (participants from multiple organizations). Key informants were identified from meeting minutes published on the United Nations High Commissioner for Refugees Regional Refugee Response Interagency portal for Lebanon [1]. Potential participants were approached via email, and a convenience sample was drawn based on responses (response rate: 80%). Three PHCs catering to large numbers of Syrian refugees were purposively selected. The PHCs were situated in the West Bekaa region of Lebanon, and the centers varied in terms of management and affiliation with the MoPH primary health care network. Participants consisted of a convenience sample of health care providers available at the PHCs on the days of the research visit (response rate: 100%). Recruitment was completed once data saturation was met.

A total of 17 face-to-face semistructured interviews lasting around an hour each were conducted with key informants (n=8) and health care providers (n=9) involved in the provision of health care to the refugee population (Table 1).

Data were collected between January and March 2015. Ethical approval was obtained from the institutional review board of the American University of Beirut and Newcastle University's ethics committee. All interviews were conducted in Arabic by local coauthors with a background in health system management and policy at the workplace of interviewees.

Table 1. Description of study participants detailing the gender and role of participants within the humanitarian health system. Participant code presented is used in the reporting of results.

Participants	Number	Gender	Participant code
Key informant			
Public sector employees	2	Male (1), female (1)	Key informant 1,2
Health center directors	2	Male (2)	Key informant 3,4
Academics	2	Male (1), female (1)	Key informant 5,6
Nongovernmental employees	2	Female (2)	Key informant 7,8
Health care provider			
Nurses	5	Male (2), female (3)	Nurse 1-5
Medical doctors	2	Male (2)	Doctor 1,2
Social workers	2	Female (2)	Social worker 1,2

Key Informant Interviews

Key informant interviews were conducted with PHC directors, nongovernmental organization employees, public sector employees, and academics working on eHealth within the Lebanese health care system. It is important to note that both PHC directors were also physicians. We probed on the current technologies being used, participant views on technology use, possible future use of technology, and willingness of their organizations to support the use of technology. This involved ascertaining equipment availability and organizational technological literacy. Interviewees were then asked to brainstorm about possible technologies that may be useful and the feasibility of these technologies within their current context.

Health Care Provider Interviews

Nurses, doctors, and social workers constituted this subset of study participants. In these interviews, questions focused on the day-to-day processes and challenges of health service provision to Syrian refugees and the interactions between health care providers and Syrian refugees. Interactions explored included those within the PHC, those within the community, and those conducted through technology. We also explored the ways in which health care providers are currently using technology, both formally and informally, and their comfort and willingness to use different forms of technology for refugee health care provision.

Data Analysis

Simultaneous translation, from Arabic to English, and transcription of the audio-recorded interviews was undertaken. Thematic analysis was conducted using NVivo 10 (QSR International) by two of the coauthors [16]. Codes were then grouped into themes and validated by a third coauthor. Data collection was concluded once data saturation was met and no new themes were emerging from the (data) interviews.

Results

Findings

Our analysis identified differences in levels of resources available across PHCs to be a key factor that needs to be

XSL•FO
RenderX

considered when integrating and scaling a health technology within the primary health care system. Additionally, the mobility of refugee communities was found to be a disincentive consideration for the integration of technologies specifically for refugee reproductive health. Lastly participants recounted experiences of engaging with refugees in which they identified varying views regarding refugee technology and health literacies. Such views are key considerations that should be accounted for when designing technologies that specifically connect refugees to the health care system.

Varying Resources Available at Primary Health Care Centers

A key consideration that needs to be accounted for when considering integrating and scaling up technologies within the refugee health care response system is the shortage in human resources as reported by key informant 2.

> *...they [center staff] might tell you we want to invest in more staff than in smartphones because we have a shortage in staff.* [Key informant 2]

In only one out of the three health centers visited by the research team did participants not identify shortages in staff and equipment as issues. Staff shortage was compounded by the increase in patient load due to the influx of Syrian refugees, resulting in nurses having to take work home with them.

> *[Since the refugee crisis began] in a short time my working hours increased [by] 6 hours, which then doubled and tripled.* [Nurse 3]

High patient load led to time constraints that discouraged the use of available health information systems (HISs).

> *We do not always have the time [to use the HIS]. We do a hundred things at the same time.* [Nurse 2]

Limited equipment available was also indicated to be a barrier to the use of existing technologies in PHCs.

> *We both are using one laptop...when she is working on the laptop, I cannot work on it... I have to enter 400 files and I'm 6 days behind... The time she [my colleague] spends on the laptop is time lost from my work.* [Nurse 3]
>
> *The main issue is [we need] to have more computers. All the work is being assigned to me [because there is one laptop] and I have to take the work home sometimes when this is not necessary.* [Nurse 1]

Interviews conducted at the PHC, which had an advanced HIS and was well resourced, revealed that shortage of resources was not a barrier for the integration of technology. The varying resources available at the different PHCs were attributed to the diversity of stakeholders involved in the health care provision of refugee reproductive health services. While some of the PHCs were part of the MoPH's primary health care network, others were not.

> *The current process is that we are working with MoPH but not all primary health care centers are within the network. Not all PHCs are supported by the MoPH.* [Key informant 2]

This resulted in varying levels of support regarding the health center management and funding received as well as the use of varying eHealth systems. Interviewees reported that while several health centers were in the process of transitioning at the time of data collection to the use of a new HIS designed and supplied by the MoPH, other health centers were not as they have their own systems. In fact, the MoPH was reported to be incentivizing PHCs to adopt health information technologies by adding them to their accreditation standards and making HIS integration a key compliance criterion for MoPH contracting.

> *[Implementation of] the information system is a main criterion of the [MoPH] contract with PHCs. The centers can use their own system or the MoPH's.* [Key informant 2]

However, both the contracts and accreditation are not extended to PHCs outside the MoPH primary health care network. The MoPH is also encouraging the use of HIS by providing training for health care providers on the system that they have developed. An interviewee indicated that continuous training is needed due to the high staff turnover at the PHCs.

> *Training sessions are continuously held centrally with follows-ups at PHCs* [Key informant 2]

Academics working on an eHealth project in Lebanon stated that they were collaborating with health centers managed by one stakeholder to overcome the variations in human and technological resources available among multiple stakeholders.

Mobility of Refugee Populations as a Disincentive to Investing in Refugee Health Technologies

The use of technology specifically for refugee populations was discouraged by participants when they considered refugees' unstable political and physical environments.

> *We are facing a difficulty in this [using mobile technology] actually. They might have a phone but are afraid of giving [their number]. For [security] reasons they don't give you the real number.* [Doctor 2]
>
> *Now regarding maintaining it [the technology], it might be low because they may go from one place to another or go back to Syria or change the location of the tents. You know things are happening, like a fire in a tent or the flooding of a tent. Forces of nature are impacting them a lot so they are [moving] quickly.* [Social worker 1]

Pilot eHealth interventions have excluded Syrian refugees due to their mobility.

> *Because they are a mobile population. For ethical reasons if a community health worker encounters Syrian refugees, they will be screened [for chronic diseases] but not included in our database.* [Key informant 6]

A PHC director attributed his hesitance in investing in using a technological system to improve the provision of health care for refugees to the high mobility of the community and their unstable presence in Lebanon.

XSL•FO
RenderX

So let us say we make something advanced and then they leave; everything we have done would be lost. [Key informant 3]

Varying Views of Refugee Technology and Health Literacies

Participants recounted experiences of engaging with refugees that resulted in varying views regarding refugee technology and health literacies. Participants highlighted that the low level of education among refugees limits their ability to use technologies as well as actively engage with the health care system. This notion was disputed by two participants who indicated that refugee women they are encountering are highly educated and highly concerned with their health and that incentivization may support refugees in engaging with health technologies.

Experiences of Technology Literacy of Refugees

All participants except two questioned the ability of Syrian refugees, who were reported to be of lower technology literacies and not capable of engaging through digital technologies with service providers.

You don't want to forget that they are of low educational status with all my respect for them. They [Syrians] have educated people but the pregnant women and the refugees...are of a low educational status... They would not know how to [create a voice] record as there is a lot of ignorance. [Nurse 5]

A participant, a social worker who visits refugee settlements, contradicted this view and stated that technological interventions would be feasible given the high education among a large proportion of female refugees.

There are very few ignorant women. Did you know that most of the women in camps have finished their college education? [Social worker 1]

Experiences of Health Literacy of Refugees

Although participants reported that many Syrian refugees attended health centers to address their health concerns, they indicated that there was a lack of motivation among the refugee community to engage with ways that facilitated access and provision of health care. Participants used Syrian refugees' lack of maintenance of medical papers and documentation as an analogy to refugees' inability to maintain technologies. When brainstorming the feasibility of using a technology that embeds audio recording hardware into a health education book/health journal, several participants stated:

They don't maintain a paper, I don't think they would maintain a file. [Nurse 4]

They barely maintain their ID; they bring it torn. Even the vaccination record they bring it ripped apart. [Key informant 1]

Another participant indicated that refugees are health literate however incentives are needed to support them in adopting and participating in any technology that connects them to health care providers.

You should tell them...if you take care of it [the technology], you will be compensated somehow in the end. [Key informant 7]

Discussion

Principal Findings

Our findings indicate that in the context of integrating technology into the provision of health care for refugees in a low or middle income country such as Lebanon, some barriers for technology adoption are common with those found elsewhere [17]. However, certain factors that hinder the use of technology were found to be specific to the provision of refugee health care. Such distinctions have not been made in previous studies assessing the eHealth readiness of Lebanese primary health care centers [18] nor in the literature on the use of technologies to improve refugee health [4,13,14].

The multitude of stakeholders involved within the underresourced health care system poses a huge barrier for the consistent use of technology across the primary health care system. Additionally, participants indicated time constraints to be a major factor that contributes to the perception of technology as a burden within health centers. Such barriers are consistent with literature on technology integration into health care systems [19] and with studies conducted in Lebanon where mHealth technologies situated in PHCs were found to be redundant given the presence of parallel technological systems already in use in the PHCs [13]. Additionally, previous eHealth and mHealth [4,13,14] studies in Lebanon have been primarily conducted with one organization managing multiple PHCs thus providing limited insight on the barriers of scaling up such technologies to the wider primary health care system. Within such contexts, Khalifehsoltani et al [19] recommends practical implications that constitute a top-down policy approach as a means of overcoming barriers to the integration of health technologies introduced by the fragmentation and underresourcefulness of the health care system. For successful eHealth integration in low and middle income countries, policies should be put in place that strengthen the government's capacity to plan, manage, regulate, and enforce eHealth policies in a way that incentivizes or mandates third party actors to adopt eHealth technologies [19]. Furthermore, a top-down approach to eHealth integration would allow for the exploration of interoperability of programs within health care systems that include a multiplicity of stakeholders [17,19]. Yassin et al [10] have indicated the Lebanese MoPH has begun introducing policies to encourage the uptake of eHealth and mHealth technologies in PHCs in a manner that would aid in overcoming the systematic challenges identified in our study.

Moreover, our findings provide theoretical contributions through highlighting that when it comes to integrating technologies for the provision of health care for refugees specifically, new factors are at play which require further consideration. Participants justified the unfeasibility of using technology within refugee health care based on their view of refugees as a mobile community that is of low health and technology literacy. Such notions were disputed by other participants who viewed refugees to be of high literacy and in need of support to engage in health

XSL•FO
RenderX

technologies. Such factors have not been previously reported within literature on integrating eHealth.

In light of such evidence, practical implications include raising awareness among health care providers and PHC management that existing research indicates that Syrian refugees in Lebanon have high technological and health literacies [8,9,20] and creating spaces in which refugees may practice and demonstrate their technological and health literacies. Interactions between health care providers and refugees mediated through technological interventions themselves may create spaces in which refugees are empowered within the health care system and can work toward debunking the negative views surfaced in our study. Virtual nurse agents were explored by Bickmore et al [21] as a means of empowering low health literacy hospital patients, while medical information visualization has been used to better inform patients of their medical conditions and consequently facilitate their communication with health care providers and inform decision making processes [22]. Such technologies aim at empowering patients by providing them with information regarding their health issues. Other technologies aim to empower patients in the health care system by facilitating their offline communications with health care providers. Jacobs et al [23] and Mirkovic [24] highlighted the potential for health technologies to allow cancer patients to prioritize the issues they would like to discuss during their face-to-face engagements with health care providers, while Colley et al [25] introduced a dual-sided tablet to be used during clinical consultation where one side of the tablet provides constructive information to the patient regarding what the doctor is inputting.

Technological designs may allow us to reconfigure the current modalities of communication to allow refugees to take on a more proactive role within their health care rather than being just recipients of a service [26]. Indeed, technology may be situated within refugee communities where they could initiate the use of the technology themselves, in order to communicate with health care providers and/or other parts of the humanitarian system. Such an approach would overcome the barriers of refugee mobility by giving refugees the possibility of initiating contact with the health care system regardless of where they are. Moreover, such a reconfiguration may play a role in changing the views of health care providers toward refugees by creating more open communication mediums where refugees can demonstrate their technological literacy and health agency [15]. Previous work with Syrian refugees has shown that such configurations mediated through refugee-led community radio shows provide health care providers with a deeper understanding of the refugee community and enable refugees to vocalize their health concerns [15].

Strengths and Limitations

This study provides an understanding of the barriers to integrating refugee health technologies in underresourced health systems such as Lebanon. The study was conducted in the West Bekaa region where a large number of refugees live; however, it may not represent the realities of other geographies in this context. Further research is thus needed in order to ascertain the generalizability of the results. Additionally, future studies and eHealth endeavors in similar contexts should actively engage with technology designers and developers who are stakeholders playing key roles in the configuration and deployment of health technologies.

The diversity of participants in this study reflects the different stakeholders and actors who engage in the provision of primary health care for refugees in Lebanon. The selection of a range of different stakeholders and actors ensures that the varying perspectives are reflected and information can be triangulated for reliability and validity. Our findings highlight factors that need to be further investigated before the implementation of eHealth for refugees.

Conclusions

The introduction of technologies into health care systems in underresourced countries for the provision of health services to refugee communities is likely to face barriers similar to those we identified in this study, particularly in contexts with fragmented and underresourced health systems. Our study provides theoretical contributions by highlighting barriers specific to how refugees are perceived by health care providers and actors within the health care system. While there is evidence that may disqualify such views, the technological literacy and health motivation of refugees should be made more visible to health care providers and other actors within the system. Technologies that empower refugees when accessing health care and those that challenge the current modalities in which refugees communicate with health care providers could play a role in debunking such perceptions in the future.

Acknowledgments

This research was funded by United Kingdom Engineering and Physical Science Research Council award EP/L016176/1 (EPSRC Centre for Doctoral Training in Digital Civics) and the Newcastle University Research Investment Fund. HG and CA received funding from the International Development Research Centre, Canada (grant number 106981-001) to contribute to this research.

Authors' Contributions

RT, SM, HG, PO, and MB conceived and designed the study. RT and SM collected the data and performed the analysis and were supervised and guided by CA. All authors contributed to the writing of the paper.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Figures at a glance. United Nations High Commissioner for Refugees. 2019. URL: http://www.unhcr.org/figures-at-a-glance.html [accessed 2019-08-16]
2. Ammar W, Kdouh O, Hammoud R, Hamadeh R, Harb H, Ammar Z, et al. Health system resilience: Lebanon and the Syrian refugee crisis. J Glob Health 2016 Dec;6(2):020704 [FREE Full text] [doi: 10.7189/jogh.06.020704] [Medline: 28154758]
3. Ammar W, Radi A, El-Jardali F. Comments on the article: Syrian refugees in Lebanon: the search for universal health coverage. Confl Health 2016;10:22. [doi: 10.1186/s13031-016-0090-9] [Medline: 27471548]
4. Saleh S, Alameddine M, Farah A, El Arnaout N, Dimassi H, Muntaner C, et al. eHealth as a facilitator of equitable access to primary healthcare: the case of caring for non-communicable diseases in rural and refugee settings in Lebanon. Int J Public Health 2018 Jun;63(5):577-588. [doi: 10.1007/s00038-018-1092-8] [Medline: 29546440]
5. Al-Shorbaji N. Telemedicine. QJM An Int J Med 2018;111(suppl_1):i90 [FREE Full text] [doi: 10.1093/qjmed/hcy200.206]
6. Mesmar S, Talhouk R, Akik C, Olivier P, Elhajj I, Elbassuoni S. The impact of digital technology on health of populations affected by humanitarian crises: recent innovations and current gaps. J Public Health Policy 2016;37(S2):167-200 [FREE Full text] [doi: 10.1057/s41271-016-0040-1]
7. Maitland C. Introduction. In: Digital Lifeline? ICTs for Refugees and Displaced Persons. Cambridge: MIT Press; 2018:1-14.
8. Lost: Syrian refugees and the information gap. Internews. 2013. URL: https://www.internews.org/sites/default/files/resources/Internews_Lost_SyriaReport_Nov2013_web.pdf [accessed 2020-06-03]
9. Talhouk R, Mesmar S, Thieme A, Balaam M, Olivier P, Akik C. Syrian refugees and digital health in lebanon: opportunities for improving antenatal health. Proc CHI Conf Human Factors Comput Syst 2016:331-342. [doi: 10.1145/2858036.2858331]
10. Yassin N, Khodor R, Baroud M. m-Health for healthcare delivery reform: prospects for Lebanese and refugee communities. In: Braithwaite J, Mannion R, editors. Healthcare Systems: Future Predictions for Global Care. Boca Raton: CRC Press; 2018:299-305.
11. Hanafy S. Digital health platforms to improve health of Syrian refugees in Lebanon. Int J Health Sci. 2017 Jun 10. URL: http://riss-ijhs.ca/archives/3271 [accessed 2020-06-03]
12. Saleh S, Sinha C. Using digital tech to improve life for refugees. International Development Research Centre. 2016 Dec 20. URL: https://www.idrc.ca/en/resources/perspectives/using-digital-tech-improve-life-refugees [accessed 2020-06-03]
13. Doocy S, Paik KE, Lyles E, Hei TH, Fahed Z, Winkler E, et al. Guidelines and mHealth to improve quality of hypertension and type 2 diabetes care for vulnerable populations in Lebanon: longitudinal cohort study. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e158 [FREE Full text] [doi: 10.2196/mhealth.7745] [Medline: 29046266]
14. Saleh S, Farah A, Dimassi H, El Arnaout N, Constantin J, Osman M, et al. Using mobile health to enhance outcomes of noncommunicable diseases care in rural settings and refugee camps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 13;6(7):e137 [FREE Full text] [doi: 10.2196/mhealth.8146] [Medline: 30006326]
15. Talhouk R, Bartindale T, Montague K, Mesmar S, Akik C, Ghassani A. Implications of synchronous IVR radio on Syrian refugee health and community dynamics. Proc 8th Int Conf Communities Technol 2017:193-202 [FREE Full text] [doi: 10.1145/3083671.3083690]
16. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
17. Gerber T. eHealth policy in LMICS: national frontiers, global challenges. In: Ranck J, editor. Disruptive Cooperation in Digital Health. Cham: Springer; 2016:157-164.
18. Saleh S, Khodor R, Alameddine M, Baroud M. Readiness of healthcare providers for eHealth: the case from primary healthcare centers in Lebanon. BMC Health Serv Res 2016;16(1):644 [FREE Full text] [doi: 10.1186/s12913-016-1896-2]
19. Khalifehsoltani S, Gerami M. E-health challenges, opportunities and experiences of developing countries. 2010 Presented at: International Conference on e-Education, e-Business, e-Management and e-Learning; 2010; Sanya p. 264-268 URL: http://ieeexplore.ieee.org/document/5432436/ [doi: 10.1109/ic4e.2010.119]
20. Talhouk R, Balaam M, Toombs A, Akik C, Ghattas H, Araújo-Soares V. Involving Syrian refugees in design research: lessons learnt from the field. Proc 2019 Designing Interact Syst Conf 2019:1583-1594 [FREE Full text] [doi: 10.1145/3322276.3322335]
21. Bickmore T, Pfeifer L, Jack B. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. Proc SIGCHI Conf Human Factors Comput Syst 2009:1265-1274 [FREE Full text] [doi: 10.1145/1518701.1518891]
22. Rajwan Y, Kim G. Medical information visualization conceptual model for patient-physician health communication. Proc Int Conf Health Informatics 2010:512 [FREE Full text] [doi: 10.1145/1882992.1883074]
23. Jacobs M, Clawson J, Mynatt E. Comparing health information sharing preferences of cancer patients, doctors, and navigators. Proc Conf Computer Supported Cooperative Work Soc Comput 2015:808-818 [FREE Full text] [doi: 10.1145/2675133.2675252]
24. Mirkovic J. Utilizing emerging technologies to promote more efficient face-to-face patient-clinician communication. Proc Conf Pervasive ubiquitous Comput 2013:1393-1396 [FREE Full text] [doi: 10.1145/2494091.2497362]
25. Colley A, Rantakari J, Häkkilä J. Dual sided tablet supporting doctor-patient interaction. Proc Conf Companion Computer Supported Cooperative Work Soc Comput 2015:13-16 [FREE Full text] [doi: 10.1145/2685553.2702672]

XSL•FO
RenderX

26. Olivier P, Wright P. Digital civics. 2015. URL: https://interactions.acm.org/archive/view/july-august-2015/digital-civics [accessed 2020-06-03]

Abbreviations

eHealth: electronic health
EPSRC: Engineering and Physical Science Research Council
HIS: health information system
mHealth: mobile health
MoPH: Ministry of Public Health
PHC: primary health care center
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 08.04.19; peer-reviewed by T Gerber, A Davoudi, H Yang; comments to author 14.09.19; revised version received 04.02.20; accepted 12.03.20; published 06.07.20.

Please cite as:
Talhouk R, Akik C, Araujo-Soares V, Ahmad B, Mesmar S, Olivier P, Balaam M, Montague K, Garbett A, Ghattas H
Integrating Health Technologies in Health Services for Syrian Refugees in Lebanon: Qualitative Study
J Med Internet Res 2020;22(7):e14283
URL: https://www.jmir.org/2020/7/e14283
doi:10.2196/14283
PMID:32628121

©Reem Talhouk, Chaza Akik, Vera Araujo-Soares, Balsam Ahmad, Sandra Mesmar, Patrick Olivier, Madeline Balaam, Kyle Montague, Andrew Garbett, Hala Ghattas. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Internet-Based Sharing Nurse Program and Nurses' Perceptions in China: Cross-Sectional Survey

Rendong Huang[1], BSc; Mei Xu[2], BSc; Xiuting Li[3], MSc; Yinping Wang[4], BSc; Bin Wang[1]; Naixue Cui[1], PhD

[1]School of Nursing, Cheeloo College of Medicine, Shandong University, Jinan, China

[2]Jinan Shizhong People's Hospital, Jinan, China

[3]Jinan Central Hospital, Jinan, China

[4]Peking University Health Science Center, Beijing, China

Corresponding Author:
Naixue Cui, PhD
School of Nursing
Cheeloo College of Medicine
Shandong University
No 44 West Wenhua Road
Jinan, 250012
China
Phone: 86 531 88382268
Email: cnx@sdu.edu.cn

Abstract

Background: China is currently piloting a "Sharing Nurse" program that aims to increase the accessibility of nursing services to at-home patients by enabling patients to order nursing services using mobile apps or online platforms.

Objective: This study aims to assess nurses' perceptions of the Sharing Nurse program, including their acceptance, concerns, needs, and willingness to take part in the program.

Methods: A total of 694 nurses participated in the questionnaire survey. The survey collected their sociodemographic and work-related information and their perceptions of the Sharing Nurse program using a self-developed questionnaire.

Results: The 694 respondents agreed that the Sharing Nurse program could provide patients with better access to nursing care (n=483, 69.6%). Their main concerns about the program were unclear liability division when medical disputes occur (n=637, 90.3%) and potential personal safety issues (n=604, 87%). They reported that insurance (n=611, 88%), permits from their affiliated hospital (n=562, 81.0%), clear instructions concerning rights and duties (n=580, 83.6%), real time positioning while delivering the service (n=567, 81.7%), and one-key alarm equipment (n=590, 85.0%) were necessary for better implementation of the program. More than half of the respondents (n=416, 60%) had an optimistic attitude toward the development of the Sharing Nurse program in China. However, only 19.4% (n=135) of the respondents expressed their willingness to be a "shared nurse." Further analyses found that nurses with a master's degree or above (χ^2_3=28.835, P<.001) or from tertiary hospitals (χ^2_3=18.669, P<.001) were more likely to be aware of the Sharing Nurse program and that male nurses were more willing to be shared nurses (Z=–2.275, P=.02).

Conclusions: The Chinese Sharing Nurse program is still in its infancy and many refinements are needed before it can be implemented nationwide. Generally, Chinese nurses are positive about the Sharing Nurse program and are willing to participate if the program is thoroughly regulated and supervised.

(J Med Internet Res 2020;22(7):e16644) doi:10.2196/16644

KEYWORDS

sharing nurse; home visiting; internet plus nursing program; perception; China

Introduction

Home visiting is a strategy to deliver various nursing or medical services to families or individuals within their home environment for an extended period of time [1,2]. Nurse home visits are employed in a number of settings, from managing acute or chronic illnesses to aiding the overall health and well-being of the patient. Home visits are also used to ease

XSL•FO
RenderX

adults into outpatient settings from inpatient settings. Home visits play an important role in enhancing health and independency, as well as providing high quality end-of-life care to older adults [3,4].

Home visiting nurses, in the provision of their services, have the opportunity to tailor the services to meet a family's unique needs more effectively [5,6]. Home visiting programs also allow qualified professionals to build a rapport with families, which may not be possible during other types of interventions [7,8]. Home visiting has also been shown to lead to many positive health outcomes. These include improved daily functioning among adults who are disabled, which may reduce Medicaid medical costs [9]; increase quality of life and increased number of days free of asthma symptoms for patients with asthma [10]; reduce child abuse and neglect; and improve child development and parenting outcomes [11-14].

Home visiting has been widely practiced in the United Kingdom, the United States, and many other developed countries. In the United Kingdom, home visiting programs have been expanded since 2005, delivering services to 16,000 disadvantaged new parents each year [15]. In the United States, about US $1.5 billion had been proposed by President Barack Obama's 2014 Budget to maintain and expand evidence-based voluntary home visiting services [16]. In Germany, since 2012, the federal and local governments have spent €102 million (US $115 million) each year to expand home visiting programs [17].

Nurse home visiting services in China are generally provided by community health care nurses, which began in the late 1990s [18]. Community health care involves a wide range of services such as disease prevention, rehabilitation, health education, and family planning guidance [19]. However, it is seen that community health care nurses remain at community health centers or stations, waiting passively for patients to arrive [18], instead of going into homes to provide health care, even though home visiting is one of their responsibilities. As a result, the traditional community health care service cannot meet the rising needs of at-home-services propelled by the large and continuous increase in the ageing population [20]. In response to this situation, the online nursing service platform (or "Sharing Nurse" platform)—where patients can request registered nurses to visit them at home and provide nursing services—emerged in China.

The Sharing Nurse platform originated as early as 2016 in the Shandong Province and has since spread to other provinces as well as first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen [21]. However, it did not receive official recognition until 2019. On February 12, 2019, the National Health Commission announced that they were launching a pilot Sharing Nurse program as part of the "Internet Plus" initiative for nursing care in six provinces and metropolitan cities (ie, Beijing, Tianjin, Shanghai, Jiangsu, Zhejiang, and Guangdong), as well as some capital cities and regions across the country [22].

The Sharing Nurse Program refers to the use of internet technology (ie, mobile apps, online websites, or other internet-based platforms) to facilitate home visiting nursing service delivery. To our knowledge, there are currently more than 20 Sharing Nurse apps in China. These apps or websites collect and verify the information of registered nurses who want to be part of the program ("shared nurses"). Clients register on the apps or websites to make an online appointment to reserve or order medical services, choose nurses to deliver services based on their preferences, and decide when and where the services are expected to be delivered. The appointed nurse then arrives at the scheduled time with the required medication and equipment to provide the ordered services [23]. After the service is completed, clients can evaluate the nurse's service and provide feedback on the platform. Currently, services provided by shared nurses include intramuscular injection, intravenous injection, collecting blood samples, and special care like newborn examinations or urethral catheterization, among others [24]. Unlike the home visiting nurses in many developed countries who are employed by agencies specializing in home visiting services, the Chinese shared nurses, in addition to delivering home visiting services through the sharing nurses platforms, are registered nurses affiliated with approved medical institutions (hospitals, community health care centers, and private clinics), and routinely serve in the medical institutions in which they are employed.

The Sharing Nurse program enables clients to actively use health care services without going to the community care centers or stations. As a new health care delivery mode, the program's quality and effectiveness is heavily dependent on both nurses and clients. Nurses' perceptions, such as acceptance of the program, as well as concerns, needs, and perceived benefits of being a shared nurse can greatly influence their willingness and motivation to participate in the program and to provide high-quality care. Eventually the success of the program also depends on how the government addresses the nurses' concerns and needs. However, research regarding nurses' perceptions (ie, acceptance, concerns, needs, and perceived benefits) of the Sharing Nurse program is still scarce. To address this gap, we conducted a questionnaire survey to assess nurses' perceptions of the program, as well as their willingness to be a shared nurse.

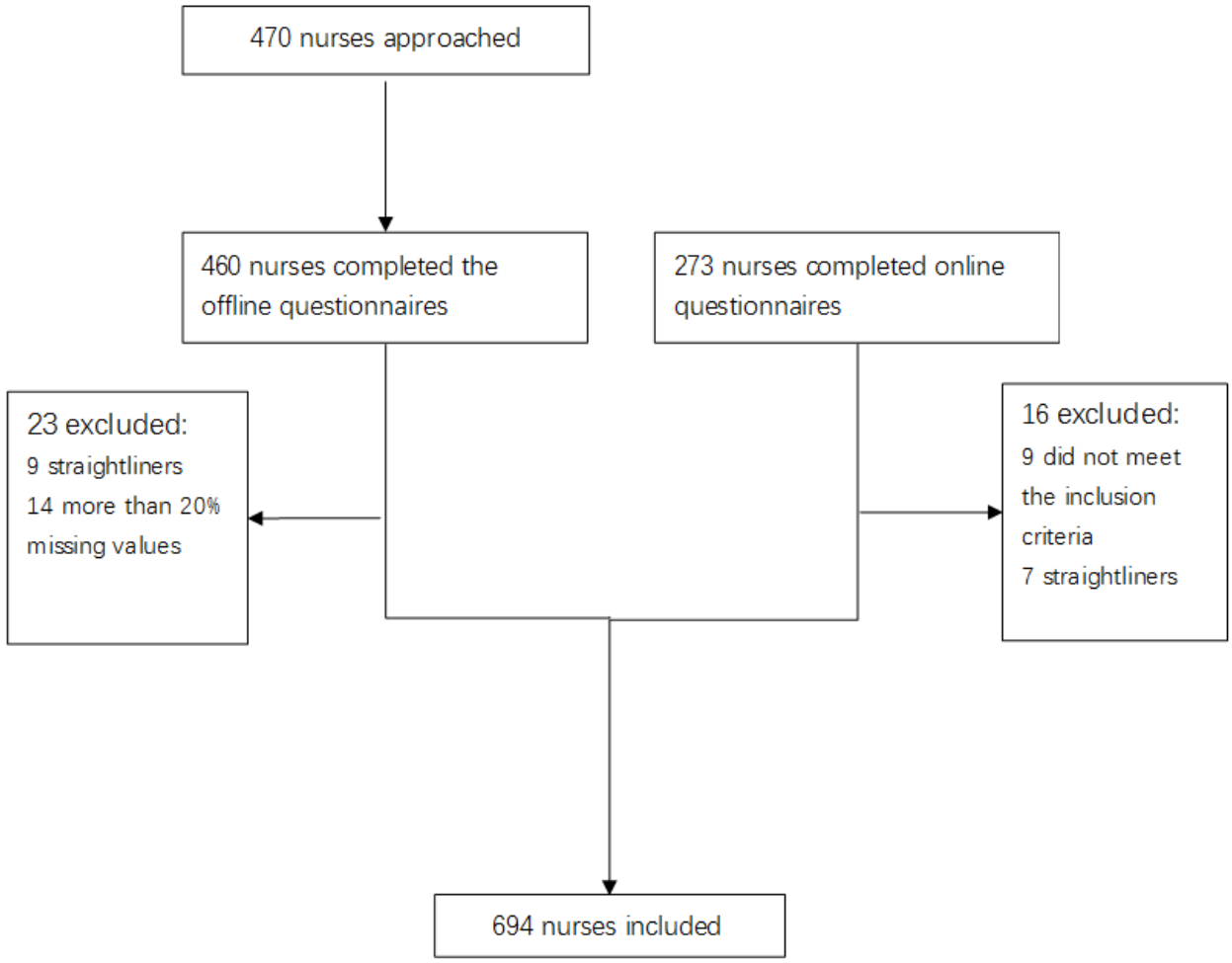
Methods

Design and Participants

A cross-sectional survey was conducted, both online and offline, in Jinan, China, from March to May 2019. A convenience sampling method was used to recruit registered nurses from two tertiary hospitals, one secondary hospital, seven community health centers, and four private hospitals in the capital city of Shandong Province, where the Sharing Nurse platform originated. The eligibility criterion for the nurses' participation in this study was more than 1 year of bedside nursing experience. A final sample of 694 nurses completed the questionnaires. The method of inducting participants is illustrated in the flow chart (Figure 1).

XSL•FO
RenderX

Figure 1. Flow diagram of the participants recruitment and questionnaires screening.



Measurements

Sociodemographic Information

The sociodemographic items were age, gender, professional title (nurse, senior nurse, nurse-in-charge, associate director, or director of nurses), years of work experience, educational background, and level of the affiliated hospital (tertiary hospital, secondary hospital, community hospital, or private hospital).

Nurses' Perceptions of the Sharing Nurse Program

We developed a questionnaire to assess nurses' perceptions of the Sharing Nurse program following these steps. First, we created an initial version of the questionnaire by developing items based on existing literature and interviews with a panel of 20 clinical nurses. Second, an expert panel comprising of two health care researchers, a senior nurse, and a community nurse refined these items. Third, five nurses took a pretest of the questionnaire and modifications were made according to their feedback to obtain the final version of the questionnaire.

The final version of the questionnaire contained five questions concerning nurses' familiarity with the Sharing Nurse program, whether they have ever been a shared nurse before, and reasons why they do not try to become a "shared nurse"; seven questions about their concerns about the program, rated on a 5-point Likert scale (1="strongly unconcerned" to 5="strongly concerned"); seven questions about the possible needs of shared nurses, rated on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"); four questions about the perceived benefits of the program, rated on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"), and one question about their willingness to be a shared nurse (response options including "willing to," "not willing to until regulations are refined," and "not willing to even though regulations refined").

Data Collection

First, hospitals of different levels were selected to enable us to explore possible variations in nurses' perceptions of the program across all hospital levels. At each chosen hospital, researchers distributed the questionnaires to nurses on site. To reach out to more nurses, such as those who were not on duty, we also

XSL•FO
RenderX

distributed electronic questionnaires (e-questionnaires) in the WeChat work group of the hospitals by using an online questionnaire platform Wenjuanxing. In the e-questionnaire, we added an extra question: "Have you ever attempted this questionnaire before?" at the end of the questionnaire to filter duplicate respondents. The willingness to participate, and thus provide informed consent, was determined at the receipt of the completed questionnaire. This study was approved by the Ethical Committees of Shandong University School of Nursing (Reference Numbers: 2019-R-009).

Data Analysis

Descriptive statistics were used to describe the respondents' sociodemographic and work-related information, as well as their responses to the questions regarding their perceptions of the Sharing Nurse program. A comparison analysis was conducted to test whether the perception differed by age, gender, educational background, and hospital levels using a chi-square test, Kruskal-Wallis tests, or the Mann-Whitney U test. For all tests, the significance level was set at $P \leq .05$. Data analysis was performed using SPSS 22.0 software (IBM Corp).

Results

Demographic and Work-Related Characteristics of the Participants

Of the 694 participants, 672 were female and 22 were male. The age of the participants ranged from 20 to 55 (mean 30.5, SD 6.8) years. The participants had been practicing clinical nursing for an average of 8.5 (SD 7.1) years. The highest percentage of the total 694 nurses (n=313, 45.1%) were from tertiary hospitals, 207 (29.9%) were from secondary hospitals, 81 (11.7%) were from community health care centers, and 114 (16.4%) were from private hospitals. Regarding their educational background, 210 (30.2%) held an associate's degree, 449 (64.7%) held a bachelor's degree, and 13 (1.9%) held a masters' degree or above. Concerning the distribution of nurses according to their different titles, 260 (37.5%) were nurses (junior title for the nursing profession), 275 (39.6%) were senior nurses, 143 (20.6%) were nurses-in-charge, and 7 (1.0%) were either associate directors or directors of nurses.

Perceptions of the Sharing Nurse Program

Close to half of the participants (300/694, 43.2%) were not aware of the Sharing Nurse program. Considering the nurses who knew about the program, only 6 had actually delivered services through a "Sharing Nurse" app. According to their answers, the factors that kept nurses from being a shared nurse included concern about personal safety (128/688, 18.6%), unclear regulations (126/688, 18.4%), lack of time (289/688, 42%), and dearth of energy (275/688, 40%). The concerns of the nurses are presented in Table 1.

Table 1. Concerns about home visiting as a shared nurse (N=694).

Concerns	Strongly unconcerned, n (%)	Partially unconcerned, n (%)	Neither concerned nor unconcerned, n (%)	Partially concerned, n (%)	Strongly concerned, n (%)	Missing, n (%)
Personal safety	16 (2.3)	9 (1.3)	57 (8.2)	82 (11.8)	522 (75.2)	8 (1.2)
Information security	26 (3.7)	38 (5.5)	97 (14)	143 (20.6)	383 (55.2)	7 (1)
Service quality supervision	18 (2.6)	25 (3.6)	120 (17.3)	164 (23.6)	357 (51.4)	10 (1.4)
Public recognition about shared nurse	33 (4.8)	35 (5.0)	128 (18.4)	145 (20.9)	345 (49.7)	8 (1.2)
Division of liability in medical disputes	11 (1.6)	8 (1.2)	41 (5.9)	59 (8.5)	568 (81.8)	7 (1.0)
Medical waste management	56 (8.1)	63 (9.1)	147 (21.2)	120 (17.3)	301 (43.4)	7 (1.0)
Service quality at home	18 (2.6)	25 (3.6)	120 (17.3)	164 (23.6)	357 (51.4)	10 (1.4)

Regarding the services that the participants think a shared nurse can deliver at home, the top three choices were blood collection for laboratory tests (489/694, 70.6%), intramuscular injection (473/694, 68.1%), and health education (508/694, 73.3%). The majority of the participants thought that at least 5 years of experience in bedside practice and the title of senior nurse were necessary qualifications to be a shared nurse.

Participants stated that they would consider being a shared nurse if they could obtain customized insurance (611/694, 88%), program training (488/694, 70.3%), clear statements on shared nurses' rights and duties (580/694, 83.6%), real time position tracking services while delivering service (567/694, 81.7%), and alarm equipment (590/694, 85.0%). A few nurses also regarded video recording of the service procedure as essential. Table 2 presents the full details of these answers.

XSL•FO
RenderX

Table 2. Needs to meet to consider to be a shared nurse of participants (N=694).

Needs	Strongly disagree, n (%)	Partially disagree, n (%)	Neither agree nor disagree, n (%)	Partially agree, n (%)	Strongly agree, n (%)	Missing, n (%)
Permits from the hospital	12 (1.7)	9 (1.3)	50 (7.2)	49 (7.1)	562 (81.0)	12 (1.7)
Insurance	12 (1.7)	4 (0.6)	22 (3.2)	34 (4.9)	611 (88.0)	11 (1.6)
Video recording of the service procedure	30 (4.3)	30 (4.3)	90 (13.0)	104 (15.0)	427 (61.5)	13 (1.9)
Clear statements about shared nurses' rights and duties	9 (1.3)	6 (0.9)	28 (4.0)	58 (8.4)	580 (83.6)	13 (1.9)
Real time position tracking while delivering service	14 (2.0)	8 (1.2)	37 (5.3)	57 (8.2)	567 (81.7)	11 (1.6)
Alarm equipment	10 (1.4)	3 (0.4)	26 (3.7)	54 (7.8)	590 (85.0)	11 (1.6)
Program training	11 (1.6)	12 (1.7)	69 (9.9)	101 (14.6)	488 (70.3)	13 (1.9)

More than half of the 694 participating nurses regarded the program as beneficial in terms of increasing public access to nursing care (n=483, 69.6%), extra income for nurses (n=402, 57.9%), preventing nurse turnover (n=371, 53.3%), and saving on health care costs (n=375, 54.0%). More than half (n=416, 60.0%) of the participants were optimistic about the program and were willing to be a shared nurse if there were clear and thorough regulations and supervision (Table 3).

Table 3. Perceived benefits toward sharing nurse program in China of participants (n=694).

Benefits	Strongly disagree, n (%)	Partially disagree, n (%)	Neither agree nor disagree, n (%)	Partially agree, n (%)	Strongly agree, n (%)	Missing, n (%)
More access to nursing care	40 (5.8)	29 (4.2)	129 (18.6)	140 (20.2)	343 (49.4)	13 (1.9)
Extra source of income for nurse	58 (8.4)	45 (6.5)	175 (25.2)	145 (20.9)	257 (37.0)	14 (2.0)
Preventing nurse turnover	82 (11.8)	66 (9.5)	160 (23.1)	119 (17.1)	252 (36.2)	15 (2.2)
Saving health costs for patients	66 (9.5)	57 (8.2)	182 (26.2)	129 (18.6)	246 (35.4)	14 (2.0)

Variations in Perceptions of the Sharing Nurse Program

It was seen that older nurses (χ^2_3=32.926, P<.001), nurses with higher levels of education (χ^2_3=28.835, P<.001), and nurses from tertiary hospitals were more likely to be aware of the program. Male nurses were more willing to be shared nurses than female nurses (Z=–2.275, P=.02; see Tables 4 and 5).

XSL•FO
RenderX

Table 4. Nurses who were aware of the sharing nurse platform.

Groups	Participants, n/N (%)	Chi-square (*df*)	*P* value	Post hoc
Age (years)		32.926 (3)	<.001	C>B, B>A
A <25	47/120 (39.2)			
B 25-34	235/423 (55.6)			
C 35-44	81/108 (75.0)			
D ≥45	30/43 (69.8)			
Gender		3.913 (1)	.048	N/A[a]
Male	17/22 (77.3)			
Female	376/672 (56.0)			
Education background		28.835 (3)	<.001	D>C, C>B, D>A
A Diploma degree	8/22 (36.4)			
B Associate's degree	93/210 (45.1)			
C Bachelor's degree	271/449 (63.1)			
D Master's degree or above	13/13 (100.0)			
Levels of working hospital		18.669 (3)	<.001	A>B, A>D
A Tertiary hospital	205/313 (65.5)			
B Secondary hospital	87/181 (48.1)			
C Community health center	42/81 (51.9)			
D Private hospital	57/114 (50.0)			

[a]N/A: not applicable.

XSL•FO
RenderX

Table 5. Attitude toward the perspective of the sharing nursing program.

Groups	Likert scale[a], median (IQR)	Kruskal-Wallis test	*Z* score	*P* value
Age (years)		0.008	N/A[b]	>.99
<25	2 (2-3)			
25-34	2 (2-3)			
35-44	2 (2-3)			
≥45	2 (2-3)			
Gender		N/A	–0.140	.89
Male	2 (2-3.25)			
Female	2 (2-3)			
Education background		0.824	N/A	.84
Diploma degree	3 (1-3)			
Associate degree	2 (2-3)			
Bachelor degree	2 (2-3)			
Master degree or above	2 (2-3)			
Levels of working hospital		0.095	N/A	.99
Tertiary hospital	2 (2-3)			
Secondary hospital	2 (2-3)			
Community health center	2 (2-3)			
Private hospital	2 (2-3)			

[a]The item was rated on a 5-point Likert scale (1: extremely optimistic, 2: optimistic, 3: neither optimistic nor pessimistic, 4: pessimistic, 5: extremely pessimistic).

[b]N/A: not applicable.

Discussion

To the best of our knowledge, this is the first study conducted to assess nurses' perceptions of the Sharing Nurse program in China. Nurses were specifically targeted because they are the executive body of the Sharing Nurse program; therefore, their perceptions and insights are of vital importance to policy makers and for the effective implementation of the program.

We found that almost half (300/694, 43.2%) of the participants did not know about the ongoing Sharing Nurse program. This may be because the survey was conducted less than 1 month after the program was officially piloted, and therefore, there was limited time for the nurses to learn about it. It may also be because the official announcement had not been disseminated widely and efficiently, or the employer hospitals or community health care centers had not made attempts to take part in the program.

Home visiting should be a part of community-based family support services [1]. In fact, the health care system in China emphasizes the central role of community health care nurses in primary care and home visiting services [25,26]. However, the results showed that nurses from the community health centers were less likely to be aware of the Sharing Nurse program. Cogdill [27] found that nurses with a master's degree were more perceptive to information needs more frequently than their colleagues. The low rate of community health care nurses who were aware of the Sharing Nurse program indicates, on the one hand, that community nurses, most of whom hold a diploma or associate's degree, lack information. On the other hand, this also reflects a flaw in policy promotion among community health care centers, which should be the frontier of the Sharing Nurse program implementation.

The main reasons for the respondents not willing to be a shared nurse were work stresses, as 40% (275/688) of nurses reported they did not have time or energy after work. This is consistent with previous study findings in England and Australia, where nurses have reported that administrative work not related to the practical task of delivering home visits was a significant barrier to home visiting work [28]. In the study city, shared nurses affiliated to approved medical institutions have to sacrifice their own time to deliver home visiting services through Sharing Nurse platforms, which adds more stress to their already high level of workload [29]. China also faces a shortage of nurses, and the growth in the number of health care professionals does not meet the increasing demand [30]. The Sharing Nurse program is expected to reduce this imbalance in a more efficient way. Therefore, the number of nurses should be expanded to attract more individuals to the nursing profession and better motivate nurses to engage in the program.

Safety issues are also a major concern about the program that the respondents shared. Studies in other countries have found that clients who received home visiting services without prior knowledge were inclined to report concerns about their safety

XSL•FO
RenderX

because they did not seek help but had been identified by others as potential beneficiaries of home visiting programs [31,32]. This can result in clients feeling vulnerable and powerless when they allow service providers into their homes [33]. Although we did not find any literature on clients' safety concerns related to the Sharing Nurse program, we found that more than two-thirds of nurses were concerned about their personal safety. Thus, the safety of visiting homes is a concern not only for the clients but also for the nurses. Hence, to better implement the Sharing Nurse program, efforts should be made to ensure the nurses' personal safety. There are many suggested measures to enhance home visiting safety. For example, the Sharing Nurse platforms verify the accuracy of information the clients upload, such as location, service requirements, and identity. The government may carry out strict inspections of the platforms. Necessary protective equipment such as real time positioning and alarm equipment could be provided to the shared nurses while delivering the home visiting service.

Liability division, when medical disputes occur while delivering a home visiting service through a Sharing Nurse platform, was found to be another major concern for the participating nurses. This concern relates to problems such as defining and dealing with medical disputes related to the services at home and whether accidents that happen to nurses when they are on the way to or back from the clients' home are workplace accidents or not. This demonstrates that there is a need for clear regulations and practical guidelines. This is supported by the finding that over 80% (580/694) of the respondents agreed that "clear statements about shared nurses' rights and duties" is necessary to the consideration of being a shared nurse. Previous findings also showed that the absence of related policies and practical guidelines was one of the major organizational barriers for Korean nurses to implement home visiting services [34].

Our study found that insurance is regarded by nurses to be essential to improve the program. Home visiting nurses face unique and complex risks in their workplaces as well as uncertainties on their way to and from clients' homes. These nurses need insurance support including traffic accident insurance and personal accident insurance. Previous studies have also pointed out that a lack of professional indemnity insurance is an obstacle both for the employment of nurses and for the fulfilling of their professional role [35]. Additionally, during a labor shortage, employment-based benefits such as insurance can be used to help recruit and retain nurses [36]. Therefore, we recommend the formulation of customized insurance options for shared nurses by policy makers and relevant agencies.

Well-planned and effective preservice and ongoing training contributes to home visitor readiness and prepares them for the demands of the job [37]. According to our participants, program training is also needed. We, therefore, suggest tailoring nursing education and training to include aspects of home visiting for the benefit of prospective home visiting nurses.

Additionally, more than half of the respondents agreed that the program could be beneficial in saving on health care costs for patients, providing extra income for nurses, and preventing nurse turnover. Considering patients' possible health care savings, the result is consistent with previous findings that home visiting could save costs such as parking fees, travel time, and time off work associated with travelling to a clinic [38]. It also cancels out the waiting time in hospital lines [24]. Shared nurses on the sharing nurse apps generally earn ¥100-300 (US $14-42) per visit for their services, which is more than what they could earn for their daily work [39]. Evidence [40,41] has shown that higher wages are associated with less turnover. The extra income, more flexibility, and the opportunities that the Sharing Nurse program offers make it more attractive to join the program.

The limitations of this study cannot be ignored. First, the participants in this study all came from one city in Shandong Province, which is not an official piloting area of the Sharing Nurse program. Nevertheless, Jinan is one of the earliest cities where the platform originated. Second, the convenience sampling method limits the generalizability of our findings. However, the sample of this study was purposefully selected to cover various levels of health care that are typical to China to explore differing perceptions among clinical nurses from different settings. Third, we only explored the perceptions of the Sharing Nurse program from nurses' perspective, not from clients' perspective. Future studies focusing on clients' perceptions and the actual effectiveness of the program are warranted.

China's Sharing Nurse program is a new way to facilitate home visiting services. However, in the early stage of the program, the awareness of the nurse program was low. The major concerns of the program include unclear liability division when medical disputes occur and potential personal safety issues. The participating nurses identified several needs that must be met for them to become a shared nurse, including insurance, permits from their employers, clear instructions concerning rights and duties, real time positioning while delivering the service, and one-key alarm equipment. Although the participating nurses noted many barriers of becoming a shared nurse, most of them held an optimistic view of the program and expressed their willingness to be part of it once regulations and practical guidelines are refined. The findings contribute to promoting knowledge about China's Sharing Nurse program and, hopefully, provide implications for further refinement and regulations of the program.

Acknowledgments

We would like to thank all the participants of the survey and the nurses who participated in the questionnaire refinement. We extend our gratitude to the Supporting Grant for Candidates for "Qilu Outstanding Scholars" (2018WLJH46) from the Shandong University and Zhenhui Wang for providing critical information about Sharing Nurse program.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Gomby DS, Culross PL, Behrman RE. Home visiting: recent program evaluations--analysis and recommendations. Future Child 1999;9(1):4-26, 195. [Medline: 10414008]
2. Gomby DS, Larson CS, Lewit EM, Behrman RE. Home visiting: analysis and recommendations. Future Child 1993;3(3):6. [doi: 10.2307/1602540]
3. Bouman A, van Rossum E, Nelemans P, Kempen GI, Knipschild P. Effects of intensive home visiting programs for older people with poor health status: a systematic review. BMC Health Serv Res 2008 Apr 03;8:74 [FREE Full text] [doi: 10.1186/1472-6963-8-74] [Medline: 18387184]
4. Nakanishi M, Niimura J, Nishida A. Factors associated with end-of-life by home-visit nursing-care providers in Japan. Geriatr Gerontol Int 2017 Jun;17(6):991-998. [doi: 10.1111/ggi.12822] [Medline: 27389916]
5. Wasik BH. Staffing issues for home visiting programs. Future Child 1993;3(3):140. [doi: 10.2307/1602547]
6. Weiss HB. Home visits: necessary but not sufficient. Future Child 1993;3(3):113. [doi: 10.2307/1602545]
7. Beer NI. The role of the home visiting nurse in the total education programme of spinal cord injured persons. Paraplegia 1984 Oct;22(5):311-315. [doi: 10.1038/sc.1984.50] [Medline: 6493799]
8. Kaplan L. Working with multiproblem families. In: Children and Youth Services Review. Lexington, MA: Lexington Books; Sep 1986:382-384.
9. Szanton SL, Alfonso YN, Leff B, Guralnik J, Wolff JL, Stockwell I, et al. Medicaid cost savings of a preventive home visit program for disabled older adults. J Am Geriatr Soc 2018 Mar;66(3):614-620. [doi: 10.1111/jgs.15143] [Medline: 29165789]
10. Mayor S. Home visits improve asthma control in low income adults, study shows. BMJ 2014 Nov 25;349:g7165. [doi: 10.1136/bmj.g7165] [Medline: 25428421]
11. Chaiyachati BH, Gaither JR, Hughes M, Foley-Schain K, Leventhal JM. Preventing child maltreatment: examination of an established statewide home-visiting program. Child Abuse Negl 2018 May;79:476-484 [FREE Full text] [doi: 10.1016/j.chiabu.2018.02.019] [Medline: 29558714]
12. Howard KS, Brooks-Gunn J. The role of home-visiting programs in preventing child abuse and neglect. Future Child 2009;19(2):119-146. [doi: 10.1353/foc.0.0032] [Medline: 19719025]
13. Sawyer MG, Frost L, Bowering K, Lynch J. Effectiveness of nurse home-visiting for disadvantaged families: results of a natural experiment. BMJ Open 2013;3(4):e002720 [FREE Full text] [doi: 10.1136/bmjopen-2013-002720] [Medline: 23619089]
14. Sweet MA, Appelbaum MI. Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. Child Dev 2004;75(5):1435-1456. [doi: 10.1111/j.1467-8624.2004.00750.x] [Medline: 15369524]
15. GOV.UK. Family Nurse Partnership Programme to Be Extended. Department of Health and Social Care. 2013. URL: https://www.gov.uk/government/news/family-nurse-partnership-programme-to-be-extended [accessed 2019-06-07]
16. HHS awards $386 million to support families through the home visiting program. U.S. Department of Health and Human Services. 2015. URL: http://www.hhs.gov/news/press/2015pres/02/20150219a.html [accessed 2019-07-08]
17. Bundeshaushaltsplan 2016, Einzelplan 17. Bundesministerium für Familie, Senioren, Frauen und Jugend. 2016. URL: https://www.bundeshaushalt.de/fileadmin/de.bundeshaushalt/content_de/dokumente/2016/soll/epl17.pdf [accessed 2019-07-08]
18. Yuan S, Peng F, Jiang X. Community health nursing in China: status, challenges, and development strategies. Nurs Outlook 2012;60(4):221-227. [doi: 10.1016/j.outlook.2012.03.002] [Medline: 22512991]
19. Ma W, Meng X, Wei Y, Li J. Roles and activities of community nurses in China: a descriptive study. J Nurs Manag 2019 Mar;27(2):433-441. [doi: 10.1111/jonm.12678] [Medline: 30443947]
20. Zhu H, Lu J, Zhang Y, Cui B. Responses to population ageing in the new era: a national condition report from China. China Popul Dev Stud 2018 Dec 4;2(3):272-283. [doi: 10.1007/s42379-018-0017-9]
21. Hu B. Shared nurse apps raise concerns in China's biggest cities. That's Mag. 2018 Jun 20. URL: https://www.thatsmags.com/shanghai/post/23810/shared-nurses-now-a-thing-in-china-s-biggest-cities [accessed 2019-07-08]
22. China pilots an internet plus nursing program. The State Council. 2019. URL: http://english.www.gov.cn/state_council/ministries/2019/02/13/content_281476520056360.htm [accessed 2019-08-10]
23. “Nurse sharing” applications bring new kind of service to China’s healthcare industry. Asia Perspective. 2018 Aug 02. URL: http://asiaperspective.net/2018/08/02/nurse-sharing-apps-new-services-healthcare-china/ [accessed 2019-08-12]
24. Hong'e M. Chinese 'nurse sharing' app to offer at-home care. CGTN. 2018. URL: https://www.cn-healthcare.com/article/20180530/content-503935.html [accessed 2019-08-13]
25. Chen Z. Launch of the health-care reform plan in China. Lancet 2009 Apr 18;373(9672):1322-1324. [doi: 10.1016/S0140-6736(09)60753-4] [Medline: 19376436]

XSL•FO
RenderX

26. Bin L. Deepening the reform of the medical and health care system. National Health and Family Planning Commission. 2016. URL: http://www.chinadaily.com.cn/m/chinahealth/2014-08/24/content_18476807_5.htm [accessed 2019-07-08]
27. Cogdill KW. Information needs and information seeking in primary care: a study of nurse practitioners. J Med Libr Assoc 2003 Apr;91(2):203-215 [FREE Full text] [doi: 10.1043/0025-7338(2003)091<0203:INAISI>2.0.CO;2] [Medline: 12883577]
28. Sawyer MG, Barnes J, Frost L, Jeffs D, Bowering K, Lynch J. Nurse perceptions of family home-visiting programmes in Australia and England. J Paediatr Child Health 2013 May;49(5):369-374. [doi: 10.1111/jpc.12197] [Medline: 23573991]
29. Yau SY, Xiao XY, Lee LYK, Tsang AYK, Wong SL, Wong KF. Job stress among nurses in China. Appl Nurs Res 2012 Feb;25(1):60-64. [doi: 10.1016/j.apnr.2011.07.001] [Medline: 21855294]
30. Wu Q, Zhao L, Ye X. Shortage of healthcare professionals in China. BMJ 2016 Sep 22;354:i4860. [doi: 10.1136/bmj.i4860] [Medline: 27659864]
31. Kitzman H, Cole R, Yoos H, Olds D. Challenges experienced by home visitors: a qualitative study of program implementation. J Community Psychol 1997 Jan;25(1):95-109. [doi: 10.1002/(sici)1520-6629(199701)25:1<95::aid-jcop7>3.0.co;2-1]
32. Stetler K, Silva C, Manning SE, Harvey EM, Posner E, Walmer B, et al. Lessons learned: implementation of pilot universal postpartum nurse home visiting program, Massachusetts 2013-2016. Matern Child Health J 2018 Jan;22(1):11-16. [doi: 10.1007/s10995-017-2385-x] [Medline: 29119476]
33. Jack SM, DiCenso A, Lohfeld L. A theory of maternal engagement with public health nurses and family visitors. J Adv Nurs 2005 Jan;49(2):182-190. [doi: 10.1111/j.1365-2648.2004.03278.x] [Medline: 15641951]
34. Park H, Park E, Park J. Barriers to chronic pain management in community-dwelling low-income older adults: home-visiting nurses' perspectives. Collegian 2016 Sep;23(3):257-264. [doi: 10.1016/j.colegn.2015.05.002]
35. Fisher JE. Mental health nurse practitioners in Australia: improving access to quality mental health care. Int J Ment Health Nurs 2005 Dec;14(4):222-229. [doi: 10.1111/j.1440-0979.2005.00375.x] [Medline: 16296989]
36. Spetz J, Adams S. How can employment-based benefits help the nurse shortage? Health Aff (Millwood) 2006;25(1):212-218. [doi: 10.1377/hlthaff.25.1.212] [Medline: 16403756]
37. Zeanah PD, Larrieu JA, Boris NW, Nagle GA. Nurse home visiting: perspectives from nurses. Infant Ment Health J 2006 Jan;27(1):41-54. [doi: 10.1002/imhj.20079] [Medline: 28640422]
38. Gehring ND, Ball GDC, Perez A, Holt NL, Neuman D, Spence N, et al. Families' perceived benefits of home visits for managing paediatric obesity outweigh the potential costs and barriers. Acta Paediatr 2018 Feb;107(2):315-321. [doi: 10.1111/apa.14101] [Medline: 28960483]
39. 'Nurse-sharing' apps offer some relief to China's aging population. Global Times. 2018 May 27. URL: http://www.globaltimes.cn/content/1104296.shtml [accessed 2019-08-10]
40. Zeytinoglu IU, Denton M, Davies S, Plenderleith JM. Casualized employment and turnover intention: home care workers in Ontario, Canada. Health Policy 2009 Aug;91(3):258-268. [doi: 10.1016/j.healthpol.2008.12.004] [Medline: 19178976]
41. Howes C. The impact of a large wage increase on the workforce stability of IHSS home care workers in San Francisco County. Labor Center, Berkeley. 2002. URL: http://laborcenter.berkeley.edu/pdf/2002/Howes.pdf [accessed 2019-10-10]

Abbreviations

e-questionnaires: electronic questionnaires

Edited by G Eysenbach; submitted 10.10.19; peer-reviewed by X Meng, Z Zhu; comments to author 07.06.20; accepted 25.06.20; published 22.07.20.

Please cite as:
Huang R, Xu M, Li X, Wang Y, Wang B, Cui N
Internet-Based Sharing Nurse Program and Nurses' Perceptions in China: Cross-Sectional Survey
J Med Internet Res 2020;22(7):e16644
URL: http://www.jmir.org/2020/7/e16644/
doi:10.2196/16644
PMID:32706711

©Rendong Huang, Mei Xu, Xiuting Li, Yinping Wang, Bin Wang, Naixue Cui. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Consequences of Gift Giving in Online Health Communities on Physician Service Quality: Empirical Text Mining Study

Li Peng[1]; Yanan Wang[2], MA; Jing Chen[3]

[1]Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[2]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[3]Department of Radiology, Central South University Xiangya School of Medicine Affiliated Haikou Hospital, Haikou, China

Corresponding Author:
Yanan Wang, MA
School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology
13 Hangkong Road
Qiaokou District
Wuhan, Hubei
China
Phone: 86 15972965520
Email: yanan_wang2020@163.com

Abstract

Background: Gift giving, which has been a heavily debated topic in health care for many years, is considered as a way of expressing gratitude and to be beneficial for the physician-patient relationship within a reasonable range. However, not much work has been done to examine the influence of gift giving on physicians' service quality, especially in the online health care environment.

Objective: This study addressed the consequences of gift giving by mining and analyzing the dynamic physician-patient interaction processes in an online health community. Specifically, gift types (affective or instrumental) based on the motivations and physician-patient tie strength were carefully considered to account for differences in physicians' service quality.

Methods: The dynamic interaction processes (involving 3154 gifts) between 267 physicians and 14,187 patients from a well-known online health community in China (haodf.com) were analyzed to obtain empirical results.

Results: Our results reveal that patient gift giving inspires physicians to improve their service quality as measured by physicians' more detailed responses and improved bedside manner, and the degree of influence varied according to the strength of the physician-patient tie. Moreover, affective gifts and instrumental gifts had different effects in improving physicians' service quality online.

Conclusions: This study is among the first to explore gift giving in online health communities providing both important theoretical and practical contributions. All of our results suggest that gift giving online is of great significance to promoting effective physician-patient communication and is conducive to the relief of physician-patient conflicts.

(J Med Internet Res 2020;22(7):e18569) doi:10.2196/18569

KEYWORDS

online health community; gift giving; affective/instrumental gifts; service quality; bedside manner; physicians; physician-patient relationship

Introduction

Background

Online health communities have recently emerged as an important channel for seeking medical information and physicians' help [1], improving the physician-patient relationship [2], and raising the level of public health [3]. The Pew Research Center reported that 59% of American adults had searched for health information online [4]. An increasing number of scholars have been paying specific attention to patient [5,6] or physician [7-9] behaviors in online health communities. However, little is known about how patient behaviors influence physician behaviors.

Patient behaviors online can be mainly divided into purchase and review behaviors. The former has already been investigated in a large number of studies. With respect to review behaviors,

XSL•FO
RenderX

positive behaviors were shown to account for almost 99% of all reviews (calculated based on data collected on Haodf.com [10]), which might be due to the unequal relationships between patients and physicians. Among the various ways of expressing gratitude, gift giving is popular among patients, although this behavior has been widely questioned by the public [11]. One cause of these debates is that the possible impact of gift giving on the service quality of physicians remains unclear. Both academia and industries have raised the question as to whether patients buy gifts for physicians in fear of receiving poor treatment rather than as a gesture of expressing gratitude [12]. To address these questions, we have been exploring patients' motivations of gift giving as well as the impact of gift giving on physician behavior (ie, service quality).

Within a reasonable rage, gift giving has benefits for emotional expression and relationship building in interpersonal communication [13] and can be considered to be a benefit for the physician-patient relationship. From the patient perspective, on the one hand, gifts can be used to express gratitude. On the other hand, gift giving helps to reassure patients that physicians will be more likely to fulfill their responsibilities after receiving gifts. From the physician perspective, on the one hand, gifts from patients make them feel respected and recognized for their efforts. On the other hand, gifts may inspire them to improve service quality.

Despite recognition of the many advantages of gift giving for both patients and physicians, these effects should also be empirically examined. To the best of our knowledge, only one study has investigated the relationships between gift giving and physicians' service delivery to date [14]. However, the influence according to gift types distinguished by motivations has not been given full consideration. Moreover, the influences of gift giving on physicians may demonstrate time effects, which have also not been analyzed. In the previous study, service quality was measured by the physicians' response speed to patients' question. Response speed, as a dimension of service quality, will ultimately influence patients' perceived service quality. However, we believe that the quality of the reply content to the question may be more important and highly valued by patients.

To extend existing studies, we systematically analyzed the motivations and effects of gift giving in online health care. First, gifts were classified into affective gifts and instrumental gifts based on different motives [15]. All gifts that physicians received were judged and categorized based on the complete interaction process (including gift giving) between physicians and patients. The different effects of affective and instrumental gifts on physicians' service quality were further examined. Second, the degree of intimacy of the physician-patient relationship (ie, the tie strength) may influence the effects of gift exchange [16] and was also considered in our study. Third, the time effects of gift giving were comprehensively analyzed to obtain conclusions that can offer more practical guidance to patients. Fourth, the level of detail and emotional support of physicians' replies were used to measure service quality. Therefore, the specific research questions in this study were as follows: (1) Does gift giving influence physicians' service quality in online health communities? (2) Are the effects of different types of gifts (affective and instrumental gifts) on physicians' service quality consistent? (3) How does the physician-patient tie strength moderate the relationships between gift giving and service quality? and (4) Does gift giving show time effects on service quality?

To address these questions, we adopted a quantitative approach and examined the effect of gift giving for the whole interaction process on physicians' service quality by analyzing longitudinal data gathered from a popular online health community in China, Haodf.com [10].

Gift Theory

Gift exchange plays a vital role in social interactions [17]. Gift theory, which involves the norm of reciprocity and explains why gift givers frequently receive return gifts, provides a suitable theoretical basis for this study.

The principle of reciprocity, which is considered as a benefit for individuals participating in social exchange, is often defined as a set of socially accepted trading rules in which one party provides resources to the other and obliges the other to return the favor [18]. The reciprocity principle is the internal cause of continuous communication between people since people will always repay each other owing to the nature of mutual indebtedness within the principle [19]. Moreover, referring to the partition manner of gifts in sociology, we divide gifts into affective and instrumental gifts based on different motives. Affective and instrumental gifts are defined as having "emotional expression" and "utilitarian purpose," respectively. However, both affective and instrumental elements, with different proportions, rarely exist in the process of gift exchange simultaneously. Hence, the gift type in our study was defined as the major element in a given exchange.

Professional Identity

Professional identity is defined as how one perceives the goals, social values, and other factors of their profession, and how this is communicated to others [20]. Relevant literature has paid more attention to the professional identity of physicians, nurses, and teachers; there are diverse definitions of professional identity in different fields [21,22], although this identity is of relevance for each worker. Only when one recognizes the occupation they are engaged in can they become more devoted to the work, exert their full potential, and realize their value in the course of the work. Studies concerning professional identity have revealed that high professional identity can be viewed as a unique way of shaping a good professional image, whereas the absence of such identity may drive one to leave the field [23]. Furthermore, professional identity can be enhanced through recognition and encouragement from others [24].

Tie Strength Theory

Following prior studies, tie strength is defined as the level of intimacy and interaction between individuals. Granovetter [25] conducted a comprehensive study to define and classify tie strength, which was categorized according to weak and strong ties defined by "unfamiliar interpersonal relations" and "familiar interpersonal relations," respectively. Thus, strong ties reflect closer relationships and more frequent interactions compared to weak ties [26]. Therefore, in the context of online health

XSL•FO
RenderX

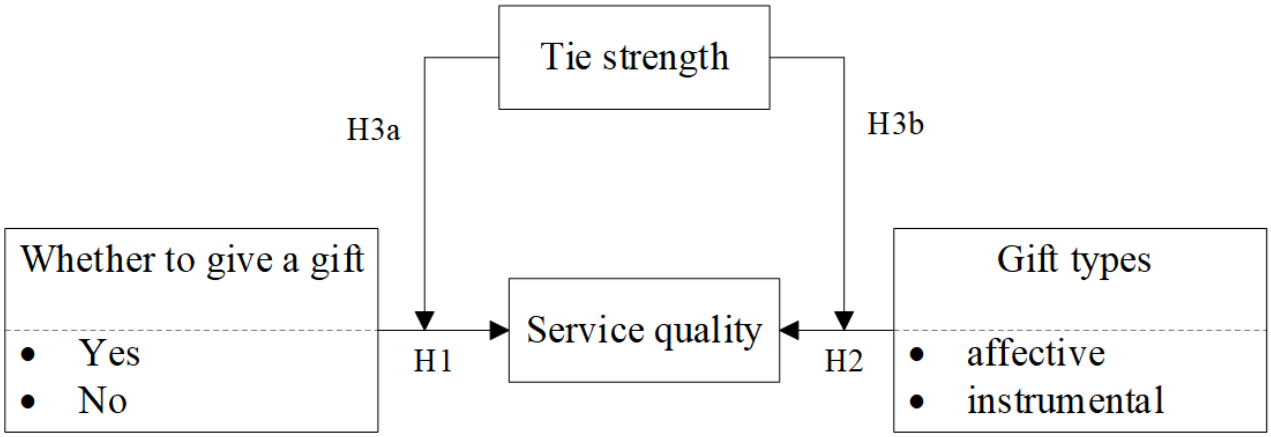
communities, tie strength can be considered to consist of two dimensions, intimacy and interaction, which are measured by the cumulative number of gifts and interaction times, respectively. Strong ties can provide material and reliable support, whereas weak ties, along with multiple information, cannot [27]. Hence, the strength of ties determines the quality of the information available and the likelihood that individuals will achieve their goals [28].

We therefore framed our study based on these perspectives to examine how tie strength influences the effect of giving gifts in online health communities. The judgment on gift types and the subsequent attitude about gift giving will both be influenced by the relationships between receivers and givers [29]. Specifically, gift giving is more likely to be accepted when the two sides are in a close relationship such as family, romantic partnership, friendship, and even geographical proximity. However, the receivers can easily suspect the motives of the gift, which will then be refused when the receivers and givers are strangers. Hence, we surmise that tie strength will not only moderate the impact of gift giving on physicians' service quality but will also influence the physicians' judgment of gift types.

Research Model and Hypotheses

Figure 1 shows the proposed research model based on the above theoretical background. To understand the relationships among several constructs, we proposed and tested several hypotheses.

Figure 1. Conceptual model.



Gift theory states that receivers will have the moral norm and obligation of returning the favor because of the reciprocity norm [30]. Online virtual gifts provide physicians with economic and honorary utility. Therefore, physicians always face the obligation and motivation to reciprocate patients after receiving a gift. Considering the online health care environment, we believe that physicians may improve their service quality to patients after receiving a gift, leading to the following hypothesis (H1): gifts help improve physicians' service quality in online health communities.

Previous studies based on the offline environment found that gift exchange offers a way to connect socially [31]. However, gifts provided at different times and different communication stages may have different effects. Even if we consider that all gifts are beneficial to the improvement of physicians' service quality, gifts with different purposes may have different effects. Kirchler and Palan [32] tested the value of unconditional nonmonetary gifts as a way to improve health worker performance in a low-income country health setting, and found that unconditional nonmonetary gifts enhanced performance by 20% over 6 weeks compared to conditional nonmonetary gifts.

Affective gifts are used to express the givers' appreciation, regardless of the possibility of return, whereas instrumental gifts are often given for a utilitarian purpose. Compared with instrumental gifts, affective gifts can provide not only economic benefits and improvement of online reputation [33] but also encouragement and recognition from patients, which can in turn enhance the professional identity of physicians and then inspire them to improve their service quality. Based on these perspectives, we developed our second hypothesis (H2) as follows: compared with instrumental gifts, affective gifts will be more effective in improving physicians' service quality.

Tie strength affects the quality of the information available and the likelihood that individuals will achieve their goals; a strong tie helps to obtain a gift that is more expensive and valuable [27]. When a patient and a physician know each other well, affectivity and compulsoriness will play major roles in their interaction process [26]. In such a situation, physicians always serve patients conscientiously. Therefore, a strong tie may weaken the effect of gift giving on physicians' service quality, leading us to establish the following hypothesis (H3a): tie strength negatively moderates the relationships between patients' gift giving and physicians' service quality.

The lower the degree of intimacy between givers and receivers, the more likely people regard gifts with utilitarian features [29]. Therefore, tie strength may affect the receiver's judgment on the motivations of gifts. Specifically, strong ties between physicians and patients decrease the feelings of physicians on the differences between affective and instrumental characteristics. A strong tie between a physician and patient will mitigate the effects of gift types on the physician's behavior. Based on these perspectives, we developed the following hypothesis (H3b): tie strength mitigates the difference of the effect of gift type on physicians' service quality.

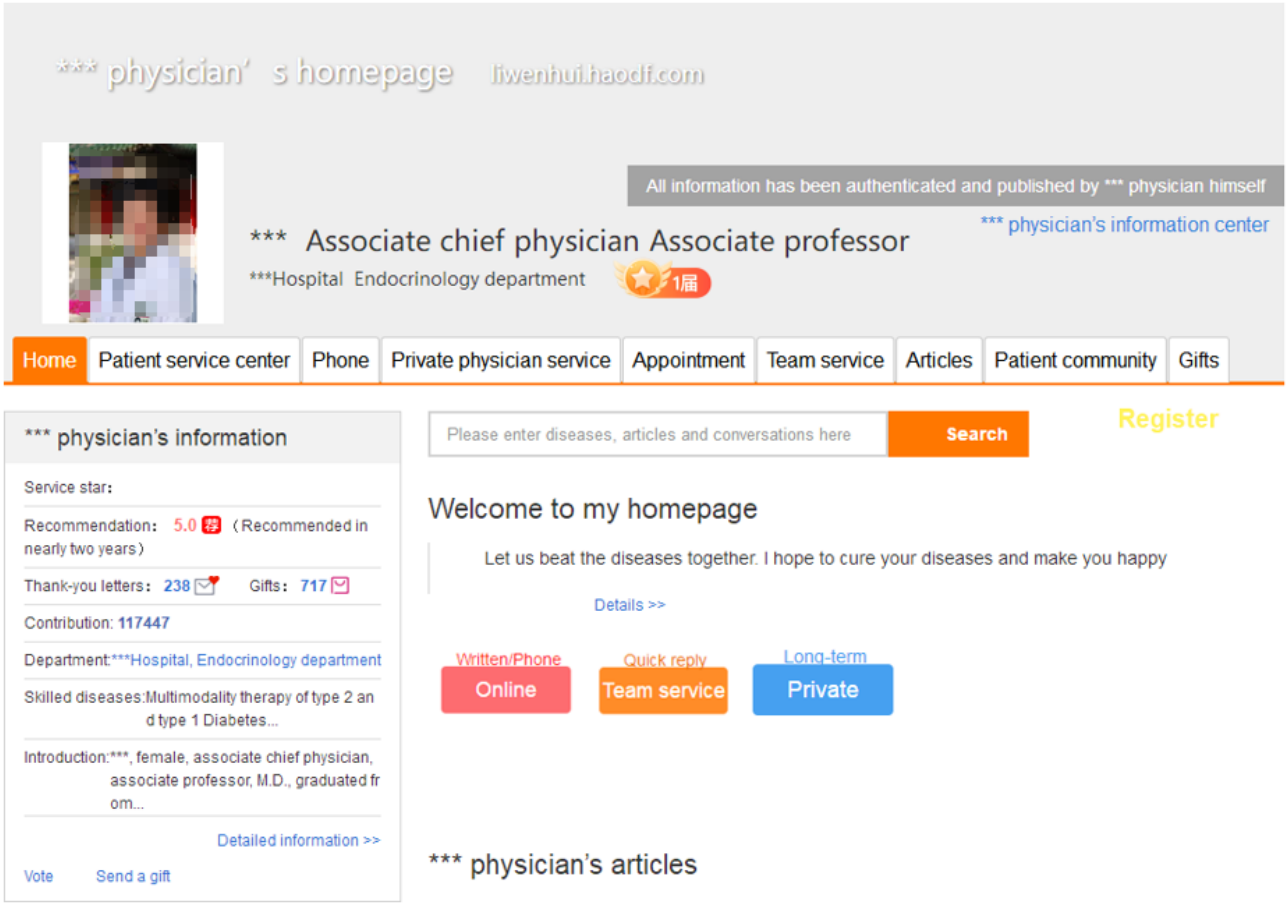
XSL•FO
RenderX

Methods

Research Context

To test the research hypotheses, we collected data from one of the most popular online health communities, Haodf.com [10], which was established in 2006 in China. More than 490,000 physicians from 7500 hospitals are currently working on the site. Haodf.com creates a homepage for each physician (see Figure 2; all screenshots of web pages were translated to English for clarity). For each physician, the entire interaction process with all patients is archived on the site. The interaction process includes the communicator (physician or patient), time, and content (physician-patient dialog or patient gift giving) for each interaction. An example of the interaction process is shown in Figure 3. Data collection was facilitated since most of the physician-patient interaction processes on Haodf.com are public, except for a few processes that are encrypted for privacy protection.

Figure 2. Physician's homepage.



XSL•FO
RenderX

Figure 3. Physician-patient interaction process.



Sample and Data Collection

To reduce the influence of disease types, we only included physicians who treat patients with diabetes as our sample. We developed a crawler to automatically download information, including physicians' information and each physician-patient interaction content, on Haodf.com. We collected the data on March 25, 2018, and the entire process was conducted for 1 week. After downloading, cleaning up, and matching information processes, 217,458 interactions (X_1) between 14,187 patients and 267 physicians were included in our model.

The data cleaning and processing (see Figure 4) were broken down into the following steps: (1) cleaning missing data, in which the records for patients' purchase behaviors (see example in Figure 5) and physicians' replies by voice messages (see example in Figure 6) were deleted, resulting in 212,303 records (X_2) retained; (2) records integration (X_3) from each physician-patient interaction process per day to ensure communication continuity; and (3) gift giving judgment, in which each record in X_3 was divided into two records based on the first gift giving behavior, namely prior to and after gift giving. Finally, 28,546 records (X_4) were included in our model.

XSL•FO
RenderX

Figure 4. Data cleaning and processing.

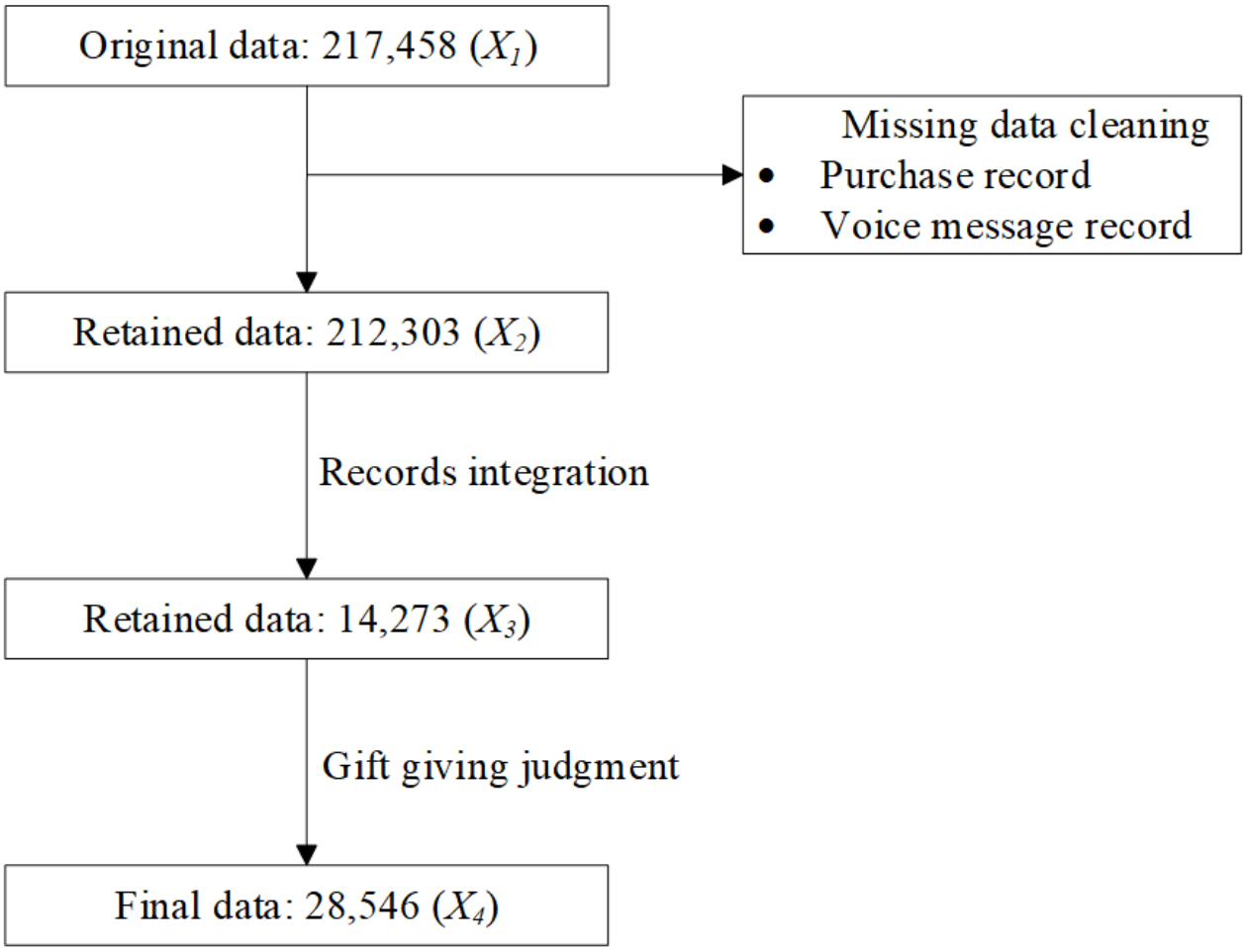
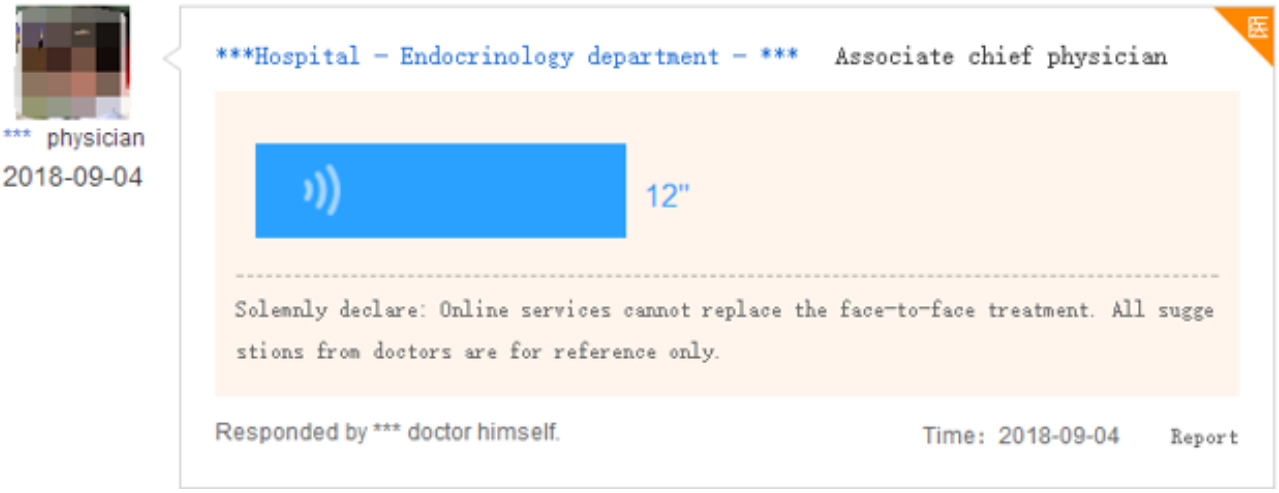


Figure 5. Patient's purchase process.



XSL•FO
RenderX

Figure 6. Physician's reply by voice message.



Judgment of Gift Types and Tie Strength

Judgment of gift types and tie strength was based on text mining.

Manual Coding for Gift Types

Gift types were judged according to whether gift giving occurred in conjunction with patients' questions. An instrumental gift is one that is purposefully given to physicians (eg, with the goal of obtaining better service). Therefore, gifts that are given close to a patient's questions were considered to be instrumental gifts (see Figure 7 for an example); otherwise, the gift was considered to be affective (see Figure 8 for an example). With respect to the judgment of whether the dialog involved a question, two assistants with a research background in medical informatics were trained to recognize patients' questions based on the keywords shown in Multimedia Appendix 1. One hundred whole interaction processes, including 300 dialogs between patients and physicians, were chosen at random and assigned to be judged. The two assistants coded the contents of the 300 dialogs independently, and consistent judgment was obtained for 296 (98.7%) dialogs. After analyzing the remaining ambiguous dialogs, consistent judgments were reached. Finally, one assistant was assigned to code the remaining contents.

Figure 7. Instrumental gift example.



XSL•FO
RenderX

Figure 8. Affective gift example.



Measurement of Tie Strength

Tie strength was measured based on the cumulative quantity of gifts and conversations. A conversation includes a question and a reply. Therefore, the tie strength will increase as the number of physician replies increases. We standardized the cumulative quantity of gifts and conversations, respectively, and tie strength was calculated according to the following formula: standardized number of gifts + standardized number of physician replies.

Variables and Empirical Models

The empirical variables included in our model are shown in Table 1. As the dependent variable, we used the ratio of average word count for a physician to respond to the patient in dialog *i* as a measure of service quality. More words contain more information. Specifically, a long reply may cover solutions for the question from all aspects and offer a detailed explanation of each aspect. In addition, a long reply may reflect the physician's patience and serious attitude to their patients. We also considered that a long question may lead to a long reply. Therefore, we also controlled for the word count from the patient's question. The formula is as follows: word count ratio_i=ln(average word count for physician dialog *i*)/ln(average word count for patient in dialog *i*).

XSL•FO
RenderX

Table 1. Variables description.

Variable	Explanation
Ratio_WordCount$_{ij}$ (Dependent variable)	The ratio of average word counts for physician *j* to that of the patient in dialog *i*
Independent variables	
GiftGiving $_{ij}$	Dummy variable in which "1" represents that a gift has been sent and "0" represents no gift
GiftType $_{ij}$	Dummy variable in which "1" represents an instrumental gift and "0" represents an affective gift
TieStrength $_{ij}$	Summation of standardized interaction times and the standardized cumulative number of gifts
Control variables	
Phone $_{j}$	Dummy variable in which "1" represents that the physician provides phone consultation and "0" represents no phone consultation
Written $_{j}$	Dummy variable in which "1" represents that the physician provides written consultation and "0" represents no written consultation
Outpatient $_{j}$	Dummy variable in which "1" represents that the physician provides outpatient service appointments and "0" represents no appointment provision
PhonePrice $_{j}$	Service price for phone consultation service set by the physician
WrittenPrice $_{j}$	Service price for written consultation service set by the physician
Thank-you Letter $_{j}$	Number of thank-you letters that physician *j* received
No. Gifts $_{j}$	Number of gifts that physician *j* received
Contribution $_{j}$	Calculated by site to measure the effort of physician *j*
No. Patients $_{j}$	Number of patients that physician *j* consulted
Recommendation $_{j}$	Total number of votes for each physician from patients
Title $_{j}$	Dummy variable in which "1" represents that physician *j* is a chief physician and "0" otherwise
City $_{j}$	Economics of the city where the hospital is located to measure patients' consumption capacity. Expressed as a dummy variable in which "1" represents that the physician works in a first-tier city and "0" otherwise
Level $_{j}$	Evaluated by the government reflecting the hospital's ability, equipment, and technology. Expressed as a dummy variable in which "1" represents that the physician works in a "III A" level hospital and "0" otherwise

Three independent variables were included in our study: (1) whether physician *j* received gifts in dialog *i* (gift giving), (2) gift type (affective or instrumental), and (3) tie strength.

To control for heterogeneity among physicians, we collected physicians' information online. Haodf.com provides three primary services for physicians to choose whether to provide, including written consultation, phone consultation, and an outpatient service appointment (details can be found in [33]). Three dummy variables were used to measure the three services. If a physician provided a written or phone consultation, the service price was also included in the model. The physician's professional title (chief, associate chief, or attending physician) evaluated by the health sector was also included. The detailed explanations for other control variables are provided in Table 1.

Our analyses were conducted in two steps. In the first step, multiple linear regression was used to explore the impacts of patients' gift giving (whether to give a gift) on physicians' service quality. In the second step, based on the situation in which gift giving occurred, the impacts of gift types and tie strength on physicians' service quality were examined. Therefore, only dialogs including gift giving were included in this step. Our empirical models for each step are as follows:

Step 1: Word count ratio = $\alpha_{ij} + \alpha_1 \times GiftGiving_{ij} + \alpha_2 \times Control_j + \varepsilon_j$

Step 2: Word count ratio = $\alpha'_{ij} + \alpha'_1 \times GiftGiving_{ij} + \alpha'_2 \times Control_j + \varepsilon'_j$

Word count ratio = $\alpha''_{ij} + \alpha''_1 GiftGiving_{ij} + \alpha''_2 \times TieStrength_{ij} + \alpha''_3 \times GiftGiving_{ij} \times TieStrength_{ij} + \alpha''_4 \times Control_j + \varepsilon''_j$

Word count ratio = $\alpha'''_{ij} + \alpha'''_1 GiftType_{ij} + \alpha'''_2 \times TieStrength_{ij} + \alpha'''_3 \times GiftType_{ij} \times TieStrength_{ij} + \alpha'''_4 \times Control_j + \varepsilon''_j$

where *i* denotes a dialog and *j* denotes a physician; α is the coefficient to be estimated. $Control_j$ represents the control variables for physician *j*, and ε_j is the standard error.

XSL•FO
RenderX

Results

Descriptive Statistics and Correlations

The descriptive statistics and correlations for the key variables are presented in Multimedia Appendix 2. Independent variables were significantly related to the dependent variable. The correlation matrix showed no serious multicollinearity, which assured obtaining reliable results.

Empirical Results

The multiple linear regression method was applied to examine the effect of patients' gift giving on physicians' service quality, and the empirical results are shown as Model 1 (adjusted R^2=0.018, F_{change}=19.774, P<.001) and Model 2 (adjusted R^2=0.027, F_{change}=27.171, P<.001). We found that the decision to give gifts positively influenced physicians' service quality, thereby supporting hypothesis H1 (Table 2).

We then investigated the impacts of gift types and tie strength on physicians' service quality. The results are shown in Model 3 (adjusted R^2=0.031, F_{change}=19.737, P<.001). Compared with instrumental gifts, affective gifts had a stronger influence on physicians' service quality, thereby supporting hypothesis H2 (Table 2).

In Model 4 (adjusted R^2=0.017, F_{change}=11.184, P<.001), we examined the moderating effect of tie strength on the relationship between gift giving and physicians' service quality. We found that the relationship between patients' gift giving and physicians' service quality was smaller when they had a strong tie (Table 2). In Model 5 (adjusted R^2=0.035, F_{change}=19.071, P<.001), a positive moderating effect of tie strength on the relationship between gift types and physicians' service quality was identified (Table 2). The relationship between gift types and physicians' service quality was smaller when there was a strong tie between the physician and patient, thereby supporting hypotheses H3a and H3b.

XSL•FO
RenderX

Table 2. Regression results on associations of gift giving and physical service quality.

Variables	Step 1, β (SE)				Step 2, β (SE)					
	Model 1	*P* value	Model 2	*P* value	Model 3	*P* value	Model 4	*P* value	Model 5	*P* value
Intercept	3.251 (0.344)	<.001	2.972 (0.344)	<.001	2.532 (-0.443)	<.001	2.775 (0.445)	<.001	2.513 (0.443)	<.001
PhonePrice	–0.034 (0.033)	.18	–0.017 (0.033)	.18	–0.137 (–0.044)	.008	–0.172 (0.044)	<.001	–0.143 (0.044)	.008
WrittenPrice	–0.040 (0.031)	.30	–0.052 (0.031)	.21	0.043 (–0.040)	.35	0.058 (0.041)	.35	0.044 (0.040)	.35
Outpatient	0.099 (0.035)	.009	0.111 (0.034)	<.001	0.024 (–0.045)	.42	0.024 (0.046)	.42	0.026 (0.045)	.42
Thank-you Letter	0.194 (0.033)	<.001	0.183 (0.033)	<.001	0.263 (0.044)	<.001	0.284 (0.044)	<.001	0.265 (0.044)	<.001
Contribution	–0.088 (0.060)	.31	–0.056 (0.060)	.31	0.137 (0.076)	.10	0.075 (0.077)	.31	0.137 (0.077)	.10
No. Gifts	–0.172 (0.030)	<.001	–0.188 (0.030)	<.001	–0.207 (0.039)	<.001	–0.156 (0.039)	<.001	–0.190 (0.039)	<.001
No. Patients	0.315 (0.060)	<.001	0.306 (0.059)	<.001	0.001 (0.078)	.01	0.002 (0.078)	.01	–0.018 (0.078)	<.001
Recommendation	–0.493 (0.076)	<.001	–0.506 (0.076)	<.001	–0.253 (0.097)	.004	–0.265 (0.098)	.004	–0.247 (0.097)	.004
Title	0.026 (0.035)	.54	0.020 (0.035)	.55	0.096 (0.046)	.05	0.084 (0.047)	.10	0.088 (0.046)	.10
City	0.083 (0.040)	.06	0.088 (0.040)	.06	-0.032 (0.053)	.57	-0.017 (0.054)	.89	-0.026 (0.053)	.55
Level	–0.292 (0.048)	<.001	–0.285 (0.048)	<.001	–0.268 (0.061)	<.001	–0.279 (0.062)	<.001	–0.256 (0.061)	<.001
GiftGiving	N/A[a]	N/A	0.300 (0.029)	<.001	N/A	N/A	N/A	N/A	N/A	N/A
GiftType	N/A	N/A	N/A	N/A	–0.415 (0.038)	<.001	N/A	N/A	–0.329 (0.051)	<.001
GiftGiving × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	–0.104 (0.025)	<.001	–0.137 (0.026)	<.001
GiftType × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.209 (0.077)	.04

[a]N/A: not applicable.

Robustness Check

The dependent variable in our study was physicians' service quality, which was measured by the ratio of the average word count for a physician in the patient in dialog *i*. Specifically, we divided the logarithm of the average word count for a physician by the logarithm of the average word count for the patient to calculate physicians' service quality. Robustness was calculated by dividing the average word count for a physician by that of the patient (see Table 3). These values were consistent with the main results of the model in Table 2, which assured the good robustness of the results.

XSL•FO
RenderX

Table 3. Robustness check.

Variables	Step 1, β (SE)				Step 2, β (SE)					
	Model 1[a]	*P* value	Model 2[b]	*P* value	Model 3[c]	*P* value	Model 4[d]	*P* value	Model 5[e]	*P* value
GiftGiving	N/A[f]	N/A	1.154 (0.692)	.10	N/A	N/A	N/A	N/A	N/A	N/A
GiftType	N/A	N/A	N/A	N/A	–7.332 (0.828)	<.001	N/A	N/A	–5.732 (1.114)	<.001
GiftGiving × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	–1.707 (0.533)	.009	–2.316 (0.561)	<.001
GiftType × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3.866 (1.681)	.06

[a]Adjusted R^2=.014, F_{change}=15.907, P<.001.

[b]Adjusted R^2=0.014, F_{change}=14.816, P<.001.

[c]Adjusted R^2=0.027, F_{change}=17.007, P<.001.

[d]Adjusted R^2=0.017, F_{change}=11.226, P<.001.

[e]Adjusted R^2=0.029, F_{change}=15.899, P<.001.

[f]N/A: not applicable.

Posthoc Analysis

We further tested the impacts of gift giving on physicians' bedside manner, which is also important for patients and a dimension of service quality. Text mining was used to analyze physicians' replies and to judge their attitude. TextMind, a psychoanalytic software system developed by the Chinese Academy of Sciences for language analysis (especially for Chinese), was used for this analysis. This software has been widely used to analyze various characteristics of the text content in previous studies (eg [34]). By using TextMind, we conducted a sentimental analysis for physicians' text content. The specific process is shown in Multimedia Appendix 3.

The average number of the physician positive sentiment words in each dialog was used to measure the physician's bedside manner. As shown in Table 4, we found a positive impact of gift giving on physicians' bedside manner; however, the impact was smaller when the tie strength was strong (Model 2). In addition, compared with instrumental gifts, affective gifts had a greater impact on physicians' bedside manner (Model 3), but the gap was reduced when the tie strength was stronger (Model 5). These results are consistent with our main results, indicating that gift giving works effectively with respect to both the physicians' reply and bedside manner.

Table 4. Empirical model results for physician bedside manner.

Variables	Step 1, β (SE)				Step 2, β (SE)					
	Model 1[a]	*P* value	Model 2[b]	*P* value	Model 3[c]	*P* value	Model 4[d]	*P* value	Model 5[e]	*P* value
GiftGiving	N/A[f]	N/A	0.342 (0.023)	<.001	N/A	N/A	N/A	N/A	N/A	N/A
GiftType	N/A	N/A	N/A	N/A	–0.348 (0.030)	<.001	N/A	N/A	–0.239 (0.041)	<.001
GiftGiving × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	–0.164 (0.024)	<.001	–0.196 (0.026)	<.001
GiftType × *TieStrength*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.232 (0.062)	<.001

[a]Adjusted R^2=0.028, F_{change}=16.743, P<.001.

[b]Adjusted R^2=0.064, F_{change}=34.801, P<.001.

[c]Adjusted R^2=0.056, F_{change}=19.318, P<.001.

[d]Adjusted R^2=0.078, F_{change}=26.869, P<.001.

[e]Adjusted R^2=0.091, F_{change}=27.562, P<.001.

[f]N/A: not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

The aim of this study was to determine the main motivations and effects of online virtual gift giving on physicians' service quality in online health communities using text mining and econometric methods. Specifically, we evaluated the impact of whether to give gifts and gift types systematically. Moreover, tie strength was carefully considered in the model as it may influence the effects of gift giving. Further, both the physicians' reply and bedside manner were used to measure service quality. The empirical results support all of our hypotheses.

Gift giving can enhance physicians' service quality, which was proven in our model and is consistent with prior studies [14]. As an important tool in social interactions [17], gift exchange is beneficial for individuals to participate in social contact. A reciprocity norm exists among people and is scrupulously obeyed; that is, people always repay each other for the mutual indebtedness within the reciprocity principle [19]. For example, in return for gifts from employers, workers will be more productive [35]. Therefore, we believe that physicians will improve their service quality when they have received gifts from patients under the reciprocity principle. In addition, gifts provide psychological gratification to the physician, as gifts from patients make them feel respected and recognized for their effort, and thus inspire them to improve their service quality.

Gifts with different motivations had different effects on influencing physicians' behaviors. We extended existing studies on gift types (eg, [15]) by investigating the consequences for giving different types of gifts, and found that although gift giving can inspire physicians to improve their quality overall, affective gifts play a more prominent role in enhancing quality compared with instrumental gifts. Our results are consistent with a previous study [32] that examined the impact of unconditional nonmonetary gifts on health worker performance in a low-income country in which unconditional nonmonetary gifts improved the performance of workers by 20% in 6 weeks compared with conditional nonmonetary gifts. Affective gifts are unconditional nonmonetary gifts, and instrumental gifts are similar to conditional nonmonetary gifts.

With respect to the moderating effects of tie strength between physicians and patients, we found that the more familiar a physician is with their patient, the role of whether to send gifts and gift type on physician service quality is decreased. Strong ties improve people's trust and obligation. Although weak ties play an important role in information transmission, a strong tie helps to obtain something more expensive and valuable [27]. Specifically, a strong tie between a physician and their patient will mitigate the effects of whether to send gifts and gift types on the physician's behavior. Our results suggest that a strong tie helps patients obtain high-quality medical services from physicians in an online health environment.

Based on the systematic empirical analyses on the motivation and effects of gift giving in an online health community, we have revealed the nature and mechanism of online gift giving. The value of the gift ranges from 5 to 200 RMB (approximately US $1 to $30), which is too small to be judged as "a red envelope" (cash wrapped in an envelope). Furthermore, an online virtual gift can improve the efficiency of physician-patient communication and contribute to the establishment of a good physician-patient relationship.

Theoretical Implications

This study contributes to theory from four aspects. First, to the best of our knowledge, this study is among the first to explore the motivations and consequences of gift giving in online health communities. Although previous studies have questioned the value of gift giving in the health sector (mainly in hospital settings) and even treat gift giving as a reason to explain the current tense physician-patient relationship and moral reduction in the medical field [36,37], the specific factors that influence gift giving in an online health environment have been unknown. In contrast to the study of Zhao et al [14], both the details of reply and bedside manner of physicians were included as potential consequences of gift giving in our study. Therefore, we empirically confirmed the role of gift giving in an online health environment.

Second, this study contributes to gift theory and tie strength theory by empirically examining the different effects of gift giving on receiver behaviors in different giver-receiver relationships, which were measured by tie strength. By using unique interaction data between physicians and patients in an online forum, this study revealed significant moderating effects of tie strength on the relationships between gift giving and physicians' behaviors. Our results suggest that environmental factors need to be carefully considered to make an accurate judgment on the effects of gift giving.

Third, to the best of our knowledge, this study is among the first to empirically explore the different effects of gift types, especially in health care. Based on the characteristics of both affective and instrumental gifts, we first coded each gift by analyzing the whole interaction process between physicians and patients and then examined their respective effects. These results contribute to the current literature on gift giving by demonstrating that these two types of gifts have different impacts on the improvement of physicians' service quality.

Fourth, this study contributes to prior studies by using a large sample with a real operation date and analysis under strict criteria. Compared with survey data, our data were collected from the open transaction platform and are therefore more practical. Moreover, text mining was used to conduct sentiment analysis of the whole interaction content between physicians and patients. Text mining can help to obtain more comprehensive and accurate information with respect to an individual's emotion.

Practical Implications

The practice of gift giving has been questioned since it was first launched on online health communities, even though patients are not obligated to buy physicians gifts. Therefore, exploring the motivations and impacts of gift giving in medical services in online health communities is important because it can provide guidance on how to improve communication efficiency. All of our results suggest that giving online virtual gifts is of

XSL•FO
RenderX

considerable significance to promoting effective physician-patient communication and is conducive to the relief of physician-patient conflicts. This study offers several important practical implications for both online health community users and administrators.

For patients, a gift can be provided to physicians when the patient desires better service, as we found that digital gifts may lead to a satisfactory reply during the interaction between patients and physicians. Moreover, patients should be aware that the gift will be more powerful when their relationship with the physician is more alienated. In addition, it is better to give an affective gift in the early stage of the interaction, as these types of gifts have a more powerful influence when the patient and the physician are strangers.

For administrators of online health communities, a more useful mechanism should be established to encourage patients to give digital gifts to the physician when necessary, such as construction of a more convenient interface. In addition, reminders can be set for patients to send a gift to physicians at the right time. Moreover, the site should not immoderately encourage patients to give gifts to physicians. The original objective of the gift-giving function is to provide a channel for patients to express their appreciation to physicians. Excessive gift giving will obstruct the development of good relations between physicians and patients. Therefore, we suggest that the administrators of online health communities set a suitable mechanism to moderate guidance for gift giving.

Limitations and Future Research

The limitations of this study can offer potential directions for future research. First, we did not obtain information on patients such as age, gender, and occupation, as patient information is rarely disclosed on the online health communities in China for privacy protection. Future research can attempt to obtain more thorough data or to eliminate the influences of patient characteristics using more complex techniques or methods. Second, as the number of physician replies by voice messages in the physician-patient interact process was very small, we deleted these from our analysis. However, voice messages may also include relevant information and may be more important than text messages as voice can convey emotion. Future studies should therefore attempt to obtain more data on voice content and useful related information.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Chinese terms expressing demand.
[DOCX File , 15 KB - jmir_v22i7e18569_app1.docx]

Multimedia Appendix 2
Descriptive statistics and correlations among variables.
[DOCX File , 15 KB - jmir_v22i7e18569_app2.docx]

Multimedia Appendix 3
Extraction of feature words from physician-patient dialogs.
[DOCX File , 436 KB - jmir_v22i7e18569_app3.docx]

References

1. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Inf Manag 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003]
2. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct 25;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
3. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
4. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, Participatory Medicine, and Ethical Care: A Focus Group Study of Patients' and Health Care Providers' Use of Health-Related Internet Information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
5. Wu B. Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
6. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: Empirical study from elaboration likelihood perspective. Comput Human Behav 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
7. Yang H, Guo X, Wu T. Exploring the influence of the online physician service delivery process on patient satisfaction. Decis Support Syst 2015 Oct;78:113-121. [doi: 10.1016/j.dss.2015.05.006]

XSL•FO
RenderX

8. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
9. Wu H, Deng Z. Knowledge collaboration among physicians in online health communities: A transactive memory perspective. Int J Inf Manag 2019 Dec;49:13-33. [doi: 10.1016/j.ijinfomgt.2019.01.003]
10. Haodf.com. URL: https://www.haodf.com/ [accessed 2020-07-15]
11. Chiu Y, Smith KC, Morlock L, Wissow L. Gifts, bribes and solicitions: print media and the social construction of informal payments to doctors in Taiwan. Soc Sci Med 2007 Feb;64(3):521-530. [doi: 10.1016/j.socscimed.2006.09.018] [Medline: 17070971]
12. Wang X. Are online gifts for doctors are red envelopes? Nanfang Daily. 2016. URL: http://news.163.com/16/0322/07/BIOB4SST00014AED.html [accessed 2020-07-01]
13. Ruth J, Otnes C, Brunel F. Gift Receipt and the Reformulation of Interpersonal Relationships. J Consum Res 1999 Mar;25(4):385-402. [doi: 10.1086/209546]
14. Zhao W, Guo X, Wu T, Geng J. The Impacts of Patients' Gift Giving Behavior on Physicians' Service Quality. 2017 Oct 28 Presented at: International Conference on Smart Health; 2017; Wuhan p. 42-53. [doi: 10.1007/978-3-319-67964-8_5]
15. Mitrut A, Nordblom K. Social norms and gift behavior: Theory and evidence from Romania. Eur Econ Rev 2010 Nov;54(8):998-1015. [doi: 10.1016/j.euroecorev.2010.03.002]
16. Lowrey TM, Otnes CC, Ruth JA. Social Influences on Dyadic Giving over Time: A Taxonomy from the Giver's Perspective. J Consum Res 2004 Mar 01;30(4):547-558. [doi: 10.1086/380288]
17. Mauss M. The Gift: The Form and Reason for Exchange in Archaic Societies. New York and London: Routledge; 2002.
18. Wu JB, Hom PW, Tetrick LE, Shore LM, Jia L, Li C, et al. The Norm of Reciprocity: Scale Development and Validation in the Chinese Context. Manag Organ Rev 2015 Feb 02;2(3):377-402. [doi: 10.1111/j.1740-8784.2006.00047.x]
19. Chen C, Hung S. To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. Inf Manag 2010 May;47(4):226-236. [doi: 10.1016/j.im.2010.03.001]
20. Pratt MG, Rockmann KW, Kaufmann JB. Constructing Professional Identity: The Role of Work and Identity Learning Cycles in the Customization of Identity Among Medical Residents. Acad Manage J 2006 Apr;49(2):235-262. [doi: 10.5465/amj.2006.20786060]
21. Beijaard D, Meijer PC, Verloop N. Reconsidering research on teachers' professional identity. Teach Teach Educ 2004 Feb;20(2):107-128. [doi: 10.1016/j.tate.2003.07.001]
22. Beijaard D, Verloop N, Vermunt JD. Teachers' perceptions of professional identity: an exploratory study from a personal knowledge perspective. Teach Teach Educ 2000 Oct;16(7):749-764. [doi: 10.1016/s0742-051x(00)00023-8]
23. Hong JY. Pre-service and beginning teachers' professional identity and its relation to dropping out of the profession. Teach Teach Educ 2010 Nov;26(8):1530-1543. [doi: 10.1016/j.tate.2010.06.003]
24. Wang GW, Shi XX, Dong-Ping MA. Relationship between professional identity and social support among general practitioners in Shandong province. Chinese J Public Health 2016:818-821.
25. Granovetter MS. The Strength of Weak Ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
26. Bian Y. Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China. Am Sociol Rev 1997 Jun;62(3):366. [doi: 10.2307/2657311]
27. Gee LK, Jones J, Burke M. Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network. J Labor Econ 2017 Apr;35(2):485-518. [doi: 10.1086/686225]
28. Song T, Yi C, Huang J. Whose recommendations do you follow? An investigation of tie strength, shopping stage, and deal scarcity. Inf Manag 2017 Dec;54(8):1072-1083. [doi: 10.1016/j.im.2017.03.003]
29. Gao H, Huang S, Brown G. The influence of face on Chinese tourists' gift purchase behaviour: The moderating role of the gift giver–receiver relationship. Tourism Manage 2017 Oct;62:97-106. [doi: 10.1016/j.tourman.2017.03.013]
30. Pai P, Tsai H. Reciprocity norms and information-sharing behavior in online consumption communities: An empirical investigation of antecedents and moderators. Inf Manag 2016 Jan;53(1):38-52. [doi: 10.1016/j.im.2015.08.002]
31. Yan Y. The Flow of Gifts: Reciprocity and Social Networks in a Chinese Village. Redwood City: Stanford University Press; May 01, 1996.
32. Kirchler M, Palan S. Immaterial and monetary gifts in economic transactions: evidence from the field. Exp Econ 2018 Aug 2;21(1):205-230 [FREE Full text] [doi: 10.1007/s10683-017-9536-1] [Medline: 29449784]
33. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: An examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
34. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing simplified Chinese psychological linguistic analysis dictionary for microblog. In: Imamura K, Usui S, Shirao T, Kasamatsu T, Schwabe L, Zhong N (eds) Lecture Notes in Computer Science volume 8211. Cham: Springer; 2013 Presented at: International Conference on Brain and Health Informatics; October 29-31, 2013; Maebashi, Japan p. 359-368. [doi: 10.1007/978-3-319-02753-1_36]
35. Kube S, Maréchal MA, Puppe C. The Currency of Reciprocity: Gift Exchange in the Workplace. Am Econ Rev 2012 Jun 01;102(4):1644-1662. [doi: 10.1257/aer.102.4.1644]

XSL•FO
RenderX

36. Zhu W, Wang L, Yang C. Corruption or professional dignity: An ethical examination of the phenomenon of "red envelopes" (monetary gifts) in medical practice in China. Dev World Bioeth 2018 Mar 21;18(1):37-44. [doi: 10.1111/dewb.12152] [Medline: 28544334]
37. Hu KJ, Xu WD. Towards the end of "red envelopes" in China? Lancet 2014 Jun 28;383(9936):2211. [doi: 10.1016/S0140-6736(14)61083-7] [Medline: 24976327]

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by Q Ye, C Zhang; comments to author 19.03.20; revised version received 24.03.20; accepted 19.04.20; published 30.07.20.

Please cite as:
Peng L, Wang Y, Chen J
Consequences of Gift Giving in Online Health Communities on Physician Service Quality: Empirical Text Mining Study
J Med Internet Res 2020;22(7):e18569
URL: http://www.jmir.org/2020/7/e18569/
doi:10.2196/18569
PMID:32729834

©Li Peng, Yanan Wang, Jing Chen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Use of Social Media to Increase the Impact of Health Research: Systematic Review

Marco Bardus[1], BA, MA, PhD; Rola El Rassi[2], MSc; Mohamad Chahrour[3], BSc; Elie W Akl[3], MD; Abdul Sattar Raslan[3], BSc; Lokman I Meho[4], BA, MA, MSLS, PhD; Elie A Akl[5], MD, MPH, PhD

[1]Department of Health Promotion and Community Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[2]Clinical Research Institute, American University of Beirut, Beirut, Lebanon

[3]Department of Pediatric and Adolescent Medicine, American University of Beirut Medical Center, Beirut, Lebanon

[4]University Libraries, American University of Beirut, Beirut, Lebanon

[5]Department of Internal Medicine, American University of Beirut Medical Center, Beirut, Lebanon

Corresponding Author:
Elie A Akl, MD, MPH, PhD
Department of Internal Medicine
American University of Beirut Medical Center
PO Box 11-0236, Riad-El-Solh
Beirut, 1107 2020
Lebanon
Phone: 961 1 374 374
Email: ea32@aub.edu.lb

Abstract

Background: Academics in all disciplines increasingly use social media to share their publications on the internet, reaching out to different audiences. In the last few years, specific indicators of social media impact have been developed (eg, Altmetrics), to complement traditional bibliometric indicators (eg, citation count and h-index). In health research, it is unclear whether social media impact also translates into research impact.

Objective: The primary aim of this study was to systematically review the literature on the impact of using social media on the dissemination of health research. The secondary aim was to assess the correlation between Altmetrics and traditional citation-based metrics.

Methods: We conducted a systematic review to identify studies that evaluated the use of social media to disseminate research published in health-related journals. We specifically looked at studies that described experimental or correlational studies linking the use of social media with outcomes related to bibliometrics. We searched the Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica dataBASE (EMBASE), and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases using a predefined search strategy (International Prospective Register of Systematic Reviews: CRD42017057709). We conducted independent and duplicate study selection and data extraction. Given the heterogeneity of the included studies, we summarized the findings through a narrative synthesis.

Results: Of a total of 18,624 retrieved citations, we included 51 studies: 7 (14%) *impact studies* (answering the primary aim) and 44 (86%) *correlational studies* (answering the secondary aim). Impact studies reported mixed results with several limitations, including the use of interventions of inappropriately low intensity and short duration. The majority of correlational studies suggested a positive association between traditional bibliometrics and social media metrics (eg, number of mentions) in health research.

Conclusions: We have identified suggestive yet inconclusive evidence on the impact of using social media to increase the number of citations in health research. Further studies with better design are needed to assess the causal link between social media impact and bibliometrics.

(J Med Internet Res 2020;22(7):e15607) doi:10.2196/15607

KEYWORDS

social media; research; bibliometrics; Altmetrics; journal impact factor; translational medical research

XSL•FO
RenderX

Introduction

Social Media and Its Impact in Health Research

Individual researchers and academic institutions use social media to disseminate their research findings to a broad audience that includes the public and health care practitioners. According to Carr and Hayes [1], social media can be broadly defined as "internet-based, disentrained, and persistent channels of mass personal communication facilitating perceptions of interactions among users, deriving value primarily from user-generated content". For example, social media can be used by clinicians who would benefit from the findings in dealing with patients. A recent scoping review of the literature on social media use in health research revealed that social media are increasingly used to recruit patients, collect data, and establish and maintain user engagement, especially in the name of research dissemination [2]. At the same time, in the era of the internet and web 2.0 technologies that allow content generation and sharing [3], social media have been increasingly used as a source for measuring the impact of research, as they contribute to knowledge generation, dissemination, and translation [4]. Indeed, social media are increasingly required to measure their research performance and demonstrate the value of their research to governments and funding organizations [5]. Researchers in the domain of health and medical sciences have been particularly concerned about demonstrating the impact of their work as it usually bears implications for public health [6]. Hence, measuring the impact of health research is essential for influencing policy-making processes, improving health systems, and health-related socioeconomic impact [6]. According to the research impact framework (RIF) for health research [4], the domain of research-related impact is generally and traditionally evaluated according to conventional and traditional bibliometric approaches, which generally include the number of citations, the impact factor (IF), or the h-index [7]. Some academic institutions in the United States, Canada, and Europe started including social media impact as an evaluation criterion in their tenure and promotion policies [8]. According to the systematic review by Cruz Rivera et al [6] on the impact of health care research, researchers should consider indicators such as "the number of reads for published articles; article download rate and number of journal webpage visits; and citations rates in non-journal media such as newspapers and mass and social media (ie, Twitter and blogs)".

Researchers can now discuss their papers and share their publications on various social networking sites [9] or with the *general public* interested in their topic. Using typology of social media by Constantinides and Fountain [10], health researchers can use 5 types of dissemination platforms: (1) blogs or web-based journals; (2) social networking sites, such as Facebook or Twitter, or professionally oriented platforms such as LinkedIn, ResearchGate, and Academia, which have recently emerged as social networking sites for academics [11]; (3) content communities; (4) forums or bulletin boards; and (5) content aggregators, such as Diigo, CiteULike, Delicious, Evernote, or through reference management software such as EndNote, Mendeley, RefWorks, Papers, and Zotero. These software companies have developed proprietary communities of users who share citations. In this way, an article can reach new and widespread audiences, broader than those of the limited subscribers of academic journals whose content is generally protected by paywalls. Some publishers have also embraced the movement of social media dissemination by including options for authors to semiautomatically share their output on academic content aggregators that include dashboards to measure the reach and impact of social media posts, based on the link or digital object identifier (DOI) associated with each article. Examples include free platforms such as the nonprofit-owned ImpactStory; or the profit-oriented Kudos, independent from publishing houses; or Publons, owned and managed by Clarivate Analytics; and PlumX [12], owned by Elsevier.

Social Media Impact Measures

How is social media impact defined and measured? In the late 1990s, researchers started studying the phenomenon of internet-based dissemination of knowledge and information, forging the terms of *webometrics* or *cybermetrics* [13]. In 2010, a group of researchers defined a new set of metrics, Altmetrics (which stands for *alternative metrics*) that include web-based metrics (eg, number of link shares, likes, tweets, and views) and qualitative data that are complementary to traditional, citation-based metrics [14]. The so-called Altmetrics attention score (AAS) includes various indices of performance of a paper, such as the number of views; the number of discussions on social media (tweets, Facebook posts, and Wikipedia pages); recommendations (eg, Faculty of 1000); saved articles on popular social bookmarking services such as Mendeley or CiteULike; and the number of citations obtained from Google Scholar, CrossRef, PubMed Central, and Scopus [15]. Despite some intrinsic limitations—as the AAS can be inflated by self-citations, automatic retweeting, or sharing on various social media platforms [7]—Altmetrics is considered the current standard for measuring the impact of research on the internet and beyond [14-16].

Similar to the number of citations, the number of social media mentions of an article is a function of time since its publication. In 2011, Eysenbach published, in this journal, a seminal paper entitled, *Can tweets predict citation metrics on social media* [17]. The author proposed a set of measures for social media impact that would account for the dimension of *time*. The paper introduced the concepts of *Tweetations*, *Twimpactfactor*, and *Twindex*. Tweetations were defined as "citations in a tweet," that is, tweets mentioning the exact journal article URL (hence excluding links to DOI, PubMed entries, or other links). The twimpact factor was defined as the "cumulative number of tweetations within n days after publication" (eg, tw7 means the total number of tweetations after 7 days) [17]. The Twindex (or tweetation index) was defined as a "metric ranging from 0 to 100 indicating the relative standing of an article compared to other articles." It is based on a percentile, rank-ordering of an article by tweetations, relative to other articles published in the same journal, which were published around the same time [17]. The author recommended that papers investigating the relationship between citations and social media mentions should adjust for time since the publication of an article (or specify a timeframe when these metrics were obtained), journal type,

XSL•FO
RenderX

seasonal variations, and other possible confounders that might explain a nonlinear distribution of social media mentions.

Social Media and Bibliometrics

How is social media dissemination related to citations? Social media dissemination is generally associated with the higher reach of an article. Some research, sponsored by Academia.edu, showed that by sharing an article on this specific academic, social networking site, a researcher could receive up to 69% more citations over 5 years [18]. In Eysenbach's (2011) paper, the author found a positive correlation between social media and subsequent citations and social media [17]. However, to what extent is there a causal link between the use of social media and subsequent citations? In other words, what is the impact of social media on citations and bibliometric indicators? To the best of our knowledge, to date, there is only one experimental study, using a randomized controlled trial (RCT) design, that has clearly shown a causal relationship between subsequent citations and the dissemination of articles through a proprietary web-based distribution platform (TrendMD) [19]. Although TrendMD does not meet the definition of social media as intended in this paper [1], as it is not based on user-generated content, but on content that is pushed by an algorithm of sponsored recommendations, the study provides a useful benchmark. The authors found significant effects on citations and Mendeley saves for the intervention group (TrendMD) compared with the control (no diffusion on TrendMD) after 6 and 12 months [19]. In addition, the intervention had positive effects on citations at both 6 and 12 months for articles in the area of health and medical sciences [19].

However, evidence from systematic reviews on this matter seems to be scarce. A basic search for systematic reviews on Altmetrics on Google Scholar, on April 2, 2019 (in titles of articles) yielded only 24 results, 3 of which were reviews: 1 systematic review mapping the evidence for marketing research [20] and 2 literature reviews on Altmetrics used for generic scholarly research output [21,22]. In terms of health research, the review evidence is also limited. To the best of our knowledge, there is only one ongoing systematic review—whose protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO) [23]—that is aimed at evaluating the use of social media to disseminate research findings among health care professionals. A similar research conducted on April 2, 2019, on PubMed, yielded 18 hits, 2 of which were systematic reviews: the first focused on medical research output and reported significant associations between Altmetrics and traditional citations, without linking measures of impact [24]; the second is a systematic review of reviews and meta-analyses and focused on the evaluation of methodological quality in the domain of skin psoriasis [25]. The latter reported that a journal's IF could predict the number of tweets, whereas the years of publication and number of Mendeley readers predicted the number of citations on Google Scholar. Nevertheless, the authors concluded there does not seem to be a connection between scientific quality, social media, activity, and article usage [25].

Therefore, the primary aim of this study was to systematically review the literature on the impact of using social media on the dissemination of health research. The secondary aim was to assess the correlation between Altmetrics and traditional citation-based metrics.

Methods

Protocol and Eligibility Criteria

We developed a priori protocol for this systematic review and registered it in the PROSPERO database (CRD42017057709) [23]. This review focuses on health research, which includes biomedical, epidemiological, clinical, public health, and health systems research [26]. The inclusion criteria were as follows:

- *Population*: The unit of analysis of this review is studies (ie, study reports) published in health-related journals, including primary and secondary research, and editorials.
- *Interventions/exposures*: For both aims, we included studies that evaluated the use of social media to disseminate health research, which reported the outcomes described below. In the protocol, we included examples of social media and web 2.0 applications, defined as "a group of internet-based applications that build on the ideological and technological foundations of web 2.0, and that allow the creation and exchange of user generated content" [27]. Web 2.0 applications included interactive websites, blogging platforms, social networking, and social sharing sites, as described in the *Introduction* section.
- *Comparator*: The comparator for the primary aim (assessing the impact) was not using social media.
- *Outcome measures*: To be included, studies had to report measures of research dissemination such as traditional bibliometrics, Altmetrics, or webometrics indicators. Traditional bibliometric indicators were defined as quantitative data and statistics to publications such as journal articles and their accompanying citation counts. Altmetrics were defined as web-sourced metrics and qualitative data that are complementary to traditional, citation-based bibliometrics [15]. Webometrics were defined as the study of the quantitative aspects of the construction and use of information resources, structures, and technologies on the web, by drawing on bibliometric and infometric approaches [13]. Studies were included if they measured Altmetrics about webometrics and traditional bibliometrics, without any restrictions to any specific types of metrics.
- *Study design*: Experimental studies (eg, RCTs and nonrandomized studies or cohort studies), case series, and case studies.
- *Publication type*: We included original research papers, including scientific meeting abstracts or research letters, if they contained sufficient information to fill the extraction forms.

Search Strategy

We searched the following electronic databases on July 12, 2017: Medical Literature Analysis and Retrieval System Online (MEDLINE; access via Ovid), Excerpta Medica dataBASE (EMBASE; access via Ovid), and Cumulative Index to Nursing and Allied Health Literature (CINAHL; access via EBSCO). We updated the searches on August 22, 2019. We developed a

XSL•FO
RenderX

search strategy with the help of a health information specialist with experience in systematic reviews. The search strategy encompassed 2 main concepts: *social media* (based on the search strategy of a related systematic review by one of the authors [28]) and impact on dissemination (including *bibliometrics*, *Altmetric* and *academic impact*). The search strategy used both keywords and medical subject heading terms. The search strategies used are provided in Multimedia Appendix 1.

We did not use language restrictions. We did restrict the search timeframe by using the start date of January 2005, the year of the introduction of the *web 2.0* concept [3]. We also reviewed the reference lists of relevant papers and searched our files for both published and unpublished studies.

Selection Process

Before starting the selection process, we conducted calibration exercises to train the reviewers and clarify the eligibility criteria. In all, 3 reviewers (EWA, MC, and ASR) and 2 research assistants screened the titles and abstracts of identified references in duplicate and independently for potential eligibility.

We obtained the full texts for all references judged as potentially eligible by at least one reviewer through our local library. For articles not found through the library, we searched on Google, Google Scholar, ResearchGate, or Academia.edu to locate self-archived copies. We contacted corresponding authors via ResearchGate or emailed them to obtain a copy of the articles. Then, the same 3 reviewers screened the full texts in duplicate and independently for eligibility, using a standardized pilot-tested full-text screening form. Once acceptable interrater reliability was achieved, the reviewers completed the selection tasks independently. All disagreements were resolved by discussion and with the help of a third reviewer (MB), who double-checked all decisions and confirmed the reasons for exclusion.

We assessed the interrater reliability for titles and abstract screening using Gwet AC1 index, as it is less prone to bias when there is a large disproportion in binary categories (ie, excluded vs included articles) that are not judged as relevant [29,30]. We assessed the interrater reliability for full-text screening using the kappa statistic.

Data Abstraction Process

The review teams abstracted data from eligible studies in duplicate and an independent manner, using a standardized and pilot-tested data abstraction form with detailed instructions. Disagreements were resolved through discussions and with the help of a third reviewer (MB and EAA). Any inconsistencies in the abstraction tables were discussed within the research team (MB, EAA, RER, EWA, MC, and ASR) until consensus was reached. The abstracted data items included:

1. general information about the report, such as the first author's name, year of publication, type of study (eg, experimental, cross-sectional, cohort, or qualitative), health area, journal, population/unit of analysis, sources of data, and period of investigation;
2. metrics reported, such as type of social media used and type of metrics (eg, social media metrics, such as Altmetric attention score; bibliometrics, such as citations; and webometrics, such as page views and number of downloads);
3. results, as reported by the study authors; and
4. funding and reported conflicts of interest.

For experimental studies, we collected additional specific information about the intervention (eg, sample size, frequency and reach of the intervention, duration and frequency of the intervention, and profile owner) and control conditions, where applicable. We also extracted the information about effects (eg, *F* tests and *t* tests) and *P* values, as reported by the authors. We used a web-based effect size calculator [31] to estimate effect sizes if they were not included in the original publication. One author (MB) checked all abstraction tables for consistency.

When correlations between social media metrics and citations were reported, following Eysenbach recommendations [17], we also extracted details about whether the paper: (1) reported social media metrics adjusted for time (eg, Twimpact factor) or provided a rationale for selecting a timeframe to assess the relationship, (2) included social media metrics that adjust for some kind of confounders (eg, using Twindex metric or stratifying by article type and/or topic), (3) the type of correlation test used (eg, using Spearman rho and/or Pearson *r*), and (4) explored the correlation using scatterplots or employed tests for nonlinear relationships (eg, log-linear and/or nonparametric tests).

Data Synthesis

Given the heterogeneity of the included studies in terms of characteristics of the population, health area, study design, and reported outcomes (including *P* values and correlation coefficients), we summarized the findings through a narrative synthesis. In the summary tables included in this review, we reported *P* values and correlation coefficients and measures of effect sizes, as explicitly mentioned by the authors of the selected studies. We followed *Journal of Medical Internet Research's* convention for reporting *P* values (3 digits) and correlation coefficients (2 digits). We included the text of the original source in quotation marks. For studies reporting correlations between social media and citation metrics, we defined the methodological quality of the paper using 4 indicators, using Eysenbach (2011) paper as a benchmark [17]: (1) appropriately adjusting for time, (2) appropriately adjusting for confounders, (3) appropriately exploring correlations, and (4) appropriately reporting nonlinear correlations tests and statistics. Appropriately adjusting for time means that the article accounted for the variability in the metric by time or specified a time when the social media metric was obtained about the time of data analysis. Appropriately adjusting for confounders means that the social media metric was adjusted for confounders, such as journal type, article type, and for the journal and the season, for example by using the Twindex metric, which is a percentile ranking relative to other articles published in the same journal and the same period. Appropriately exploring correlations involves the authors checking for correlations between social media metrics and bibliometrics by inspecting

XSL•FO
RenderX

scatterplots. Appropriate reporting of tests for nonlinear correlations, such as Spearman rho correlations, log-linear tests of relationships, or other statistics, was based on ranking for non-normal distributions in the citations and social media mentions. We used the following scoring convention: (1) accounting for the time when selecting the data timeframe or acknowledging the role of time since publication (+) and adjusting the social media metric for time (++); (2) appropriately adjusting for confounders such as article type, topic, and/or subject (+) and seasonality or time factors using Twindex or similar metrics that account for the relative ranking of the article to the journal and season (++); (3) appropriately exploring correlations by including scatterplots (+); and (4) appropriately reporting nonlinear correlation tests and statistics (+) as well as log-linear relationship tests (++).

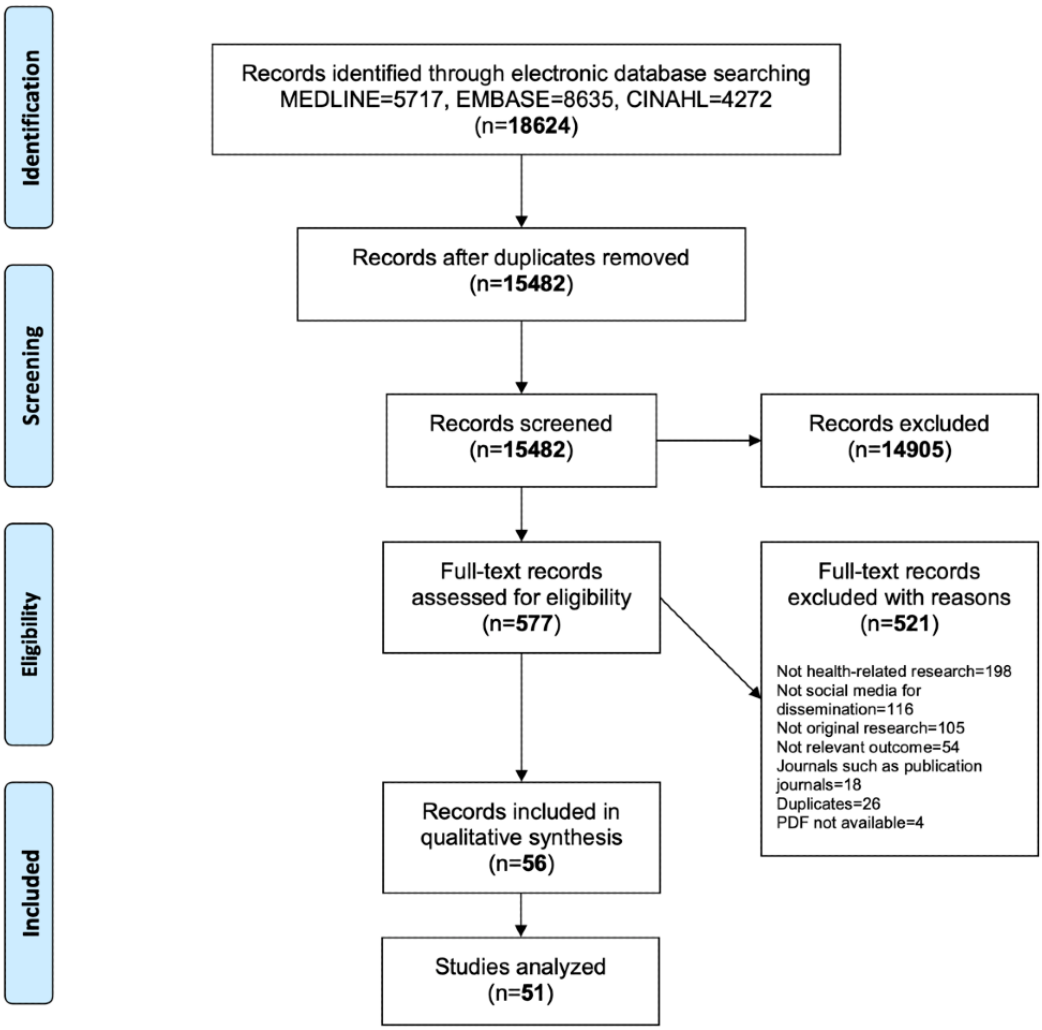
Results

Study Selection

Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. Searches in the selected electronic databases yielded 13,576 records in July 2017 and 5048 in August 2019, totaling 18,624. Of these, 577 were selected for full-text screening, following a duplicate and independent selection process. The level of agreement at the title and abstract screening phase was high (mean Gwet AC1 0.96, SD 0.03). Similarly, we excluded 521 of 577 full-text records, achieving a substantial interrater agreement (mean Kappa 0.69, SD 0.22; mean Gwet AC1 0.77, SD 0.04). We excluded these 521 articles for the following reasons (a list of excluded records is provided in Multimedia Appendix 2): 198 did not discuss health-related research; 116 discussed social media application use, but not for disseminating research; 105 were not original research articles (ie, editorials, commentaries, conceptual papers, literature reviews, praising the use of social media for research dissemination); 54 did not report relevant study outcomes (ie, focusing either on Altmetrics, social media, or citation metrics separately); and 18 were records discussing the use of social media to increase the impact of a journal. Other reasons for exclusion were as follows: 26 were duplicates and 4 were citations of conference abstracts. After consensus-seeking discussions, we judged 56 records as eligible for inclusion in this systematic review, representing 51 unique studies, as 5 studies presented the same data in different publications (an abstract followed by a publication in a journal). The studies were by Amath [32,33], Hayon [34], Knight [35,36], Nolte [37,38], and O'Connor [39,40]. A total of 5 articles included only a conference abstract but were deemed to provide sufficient data for inclusion [41-45].

XSL•FO
RenderX

Figure 1. PRISMA flow diagram. MEDLINE=Medical Literature Analysis and Retrieval System Online, EMBASE=Excerpta Medica dataBASE, CINAHL=Cumulative Index to Nursing and Allied Health Literature.



Characteristics of the Included Studies

A total of 7 of 51 studies (14%) were categorized as *impact studies* [46-52], as they presented some interventions that tested the use of social media to disseminate research articles. The main characteristics of impact studies are summarized in Table 1. In all, 86% (44/51) studies were categorized as *correlational studies* [16,17,25,32,34,36,38,39,41-45,53-83], as they described the associations between traditional bibliometrics and Altmetrics. Correlational studies were classified according to the number of quality indicators as follows: *very good quality*, with 4/4 indicators (7/44, 16% studies; Table 2); *good quality*, with 3/4 indicators (8/44, 18% studies; Table 3); *fair quality*, with 2/4 indicators (10/44, 23% studies; Table 4); *poor quality*, with 1/4 indicators (12/44, 27% studies; Table 5); and *very poor quality*, with 0/4 indicators (7/44, 16% studies; Table 6). Detailed information of all 51 studies is included in Multimedia Appendix 3. In the next paragraphs, the results are presented separately, following the primary and secondary aims of this study.

XSL•FO
RenderX

Table 1. Characteristics of studies assessing the impact of social media interventions (n=7).

References	Health research area	Unit and period of analysis	Type of study	Social media interventions	Metrics reported	Results
Allen, 2013 [48]	Clinical pain sciences	16 original research articles from *PLOS ONE* (2006-2011)	Quasi-experimental (before-after)	Blog posts shared on Facebook, Twitter, LinkedIn, and ResearchBlogging.org	Citations (Scopus); HTML views and PDF downloads	Significant increase in HTML and PDF views; no significant effect on citations approximately 1 year after publication
Cawcutt 2019 [50]	Women's health	8 original research articles published in 8 journals (2018)	Quasi-experimental (before-after)	Tweets chatted during a Physician's Weekly tweet chat event (#PWChat)	Article downloads; AAS[a]	Increased AAS; increased downloads (statistical significance not reported)
Fox, 2015 [46]	Cardiology	243 articles, 121 intervention and 122 control, available from *Circulation* journal (2013-2014)	Experimental (RCT[b])	Twitter and Facebook posts through the official circulation of social media accounts	HTML views and PDF downloads	No significant difference in 30 days' HTML views (and downloads)
Fox, 2016 [47]	Cardiology	152 articles, 74 intervention and 78 control, available from *Circulation* journal (2015)	Experimental (RCT)	Twitter and Facebook posts through the official circulation of social media accounts	HTML views and PDF downloads	No statistically significant difference in 6-day or 30-day page views (and downloads)
Hoang, 2015 [52]	Radiology	2 research articles appearing on the *American Journal of Neuroradiology* and the *American Journal of Roentgenology* (2013-2014)	Quasi-experimental (retrospective cohort)	Blog posts on Radiopedia.org; podcast shared on Twitter and Facebook	HTML views and PDF downloads	Increased page views during the intervention; no increased activity beyond the podcast
Thoma, 2018 [51]	Emergency medicine	29 articles selected for intervention and control from the *Canadian Journal of Emergency Medicine* (2016)	Experimental (RCT)	Podcast or infographic or *standard social media promotion* through Twitter and Facebook	HTML views	Using podcasts and infographics was associated with increased Altmetric scores and abstract views but not full-text article views; they did not significantly increase full-text readership
Tonia, 2016 [49]	Public health	130 articles, 65 intervention and 65 control, from the *International Journal of Public Health*	Experimental (RCT)	*International Journal of Public Health* blog, Twitter, and Facebook accounts dissemination	Article abstract, PDF views, and downloads; citations; AAS	Number of downloads and the number of citations significantly correlated for all papers, with the correlation being stronger in the intervention group

[a]AAS: Altmetrics attention score.

[b]RCT: randomized controlled trial.

XSL•FO
RenderX

Table 2. Characteristics of correlational studies of very good quality (n=7).

Study ID and reference	Health research area/unit and period of analysis	Metrics reported	Results	Methodological quality indicators[a]			
				1	2	3	4
Costas, 2015 [61]	Biomedical and health sciences; 217,115 articles in health sciences available from WoS[b] (2011-2013)	Social media: Altmetrics-Bibliometrics: Citations (WoS[b])	Positive relationship between number of Altmetrics and the average citation impact and citation scores	+	+	+	+
Delli, 2017 [64]	Dental medicine; 100 articles with highest AAS[c] from Altmetric Explorer and JCR[d] (2015)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	No significant correlation between Altmetrics and citations	+	+	+	+
Eysenbach, 2011 [17]	Medical informatics; 208 tweets including links to 286 *JMIR*[e] articles (2008-2010)	Social media: Twitter-Bibliometrics: Citations (Google Scholar and Scopus)	*Moderate* correlations	+	++	+	++
Haustein, 2014 [16]	Biomedical and health sciences; 1,431,576 biomedical and health sciences articles available on PubMed (2010-2012)	Social media: Twitter; Altmetrics-Bibliometrics: Citations (WoS[b])	*Moderate* correlations	+	+	+	+
Knight, 2014 [35,36]	Organ transplantation; 6979 articles with citation data available; 1346 with social media mention (2011-2012)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	Significant correlations between social media mentions and citations	+	+	+	+
Livas, 2018 [71]	Orthodontics; Top 200 articles in orthodontics available from Altmetrics Explorer (2017)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	No correlation was observed between Altmetrics score and citations	+	+	+	+
Maggio, 2018 [72]	Health profession education; 2486 articles with Altmetrics published in health profession education (2013-2015)	Social media: Altmetrics-Bibliometrics: Citations (WoS[b])	Significant correlations between Altmetrics and bibliometrics, but moderate effects	+	+	+	+

[a]1: appropriately adjusting for time of the social media metric (+); 2: appropriately adjusting for confounders such as article type (+) and seasonality/time factors (++); 3: appropriately exploring correlations by including scatterplots (+); 4: appropriately reporting nonlinear correlations tests and statistics (+) as well as log-linear relationship tests (++).

[b]WoS: Web of Science.

[c]AAS: Altmetrics attention score.

[d]JCR: Journal Citation Reports.

[e]JMIR: Journal of Medical Internet Research.

XSL•FO
RenderX

Table 3. Characteristics of correlational studies of good quality (n=8).

Study ID and reference	Health research area/unit and period of analysis	Metrics reported	Results	Methodological quality indicators[a]			
				1	2	3	4
Dal-Ré, 2017 [62]	Medical sciences; 410 original investigations and 182 opinion articles published in the first 4 printed issues of 4 top-ranked general medicine journals and 1 top-ranked journal on 5 different medical specialties that provide Altmetric scores (2015-2016)	Social media: Altmetrics-Bibliometrics: Citations (Google Scholar)	AAS[b] was *strongly/moderately* associated with citation count	+	+	−	+
Haustein, 2015 [66]	Biomedical and health sciences; 1,339,297 articles, of which 595,254 in biomedical and health sciences, available from WoS[c] (2012)	Social media: Altmetrics-Bibliometrics: Citations (WoS[c])	No significant correlation between Altmetrics and citations	−	+	+	+
Jabaley, 2018 [67]	Sepsis research; Top 50 articles available on PubMed (via query; 2012-2017)	Social media: Altmetrics-Bibliometrics: Citations (Scopus and WoS[c])	*Weak* to *moderate* correlations between Altmetrics and citations	+	+	+	−
O'Connor, 2017 [39,40]	Urology; Top 5 articles of top 10 journals in urology (2014-2015)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	*Weak* positive correlation between Altmetric score and citations	−	+	+	+
Rosenkrantz, 2017 [76]	Radiology; 892 articles from selected radiology journals (2013)	Social media: Altmetrics-Bibliometrics: Citations (WoS[c])	Significant but *weak* correlation between the citation count and both the AAS[b] and the number of Twitter mentions	−	+	+	+
Scotti, 2017 [78]	Not specified, hospital; 268 articles with Altmetric score out of 646 articles published in 2013 in indexed journals (with a 2012 IF[d] score) by researchers affiliated to the authors' hospital (2013)	Social media: Altmetrics-Bibliometrics: Citations (WoS[c])	Altmetrics significantly associated with IF[d] as well as Facebook, Twitter, and Mendeley	−	+	+	+
Thelwall, 2013 [81]	Not specified; 171-135,331 articles with nonzero Altmetric score and a valid PubMed ID (2011)	Social media: Altmetrics-Bibliometrics: Citations (WoS[c])	Significant correlations between most Altmetrics and citations	+	+	−	++
Thelwall, 2016 [82]	Medical sciences; 290,282 articles from 45 fields in Scopus Medicine (2009)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	Significant correlations between Mendeley readers and citations	+	+	−	+

[a]1: appropriately adjusting for time of the social media metric (+); 2: appropriately adjusting for confounders such as article type (+) and seasonality/time factors (++); 3: appropriately exploring correlations by including scatterplots (+); 4: appropriately reporting nonlinear correlations tests and statistics (+) as well as log-linear relationship tests (++).

[b]AAS: Altmetrics attention score.

[c]WoS: Web of Science.

[d]IF: impact factor.

XSL•FO
RenderX

Table 4. Characteristics of correlational studies of fair quality (n=10).

Study ID and reference	Health research area/unit and period of analysis	Metrics reported	Results	Methodological quality indicators[a]			
				1	2	3	4
Araujo, 2018 [54]	Physiotherapy; 200 randomly selected articles from physiotherapy evidence database (PEDro; 2013-2016)	Social media: Altmetrics mentioned/reader-Bibliometrics: Citations (WoS[b])	Significant correlation with citations	+	+	−	−
Calopedos, 2017 [58]	Urology; 22 urology articles in English language identified via PubMed (2010-2015)	Social media: Altmetrics-Bibliometrics: Citations (Google Scholar)	Significant correlation between Altmetrics and citations	+	−	+	−
Chang, 2019 [59]	Pediatric surgery; 140 articles appearing on 14 core journals on pediatric surgery (2012-2015)	Social media: Altmetrics-Bibliometrics: Citations (Scopus); IF[c] (JCR[d])	*Strong* correlations between Altmetric scores and citations, but not IF[c]	+	−	+	−
Dardas, 2019 [63]	Nursing; 100 articles in nursing with highest AAS[e] from WoS[b] (2012-2018)	Social media: Altmetrics-Bibliometrics: Citations (WoS[b] and Scopus)	Significant *moderate* correlation between Altmetrics and citation counts	−	+	+	−
Hassona, 2019 [65]	Dental medicine; 100 articles with highest AAS[e] from Altmetric Explorer (2018)	Social media: Altmetrics-Bibliometrics: Citations (Google Scholar and Scopus)	No significant correlation between Altmetrics and citations	−	+	−	+
Liu, 2013 [70]	Field not specified; 33,128 articles appearing in *PLOS One* (2011)	Social media: Altmetrics-Bibliometrics: HTML views, PDF downloads, and citations (Scopus, PubMed, and CrossRef)	Significant correlations between Altmetrics and bibliometrics	−	−	+	+
Nolte, 2019 [37,38]	Urology; 44 articles tweeted about the 2015 American Urological Association meeting (2015)	Social media: Twitter-Bibliometrics: IF[c]	Positive significant correlation with subsequent publication IF[c] within 18 months of presentation	−	+	+	−
Punia, 2019 [73]	Neurological research; Top 100 articles from top 5 neurology journals (2016)	Social media: Altmetrics-Bibliometrics: Citations	*Weak* positive correlation between Altmetric score and citations	−	−	+	+
Quintana, 2016 [74]	Psychiatry; 438 articles in the *American Journal of Psychiatry* (2013-2015)	Social media: Twitter-Bibliometrics: Citations (WoS[b])	Positive correlation between Twitter mentions and citations	+	−	−	++
Ruano, 2018 [25]	Psoriasis research; 164 systematic reviews or meta-analyses available from MEDLINE[f], EMBASE[g], and Cochrane databases (2016)	Social media: Altmetrics-Bibliometrics: Citations (Google Scholar)	No significant correlation between Altmetrics and citations; The number of Mendeley readers was significantly associated with citations	−	+	+	−

[a]1: appropriately adjusting for time of the social media metric (+); 2: appropriately adjusting for confounders such as article type (+) and seasonality/time factors (++); 3: appropriately exploring correlations by including scatterplots (+); 4: appropriately reporting nonlinear correlations tests and statistics (+) as well as log-linear relationship tests (++).

[b]WoS: Web of Science.

[c]IF: impact factor.

[d]JCR: Journal Citation Reports.

[e]AAS: Altmetrics attention score.

[f]MEDLINE: Medical Literature Analysis and Retrieval System Online.

[g]EMBASE: Excerpta Medica database.

XSL•FO
RenderX

Table 5. Characteristics of correlational studies of poor quality (n=12).

Study ID and reference	Health research area/unit and period of analysis	Metrics reported	Results	Methodological quality indicators[a]			
				1	2	3	4
Amath, 2017 [32,33]	Medical education: 482 articles appearing on Medical Education journal (2012-2013)	Social media: Twitter, Mendeley; Altmetrics-Bibliometrics: Citations (Scopus)	*Very strong* correlation between Tweet counts and Altmetrics score; Citations were *strongly* correlated with access counts and Mendeley downloads, and *weakly* and *moderately* correlated respectively with Twitter mentions and Altmetric scores	-	-	+	-
Azer, 2019 [55]	Medical professionalism; 50 most-cited articles in medical professionalism identified by searching WoS[b] (1994-2011)	Social media: Altmetrics-Bibliometrics: Citations (WoS[b])	No significant correlation between Altmetrics and citations	–	+	–	–
Baan, 2017 [56]	Transplantation; All articles published on transplantation in 2015 (volume 99)	Social media: Twitter-Bibliometrics: number of views and downloads	Significant correlation between downloads and Twitter activity	–	–	–	+
Batooli, 2016 [57]	Medical sciences; 533 articles published by faculty at Kashan University of Medical Sciences (1997-2014)	Social media: ResearchGate, Mendeley-Bibliometrics: Citations (Scopus)	Positive correlation between the number of views of articles in ResearchGate and citations; positive correlation between reading frequency in Mendeley and citations; number of views of articles in ResearchGate correlated with higher reading frequency in Mendeley and citations	–	–	–	+
Chen, 2019 [41]	Rheumatology; 1460 articles appearing in *Rheumatology* journal (2010-2015)	Social media: Altmetrics-Bibliometrics: Citations and downloads	*Strong* correlations between Altmetric and downloads, but not citations	–	+	–	–
Chiang, 2016 [42]	Gastroenterology; 1671 articles appearing on 5 core gastroenterology journals, 482 being tweeted (2012)	Social media: Twitter-Bibliometrics: Citations (Google Scholar)	No significant correlation between Twitter and citations	–	+	–	–
Cho, 2017 [60]	Medical sciences; 98 articles from medical sciences from Korean researchers in Scopus (2010-2014)	Social media: ImpactStory; Altmetrics-Bibliometrics: Citations (Scopus)	The more the papers are cited in the journal, the more papers saved on Mendeley	–	+	–	–
Hayon, 2019 [34]	Urology; 213 articles from 7 prominent urology journals (2014-2015)	Social media: Altmetrics-Bibliometrics: Citations (Google Scholar and Scopus)	Positive relationship between Twitter activity and Scopus citations	+	–	–	–
Jedhav, 2019 [68]	Neurointerventional surgery; 451 articles first published on the web on the *Journal of Neurointerventional Surgery* (2015-2016)	Social media: Twitter-Bibliometrics: Citations (WoS[b])	The level of evidence of the publication and the topic of research strongly predicts future citations. The number of clicks also appears to be a strong predictor of future citations, and the number of clicks increases as the number of Twitter users also grows	–	–	+	–
Jeong, 2019 [69]	Coloproctology; 404 articles published on 3 journals with Twitter profiles (2015-2016)	Social media: Twitter-Bibliometrics: Citations (WoS[b])	Significant correlations between citations and Twitter activity	–	+	–	–
Konstantiniuk, 2015 [44]	Sepsis research; 12 articles on sepsis compared with 8 articles on ICU[c] (period not indicated)	Social media: Twittter; Altmetrics; ResearchGate-Bibliometrics: Citations (Google Scholar and WoS[b])	The Altmetric score neither correlated with Google Citations nor publishing date	–	+	–	–
Shirazi, 2018 [79]	Health literacy; 615 articles with a digital object identifier and indexed in WoS[b] (2015)	Social media: Altmetrics-Bibliometrics: Citations (WoS[b])	Significant correlations between Altmetrics and citations	–	–	–	+

XSL•FO
RenderX

[a]1: appropriately adjusting for time of the social media metric (+); 2: appropriately adjusting for confounders such as article type (+) and seasonality/time factors (++); 3: appropriately exploring correlations by including scatterplots (+); 4: appropriately reporting nonlinear correlations tests and statistics (+) as well as log-linear relationship tests (++).

[b]WoS: Web of Science.

[c]ICU: Intensive Care Unit.

Table 6. Characteristics of correlational studies of very poor quality (n=7).

Study ID and reference	Health research area/unit and period of analysis	Metrics reported	Results	Methodological quality indicators[a]			
				1	2	3	4
Araujo, 2017 [53]	Parkinson disease research; Top 20 articles with highest AAS[b] appearing on the *Journal of Parkinson's Disease* (2013-2016)	Social media: Twitter, Facebook, Altmetrics-Bibliometrics: Citations (Scopus)	Qualitative summary in support of correlation	–	–	–	–
Heydarpour, 2017 [43]	Multiple sclerosis research; 4693 articles on multiple sclerosis retrieved from Altmetric Explorer and PubMed (2016)	Social media: Altmetrics-Bibliometrics: Citations (WoS[c])	*Moderate* correlations between AAS[b] and citations	–	–	–	–
Matava, 2017 [45]	Pediatric anesthesiology; Top 100 articles on pediatric anesthesiology available from Altmetrics Explorer (2016)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	No significant correlation between Altmetrics or Twitter mentions and Citations; The number of Mendeley mentions was significantly associated with citations	–	–	–	–
Ramezani-Pakpour-Langeroudi, 2018 [75]	Clinical medicine; 55 highly cited articles on Thomson Reuters' Essential Science Indicator (2015)	Social media: ResearchGate, Mendeley, Academia, LinkedIn-Bibliometrics: Citations (Scopus)	A positive direct relationship was observed between visibility at social networking sites with citation and h‐index rate	–	–	–	–
Ruan, 2018 [77]	Plastic and reconstructive surgery; 55 most-cited articles published in *Plastic and Reconstructive Surgery* (2014-2015)	Social media: Altmetrics, Mendeley-Bibliometrics: Citations (Scopus)	No significant correlation between Altmetrics and citations; The number of Mendeley mentions was significantly associated with citations	–	–	–	–
Smith, 2019 [80]	Gastrointestinal endoscopy; 2361 original research articles published in *Gastrointestinal Endoscopy* of which 2050 were cited at least once (2010-2016)	Social media: Altmetrics-Bibliometrics: Citations (Scopus)	Significant correlations between tweets and citations	–	–	–	–
Wiehn, 2017 [83]	Medical sciences; 36 Shire-sponsored articles (2016)	Social media: Altmetrics-Bibliometrics: Article downloads and IF[d]	No correlation was observed between Altmetrics score and IF[d], downloads	–	–	–	–

[a]1: appropriately adjusting for time of the social media metric (+); 2: appropriately adjusting for confounders such as article type (+) and seasonality/time factors (++); 3: appropriately exploring correlations by including scatterplots (+); 4: appropriately reporting nonlinear correlations tests and statistics (+) as well as log-linear relationship tests (++).

[b]AAS: Altmetrics attention score.

[c]WoS: Web of Science.

[d]IF: impact factor.

Is There Evidence of the Impact of Social Media?

In this section we elaborated on the findings related to the different areas of research, study type, social media intervention characteristics, metrics assessed, and reported results, as summarized in Table 1.

Areas of Health Research

Impact studies have reported the use of social media interventions to promote original research articles published in academic journals in the subject areas of clinical pain sciences (ie, *PLOS ONE*) [48], cardiology (ie, *Circulation*) [46,47], radiology [52], emergency health [51] public health (ie, *International Journal of Public Health*) [49], and women's health (various journals) [50].

Study Types

We identified 4 randomized controlled experiments, further referred to as *RCTs* [46,47,49,51], and 3 quasi-experimental trials [48,50,52].

Social Media Intervention Characteristics

Articles in the intervention conditions were shared on Twitter or Facebook social media profiles of the targeted journals using automated or manually made posts leveraging the organic (ie, unpaid) reach of each social networking site. One study used

XSL•FO
RenderX

advertising to increase the views of Facebook posts (ie, *boosted* content) to increase reach, which is the number of viewers of the post [47]. Intervention duration ranged from 12 [49], 18 [48], 34 [47], to 52 weeks [46]. The intensity varied considerably: only once the content was blogged [48], 1 post every 2 weeks and 12 weeks [49], 1 or 2 posts per week [46], or several posts per day [47]. Three studies used blogs to diffuse the studies further [48,49,52]. Thoma et al [51] tested the use of a podcast and infographics as complementary information in addition to Facebook and Twitter content sharing. The interventions varied in nature of the message posted, the profile of the social media owner(s), the duration of the posting, its intensity, and whether an incentive or paid promotion was used. For example, in one of the quasi-experimental trials [48], the authors advertised the posts related to 16 original research articles about clinical pain sciences, published on *PLOS ONE*, through a systematic intervention targeting Facebook, Twitter, LinkedIn, and ResearchBlogging.org. The authors used a systematic protocol (ie, timing the frequency of release of the messages) to direct social media users to read a web-based version of the original research article.

Metrics Assessed

These studies investigated the effect of social media interventions on subsequent access to web-based journal content and article downloads. Two studies (1 experimental [49] and 1 quasi-experimental [48]) reported the effect on the number of citations.

Reported Results

Although the quasi-experimental trials reported an increase in HTML and PDF views during the intervention [48,50,52], the RCTs reported no significant changes in the same metrics [46,47,49,51]. There were no significant effects on the number of citations in both the quasi-experiment [48] and the RCT [49].

What Is the Association Between Altmetrics and Bibliometrics in Health Research?

In this section we present the results related to the correlational studies identified.

Areas of Health Research

The 44 correlational studies evaluated the relationship between Altmetrics and bibliometrics in a variety of health research fields and disciplines, covering generic medical or biomedical research disciplines [16,57,60-62,66,70,78,81-83], or more specific disease-related research fields, such as multiple sclerosis [43], neurological research [73], Parkinson disease [53], psoriasis [25], and sepsis [44,67]. Some articles covered clinical or dental medicine [64,65,75] and different kinds of surgery [59,68,77]; others focused on specialized branches of medicine, such as urology [34,38,39,58], radiology [76], and coloproctology [69].

Study Types

Almost all correlational studies were quantitative cross-sectional studies (43/44, 97%) examining the relationship between bibliometrics and various social media metrics or monitored citation trends over time [52,83]. The association between social media use and citations was also discussed in a qualitative study [53].

Metrics Assessed

Correlational studies encompassed a wide variety of Altmetrics and bibliometrics. Some correlational studies have investigated the association between Altmetrics scores and citations in Web of Science, Scopus, or Google Scholar [36,39,53,61,62,66,70,74,76,81-83]. One study investigated the correlation between ImpactStory indices and Scopus citations [60]. Other studies focused on the usage metrics of specific social media platforms, such as Twitter [17,42,56,74], blogs, web-based posts [52,58], ResearchGate, or Mendeley [57]. A total of 4 studies reported data on the association between the Altmetric of articles and the IF of journals in which they were published [38,59,60,83]; 3 studies investigated the relationship between sharing articles on academic social media platforms, such as ResearchGate and reference management software Mendeley [44,57,75].

Methodological Quality

As shown in the last 4 columns of Tables 2 to 6, the methodological quality of the papers varied according to the type of article and amount of detail included in the publication. The 5 conference abstracts included in the list of correlational studies [41-45] did not provide sufficient information to meet most of the methodological quality indicators. Only 3 of these [41,42,44] provided different correlation results according to the type of disease area, topic, or article type; hence, they were deemed to have adjusted for confounders. Of the remaining 39 studies with full-text, 16 adjusted the social media metric for time (16/39, 41%); 22 studies adjusted for confounders such as article type or topic (22/39, 56%), but none included the seasonality and timeframe of the article publication. Some 21 studies included scatterplots to explore the type of correlation (21/39, 54%), and 18 studies reported the use of Spearman rho or other nonparametric tests when comparing social media citation counts (18/39, 46%). However, only 2 studies [74,81] reported analyses for tests based on ranking similar to log-linear correlations included in a study by Eysenbach [17]. The method used in a study by Quintana and Doan [74], but described in the paper by Thelwall et al [81], “compares a given publication against the publications that appeared immediately before and after it. A successful test occurs when the number of Twitter mentions and citations for a given publication are both higher (or both lower) than the average of Twitter mentions and citations of the two adjacent articles” [81].

A total of 6 of 39 studies (15%) [16,36,61,64,71,72] met the minimum standard in all 4 criteria mentioned above, but none followed exemplar paper by Eysenbach [17] when reporting correlations between social media metrics and citations. In all, 7 studies scored at least three criteria (7/39, 18%). Of these studies, 4 scored criteria 2-3-4 [40,66,76,78], 2 studies scored 1-2-4 [62,82], and 1 study 1-2-3 [67]. The remaining 18 studies scored less than 2 of any quality criteria (18/39, 46%).

Reported Results

Twitter was the most popular social media platform discussed in correlational studies (21 of 44 studies)

XSL•FO
RenderX

[16,17,32,34,38,42,45,53,55,56,59,62,63,68,69,72,74,76,78-80], followed by Mendeley (15 studies) [25,45,55,57,60,63,71,72,75-79,82,83]. In reference to the association between Altmetric scores and bibliometrics, the results were mixed. No significant correlations were reported in 12 of 44 studies (27%) [25,41,42,44,45,55,64-66,71,77,83]; *weak* or *moderate* correlations were reported by authors in 13 of 44 studies (30%) [16,17,39,43,53,61-63,67,72-74,76]. *Strong* positive associations were reported by the authors of half (22) of the correlational studies. In particular, strong associations were found between Mendeley readership (reads and downloads) and subsequent citations [25,45,57,60,75,77-79,82]. This is also consistent with the findings reported in one of the impact studies [49], which found significant correlations between the number of downloads of a paper and the subsequent number of citations (r=0.52; *P*<.001), which were larger in the intervention group (r=0.67; *P*<.001). Only 2 studies reported no significant associations between traditional and social media metrics [66] and between Altmetric score and journal IF, ResearchGate reads, and the number of article downloads [83].

When focusing only on the 7 high quality studies (ie, those meeting all methodological quality criteria 1-4) [16,17,36,61,64,71,72], 5 studies reported correlation coefficients of moderate size [16,17,36,61,72] and 2 studies, from the same discipline (orthodontics), reported no significant correlations [64,71].

Discussion

Is There Evidence of the Impact of Social Media?

The primary aim of this review was to evaluate the effect of social media on the dissemination of health research output. Of the 51 identified studies, only 7 were experimental studies aimed at establishing a causal link between social media use and subsequent citations. The identified impact studies provide suggestive yet inconclusive evidence on the effect of using social media to increase the number of citations, thus contributing to the dissemination of health research according to traditional bibliometric indicators. This result is consistent with findings reported in a systematic review, which was aimed at describing the interactions between bibliometric factors and social media activity on the impact of reviews in the field of psoriasis [25]. The findings suggest that although Google Scholar citations were predicted by the number of readers on Mendeley and year of publication (ie, 2015 and 2016), the number of tweets and the IF of a journal were not. Moreover, a journal's IF was the sole significant predictor of the number of tweets [25]. Careful considerations should be made, as the limited number of studies we identified does not allow us to generate strong conclusions or recommendations.

In our review, we identified four impact studies [46,47,49,51], which used randomized controlled experiments. However, unlike in the TrendMD dissemination study [19], these RCTs did not show consistent effects of the social media dissemination strategies on citations over time. This finding might be due to several limitations. First, the experiments used *social media interventions*, which had very different intensities (eg, once vs several posts per day) and duration (ie, 12-52 weeks). This fact does not allow us to determine whether the effect was due to a *dose-response* or to mere exposure to the intervention. Consequently, the long-term effects of social media interventions are still unclear as the only long-term study (52 weeks) [46] reported no difference in median 30-day page views between the intervention and control conditions. However, a larger, longitudinal RCT on TrendMD distribution showed higher citation counts in health and medical science articles after 6 and 12 months than in a control group [19]. Second, the reported social media impact outcomes differed in terms of content and type of social media channel used, frequency and intensity of social media use, type of exposure, and unit of analysis. It is important to note that careful consideration should be made when comparing social media interventions across the spectrum of platforms and types of social networking sites. A journal's or author's Facebook page or Twitter handle that originally has 100,000 followers would very likely increase the reach of an article shared, compared with a page that has only 1000 followers. This finding could have implications on the effect of increased citations in journals that already have a high IF compared with journals that have a low IF (as the accounts of the former would have a higher number of followers). In other words, the *social media interventions* were too heterogeneous to compare among each another and to perform meta-analyses.

Future impact studies should maintain a rigorous study design; consistently report social media outcomes using standard Altmetric scores; provide better and more detailed explanations about the specific timeframes, exposure, frequency, and intensity of interventions for comparability. In other words, future studies should answer research questions such as to what extent does the frequency of social media posting influence short-term indicators (eg, number of PDF downloads) and long-term indicators (eg, Altmetric score and citations)? To what extent does posting on Mendeley and Twitter, as opposed to Twitter, have an impact on the Altmetric score and citations?

This information could help researchers specialized in systematic reviews to develop accurate evidence, including meta-analyses; the information could also be helpful for researchers aimed at testing different social media intervention strategies or at comparing similar methodologies in different domains or disciplines. Other factors that can explain the lack of findings are in the topics of the information shared on social media. Some very specific health disciplines have limited readership, as they require specific knowledge to understand the content that is shared.

What Is the Association Between Altmetrics and Bibliometrics in Health Research?

Another finding of this review is that most of the available evidence focuses on describing correlations between traditional and social media metrics in health research. The included 44 correlational studies provide further support that, in general, the higher the AAS, the higher the subsequent citations will be. However, the studies reported wide variability in the magnitude of correlation coefficients and provided a variety of interpretations for the strength of these correlations, which warns some caution. Most reports did not provide in-depth evaluations

XSL•FO
RenderX

of the correlations, including, for example, confidence intervals of the correlation coefficients; an analysis of the distribution of citation counts and social media metrics, which tend to violate the assumptions of normality; and the use of visual representations such as scatterplots, as recommended by some researchers [84,85]. We identified only 6 correlational studies [16,36,61,64,71,72] meeting the minimum methodological quality criteria (1-4) described in the methodology, in addition to the seminal paper by Eysenbach, 2011 [17]. Most notably and quite surprisingly, none of the identified 44 studies followed the recommendations and metrics suggested by Eysenbach. No paper included the Twindex or the Twimpact factor or correctly adjusted the social media metrics for the time since publication or seasonality in publication and for the skewed distribution of the metric, for example, by dichotomizing result or testing for log-linear correlations.

Twitter and Mendeley seem to be the indicators that contribute the most to the Altmetric score. Mendeley and ResearchGate were positively associated with subsequent citations [25,45,77,79]. Although Twitter can be used to disseminate research output among a broader, general public, the use of Mendeley and ResearchGate seems to be restricted to specific target audience of researchers or media professionals. Unsurprisingly, the more an article is shared on ResearchGate or Mendeley, the more it will be cited. This finding might indicate that researchers use reference management software (Mendeley) to organize their libraries for research purposes and then share their research on a specialized social networking site (ResearchGate). This software allows users to discover new *related research* because the platform itself suggests new evidence based on the users' previous reads and mentions. Although this does not imply that research has been disseminated among the wider public, ResearchGate and Mendeley appear fundamental for the research community.

Our findings are consistent with some existing review evidence investigating the domain of medical research output [24], as well as other multidisciplinary research fields [21,22], or marketing [20]. Another review that focused on orthopedic research journals [86] reported that web-based mentions were weakly yet positively related to various bibliometric indices, such as the number of citations, journal IF, Eigenfactors, and h-index values for the first and last authors. In addition, a systematic review of reviews published in scientific journals related to skin psoriasis [25] found an association between Altmetrics and bibliometric indicators. The association between traditional bibliometrics (eg, number of citations) and social media metrics (eg, number of mentions) could be an indicator of a positive effect of using social media on research dissemination. However, not all correlational articles identified showed strong positive correlations. The association could be confounded by several factors, such as the value of the research paper or the popularity of the topic, as we have previously discussed. In other words, a high-value research paper could lead to both high exposure on social media and a high number of citations, depending on the intrinsic subject. An article discussing skin psoriasis, or another more common disease, might attract the attention of the wider general public, compared with coloproctology and neuro-interventional surgery, which attract audiences of specialized health care professionals. Moreover, researchers can discover articles to cite from other *traditional* sources, such as electronic databases and libraries. Once an article is published and cited, other researchers might want to share the discovery on their social networking sites. Hence, using social media to diffuse health research may generate a virtuous circle that can be beneficial for both individual researchers and journals, as this will result in an increased IF. We need more evidence of good quality experimental, *impact studies* rather than correlational studies to establish a causal link between social media use and impact on citations. We also need better reporting of correlational studies; following the suggestions included in the seminal paper by Eysenbach [17], researchers should strive to provide more insight (and data) on the actual distribution of bibliometrics and social media metrics analyzed so that more meaningful interpretations of the relationship between these variables could be drawn.

Recommendations

Scientific impact is a multidimensional concept that cannot be adequately measured by a single indicator [87]. As the AAS suggests, social media impact is even more multidimensional, as it is linked to various web-based tools that a particular researcher can use. This researcher is also embedded in a unit within an institution [88]; the researcher's work is then published in a journal, which has a specific and independent impact (IF, citations, and Altmetrics). The evaluation of the researcher's scientific impact on social media should then take into account various dimensions and indicators and be aligned with more qualitative evaluation on other domains of the RIF [4], which includes policy, service, and societal impacts.

In conclusion, is social media dissemination worth the effort? On the basis of the findings of this review, we recommend researchers in health sciences to continue using social media to disseminate their research, as there is some data suggesting its long-term impact on citations following dissemination on the internet using paid services (eg, TrendMD study) [19]. Researchers should use popular social networking platforms, such as Twitter or Facebook, first to engage with the general public and the media, to design more citizen-oriented research and democratize their findings. Researchers should also use popular social networking sites to interact with peers and discuss their research as well as disseminate their findings. Researchers should also diffuse their work on specialized social networking sites for academics, such as Mendeley, ResearchGate, and Academia. The content shared on these networks is very likely to be cited, as researchers who are on Mendeley may use it as a reference management software. Mendeley and ResearchGate databases may be used in combination with traditional electronic databases for literature reviews and similar activities. Storing copies of articles on these platforms could increase the chances of a paper being cited in the future.

Although more rigorous longitudinal research needs to establish whether social media activity can be linked to increased citations, it is important to consider social media as useful tools to reach a wider public, not just specialized audiences. Researchers should put effort into translating knowledge for

XSL•FO
RenderX

different target audiences, bearing in mind the users of each social media channel. The AAS can be a useful instrument that allows researchers and institutions to evaluate social media impact by distinguishing among the attention generated by specialized applications (Mendeley and ResearchGate) or by the wider public and traditional media (Facebook, Twitter, and blogs).

Strengths and Limitations

This review sheds light on the use of social media to disseminate health research output. The main strength of this review is the use of standard methods of a systematic review, including a comprehensive search strategy, a duplicate approach to study selection and data abstraction, and detailed data abstraction. The main limitation is that we were not able to conduct a meta-analysis because of the substantial variability in the included studies. For example, the impact studies varied in terms of the interventions evaluated (eg, type of social media used, the message posted, and its duration and intensity) and the outcomes assessed.

Conclusions

Our findings have implications for research in the field of health-related metrics. There is a need for more and better designed experimental studies testing the use of social media to increase the dissemination of health research. These studies should be of a randomized design, evaluate the appropriate use of social media, and assess a variety of outcomes (both all Altmetrics dimensions and traditional bibliometrics) over a meaningfully long period. For example, experimental trials could test different strategies to diffuse research articles on social media, by comparing paid (or *boosted*) content with nonpaid, *organic* posts. Other trials could include the diffusion of research article links on numerous social networking sites versus a limited number of sites. Once there is evidence of the effectiveness of using social media (as opposed to not using them), different approaches should be investigated.

Our findings also have implications for the practice of using social media for research dissemination. Researchers should not use social media for the sole purpose of increasing their research productivity (ie, number of publications), as there is currently no evidence to support such an effect. They can use it for other purposes, such as disseminating their findings to social media users. When using social media, they have to be attentive to details such as the content of the message, its frequency, and the use of incentives or paid promotions, as this could affect the reach of the posts.

Acknowledgments

The authors would like to acknowledge Ms Aida Farha from AUB libraries for her help in developing the search strategy and Ms Layal Hneiny, who performed the electronic database search updates. The authors would also like to thank Mr Fadel Al Turki, Dr Sanaa Baddour, Ms Maya Hammoud, and Ms Matilda Sang for their help in reviewing the abstracts.

Authors' Contributions

EAA, RER, MB, LIM, and SB conceived and designed the review, and EAA, RER, and MB coordinated it. EAA, RER, MB, and LIM were involved in developing the search strategy, and MC, EWA, ASR, and MB extracted the data. Data analyses were undertaken by EAA, RER, and MB, whereas data interpretation was done by EAA, RER, and MB. RER, EAA, and MB drafted the review. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 18 KB - jmir_v22i7e15607_app1.docx]

Multimedia Appendix 2
Records excluded at the full-text screening phase with full citation.
[DOCX File , 80 KB - jmir_v22i7e15607_app2.docx]

Multimedia Appendix 3
Detailed characteristics of included studies.
[DOCX File , 93 KB - jmir_v22i7e15607_app3.docx]

References

1. Carr C, Hayes R. Social media: defining, developing, and divining. Atl J Commun 2015 Feb 6;23(1):46-65. [doi: 10.1080/15456870.2015.972282]
2. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health researchers' use of social media: scoping review. J Med Internet Res 2019 Nov 13;21(11):e13687 [FREE Full text] [doi: 10.2196/13687] [Medline: 31719028]

XSL•FO
RenderX

3. O'Reilly T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'Reilly Media Inc. 2005. URL: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html [accessed 2016-11-28]
4. Kuruvilla S, Mays N, Pleasant A, Walt G. Describing the impact of health research: a research impact framework. BMC Health Serv Res 2006 Oct 18;6:134 [FREE Full text] [doi: 10.1186/1472-6963-6-134] [Medline: 17049092]
5. Davidson PM, Newton PJ, Ferguson C, Daly J, Elliott D, Homer C, et al. Rating and ranking the role of bibliometrics and webometrics in nursing and midwifery. ScientificWorldJournal 2014;2014:135812 [FREE Full text] [doi: 10.1155/2014/135812] [Medline: 24550691]
6. Rivera SC, Kyte DG, Aiyegbusi OL, Keeley TJ, Calvert MJ. Assessing the impact of healthcare research: a systematic review of methodological frameworks. PLoS Med 2017 Aug;14(8):e1002370 [FREE Full text] [doi: 10.1371/journal.pmed.1002370] [Medline: 28792957]
7. Scarlat MM, Mavrogenis AF, Pećina M, Niculescu M. Impact and alternative metrics for medical publishing: our experience with international orthopaedics. Int Orthop 2015 Aug;39(8):1459-1464. [doi: 10.1007/s00264-015-2766-y] [Medline: 25947897]
8. Gruzd A, Staves K, Wilk A. Tenure and promotion in the age of online social media. Proc Am Soc Info Sci Tech 2012 Jan 11;48(1):1-9 [FREE Full text] [doi: 10.1002/meet.2011.14504801154]
9. Carrigan M. Social Media for Academics. First Edition. Thousand Oaks, CA: Sage Publications; 2016.
10. Constantinides E, Fountain SJ. Web 2.0: conceptual foundations and marketing issues. J Direct Data Digit Mark Pract 2008 Jan 4;9(3):231-244. [doi: 10.1057/palgrave.dddmp.4350098]
11. Miah A. The A to Z of Social Media for Academia. Academic & University News: Times Higher Education. 2017. URL: https://www.timeshighereducation.com/a-z-social-media [accessed 2019-07-03]
12. Champieux R. PlumX. J Med Libr Assoc 2015 Jan;103(1):63-64 [FREE Full text] [doi: 10.3163/1536-5050.103.1.019]
13. Thelwall M. A history of webometrics. Bul Am Soc Info Sci Tech 2012 Aug 21;38(6):18-23. [doi: 10.1002/bult.2012.1720380606]
14. Priem J, Taraborelli D, Groth P, Neylon C. Altmetrics: A Manifesto. Altmetrics. 2011. URL: http://altmetrics.org/manifesto [accessed 2020-05-25]
15. Brigham TJ. An introduction to altmetrics. Med Ref Serv Q 2014;33(4):438-447. [doi: 10.1080/02763869.2014.957093] [Medline: 25316077]
16. Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. J Assn Inf Sci Tec 2013 Nov 26;65(4):656-669. [doi: 10.1002/asi.23101]
17. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
18. Niyazov Y, Vogel C, Price R, Lund B, Judd D, Akil A, et al. Open access meets discoverability: citations to articles posted to academia.edu. PLoS One 2016;11(2):e0148257 [FREE Full text] [doi: 10.1371/journal.pone.0148257] [Medline: 26886730]
19. Kudlow P, Dziadyk DB, Rutledge A, Shachak A, Eysenbach G. The citation advantage of promoted articles in a cross‐publisher distribution platform: a 12‐month randomized controlled trial. J Assoc Inf Sci Technol 2019 Dec 23;1:18. [doi: 10.1002/asi.24330]
20. Misirlis N, Vlachopoulou M. Social media metrics and analytics in marketing – S3M: a mapping literature review. Int J Inf Manag 2018 Feb;38(1):270-276. [doi: 10.1016/j.ijinfomgt.2017.10.005]
21. Sugimoto CR, Work S, Larivière V, Haustein S. Scholarly use of social media and altmetrics: a review of the literature. J Assoc Inf Sci Technol 2017 Jun 22;68(9):2037-2062. [doi: 10.1002/asi.23833]
22. Karanatsiou D, Misirlis N, Vlachopoulou M. Bibliometrics and altmetrics literature review. Perform Meas Metr 2017 Apr 10;18(1):16-27. [doi: 10.1108/PMM-08-2016-0036]
23. El Rassi R, Elie A, Bardus M, Meho L, Farha A. The Use of Social Media for Disseminating Health Research. University of York. 2017 Feb 20. URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017057709 [accessed 2020-05-25]
24. Patthi B, Prasad M, Gupta R, Singla A, Kumar JK, Dhama K, et al. Altmetrics - a collated adjunct beyond citations for scholarly impact: a systematic review. J Clin Diagn Res 2017 Jun;11(6):ZE16-ZE20 [FREE Full text] [doi: 10.7860/JCDR/2017/26153.10078] [Medline: 28764311]
25. Ruano J, Aguilar-Luque M, Gómez-Garcia F, Mellado PA, Gay-Mimbrera J, Carmona-Fernandez PJ, et al. The differential impact of scientific quality, bibliometric factors, and social media activity on the influence of systematic reviews and meta-analyses about psoriasis. PLoS One 2018;13(1):e0191124 [FREE Full text] [doi: 10.1371/journal.pone.0191124] [Medline: 29377889]
26. Nass S, Levit L, Gostin L. The value, importance, and oversight of health research. In: Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: The National Academies Press; 2009.
27. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]

XSL•FO
RenderX

28. Bardus M. The web 2.0 and social media technologies for pervasive health communication: are they effective? Stud Commun Sci 2011;11(1):136 [FREE Full text] [doi: 10.5169/seals-791189]
29. Gwet K. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. Stat Methods Inter-Rater Reliab Assess Ser 2002 May;2:1-9. [doi: 10.1007/springerreference_183942]
30. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Simcoe St, Toronto: Advanced Analytics; 2014.
31. Wilson DB. Practical Meta-Analysis Effect Size Calculator. The Campbell Collaboration. URL: http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD5.php [accessed 2019-07-06]
32. Amath A, Ambacher K, Leddy JJ, Wood TJ, Ramnanan CJ. Comparing alternative and traditional dissemination metrics in medical education. Med Educ 2017 Sep;51(9):935-941. [doi: 10.1111/medu.13359] [Medline: 28719136]
33. Ramnanan CJ, Ambacher K, Amanth A, O'Brien C, Wood TJ, Leddy JJ. How do alternative and traditional dissemination metrics compare in medical education scholarship? FASEB J 2016;30(1_supplement):568.1 [FREE Full text] [doi: 10.1096/fasebj.30.1_supplement.568.1]
34. Hayon S, Tripathi H, Stormont IM, Dunne MM, Naslund MJ, Siddiqui MM. Twitter mentions and academic citations in the urologic literature. Urology 2019 Jan;123:28-33. [doi: 10.1016/j.urology.2018.08.041] [Medline: 30278190]
35. Knight S. Social media and online attention as an early measure of the impact of research in solid organ transplantation. Transplantation 2014 Sep 15;98(5):490-496. [doi: 10.1097/TP.0000000000000307] [Medline: 25061904]
36. Knight SR. Social media and online attention as an early measure of the impact of research in solid organ transplantation. Transplantation 2014 Sep 15;98(5):490-496. [doi: 10.1097/TP.0000000000000307] [Medline: 25061904]
37. Nguyen KG, Gross C, Cooperberg M, Katz M, Hittelman A, Syed J, et al. MP69-07 hashtag peer-review: does early social media success correlate with conventional metrics of publication impact? J Urology 2017 Apr;197(4S):e928-e929. [doi: 10.1016/j.juro.2017.02.2303]
38. Nolte AC, Nguyen KA, Perecman A, Katz MS, Kenney PA, Cooperberg MR, et al. Association between twitter reception at a national urology conference and future publication status. Eur Urol Focus 2019 May 15:S2405-S4569. [doi: 10.1016/j.euf.2019.05.004] [Medline: 31103604]
39. O'Connor EM, Nason GJ, O'Kelly F, Manecksha RP, Loeb S. Newsworthiness vs scientific impact: are the most highly cited urology papers the most widely disseminated in the media? BJU Int 2017 Sep;120(3):441-454 [FREE Full text] [doi: 10.1111/bju.13881] [Medline: 28418091]
40. O'Connor EM, Nason GJ, O'Kelly F, Manecksha RP, Loeb S. Newsworthiness vs scientific impact: are the most highly cited urology papers the most widely disseminated in the media? BJU Int 2017 Sep 18;120(3):441-454 [FREE Full text] [doi: 10.1111/bju.13881] [Medline: 28418091]
41. Chen W, Cocskhull F, Bukhari M, Galloway J. If I tweet more about my paper, will I improve my citation index? Rheumatology 2019 Apr 12;58(3):kez107.030. [doi: 10.1093/rheumatology/kez107.030]
42. Chiang AL, Rabinowitz LG, Alakbarli J, Chan WW. Tu1108 social media exposure is independently associated with increased citations of publications in gastroenterology. Gastroenterology 2016 Apr;150(4):S845. [doi: 10.1016/S0016-5085(16)32855-4]
43. Heydarpour P, Shirazi A, Sahraian M. Poster session 2: P888 multiple sclerosis research dissemination in the web: news, blogs, or tweets? Mult Scler 2017 Oct;23(3_suppl):427-679 [FREE Full text] [doi: 10.1177/1352458517731406]
44. Konstantiniuk A, Amrein K. Scientific papers in the social media. Wien Klin Wochenschau 2015 Oct;127(19):815-816 [FREE Full text] [doi: 10.1007/s00508-015-0873-9]
45. Matava C, Salman M. The impact of social media on pediatric anesthesia papers and bibliometry. Anesth Analg 2017 May;124(1):S152 [FREE Full text]
46. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from circulation. Circulation 2015 Jan 6;131(1):28-33. [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
47. Fox CS, Gurary EB, Ryan J, Bonaca M, Barry K, Loscalzo J, et al. Randomized controlled trial of social media: effect of increased intensity of the intervention. J Am Heart Assoc 2016 Apr 27;5(5):- [FREE Full text] [doi: 10.1161/JAHA.115.003088] [Medline: 27121850]
48. Allen HG, Stanton TR, di Pietro F, Moseley GL. Social media release increases dissemination of original articles in the clinical pain sciences. PLoS One 2013;8(7):e68914 [FREE Full text] [doi: 10.1371/journal.pone.0068914] [Medline: 23874810]
49. Tonia T, van Oyen H, Berger A, Schindler C, Künzli N. If I tweet will you cite? The effect of social media exposure of articles on downloads and citations. Int J Public Health 2016 May;61(4):513-520. [doi: 10.1007/s00038-016-0831-y] [Medline: 27193574]
50. Cawcutt KA, Erdahl LM, Englander MJ, Radford DM, Oxentenko AS, Girgis L, et al. Use of a coordinated social media strategy to improve dissemination of research and collect solutions related to workforce gender equity. J Womens Health (Larchmt) 2019 Jun;28(6):849-862. [doi: 10.1089/jwh.2018.7515] [Medline: 30998087]
51. Thoma B, Murray H, Huang SY, Milne WK, Martin LJ, Bond CM, et al. The impact of social media promotion with infographics and podcasts on research dissemination and readership. CJEM 2018 Mar;20(2):300-306. [doi: 10.1017/cem.2017.394] [Medline: 28899440]

XSL•FO
RenderX

52. Hoang JK, McCall J, Dixon AF, Fitzgerald RT, Gaillard F. Using social media to share your radiology research: how effective is a blog post? J Am Coll Radiol 2015 Jul;12(7):760-765. [doi: 10.1016/j.jacr.2015.03.048] [Medline: 25959491]
53. Araújo R, Sorensen AA, Konkiel S, Bloem BR. Top altmetric scores in the Parkinson's disease literature. J Parkinsons Dis 2017;7(1):81-87 [FREE Full text] [doi: 10.3233/JPD-179000] [Medline: 28222540]
54. Araujo AC, Nascimento DP, Gonzalez GZ, Maher CG, Costa LO. Impact of low back pain clinical trials measured by the altmetric score: cross-sectional study. J Med Internet Res 2018 Apr 5;20(4):e86 [FREE Full text] [doi: 10.2196/jmir.9368] [Medline: 29622526]
55. Azer SA, Azer S. Top-cited articles in medical professionalism: a bibliometric analysis versus altmetric scores. BMJ Open 2019 Jul 31;9(7):e029433 [FREE Full text] [doi: 10.1136/bmjopen-2019-029433] [Medline: 31371297]
56. Baan CC, Dor FJ. The transplantation journal on social media: the transplantjrnl journey from impact factor to Klout score. Transplantation 2017 Jan;101(1):8-10. [doi: 10.1097/TP.0000000000001581] [Medline: 27906823]
57. Batooli Z, Ravandi SN, Bidgoli MS. Evaluation of scientific outputs of Kashan university of medical sciences in Scopus citation database based on Scopus, ResearchGate, and Mendeley scientometric measures. Electron Physician 2016 Feb;8(2):2048-2056 [FREE Full text] [doi: 10.19082/2048] [Medline: 27054017]
58. Calopedos RJ, Garcia C, Rashid P, Murphy DG, Lawrentschuk N, Woo HH. Citation indices for social media articles in urology. BJU Int 2017 May;119(Suppl 5):47-52 [FREE Full text] [doi: 10.1111/bju.13872] [Medline: 28544295]
59. Chang J, Desai N, Gosain A. Correlation between altmetric score and citations in pediatric surgery core journals. J Surg Res 2019 Nov;243:52-58. [doi: 10.1016/j.jss.2019.05.010] [Medline: 31154133]
60. Cho J. A comparative study of the impact of Korean research articles in four academic fields using altmetrics. Perform Meas Metr 2017 Apr 10;18(1):38-51. [doi: 10.1108/PMM-02-2016-0005]
61. Costas R, Zahedi Z, Wouters P. Do 'altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. J Assn Inf Sci Tec 2014 Jul 28;66(10):2003-2019. [doi: 10.1002/asi.23309]
62. Dal-Ré R, Mahillo-Fernández I. Are article's social media mentions associated to citation counts? An analysis in highly influential medical journals. Rev Clin Esp 2018;218(1):40-42. [doi: 10.1016/j.rce.2017.04.003] [Medline: 28571841]
63. Dardas LA, Woodward A, Scott J, Xu H, Sawair FA. Measuring the social impact of nursing research: an insight into altmetrics. J Adv Nurs 2019 Jul;75(7):1394-1405. [doi: 10.1111/jan.13921] [Medline: 30507052]
64. Delli K, Livas C, Spijkervet F, Vissink A. Measuring the social impact of dental research: an insight into the most influential articles on the web. Oral Dis 2017 Nov;23(8):1155-1161. [doi: 10.1111/odi.12714] [Medline: 28734099]
65. Hassona Y, Qutachi T, Dardas L, Alrashdan MS, Sawair F. The online attention to oral cancer research: an altmetric analysis. Oral Dis 2019 Sep;25(6):1502-1510. [doi: 10.1111/odi.13111] [Medline: 31033085]
66. Haustein S, Costas R, Larivière V. Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. PLoS One 2015;10(3):e0120495 [FREE Full text] [doi: 10.1371/journal.pone.0120495] [Medline: 25780916]
67. Jabaley C, Groff RT, Stentz M, Moll V, Lynde G, Blum J, et al. Highly visible sepsis publications from 2012 to 2017: analysis and comparison of altmetrics and bibliometrics. J Crit Care 2018 Dec;48:357-371. [doi: 10.1016/j.jcrc.2018.09.033] [Medline: 30296750]
68. Jadhav AP, Ducruet AF, de Leacy R, Fargen KM. Social media and predictors of traditional citations: insights from the journal of neurointerventional surgery. J Neurointerv Surg 2019 Feb;11(2):99-100. [doi: 10.1136/neurintsurg-2018-014689] [Medline: 30665984]
69. Jeong JW, Kim MJ, Oh H, Jeong S, Kim MH, Cho JR, et al. The impact of social media on citation rates in coloproctology. Colorectal Dis 2019 Oct;21(10):1175-1182. [doi: 10.1111/codi.14719] [Medline: 31124259]
70. Liu CL, Xu YQ, Wu H, Chen SS, Guo JJ. Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: dimensions and structure. J Med Internet Res 2013 Nov 25;15(11):e259 [FREE Full text] [doi: 10.2196/jmir.2707] [Medline: 24275693]
71. Livas C, Delli K. Looking beyond traditional metrics in orthodontics: an altmetric study on the most discussed articles on the web. Eur J Orthod 2018 Apr 6;40(2):193-199. [doi: 10.1093/ejo/cjx050] [Medline: 29016742]
72. Maggio LA, Leroux TC, Meyer HS, Artino AR. #MedEd: exploring the relationship between altmetrics and traditional measures of dissemination in health professions education. Perspect Med Educ 2018 Aug;7(4):239-247 [FREE Full text] [doi: 10.1007/s40037-018-0438-5] [Medline: 29949099]
73. Punia V, Aggarwal V, Honomichl R, Rayi A. Comparison of attention for neurological research on social media vs academia: an altmetric score analysis. JAMA Neurol 2019 Jul 1:1122-1124 epub ahead of print. [doi: 10.1001/jamaneurol.2019.1791] [Medline: 31260003]
74. Quintana DS, Doan NT. Twitter article mentions and citations: an exploratory analysis of publications in the American journal of psychiatry. Am J Psychiatry 2016 Feb 1;173(2):194. [doi: 10.1176/appi.ajp.2015.15101341] [Medline: 26844798]
75. Ramezani-Pakpour-Langeroudi F, Okhovati M, Talebian A. Do highly cited clinicians get more citations when being present at social networking sites? J Educ Health Promot 2018;7:18 [FREE Full text] [doi: 10.4103/jehp.jehp_69_17] [Medline: 29629379]
76. Rosenkrantz AB, Ayoola A, Singh K, Duszak R. Alternative metrics ('altmetrics') for assessing article impact in popular general radiology journals. Acad Radiol 2017 Jul;24(7):891-897. [doi: 10.1016/j.acra.2016.11.019] [Medline: 28256440]

XSL•FO
RenderX

77. Ruan QZ, Chen AD, Cohen JB, Singhal D, Lin SJ, Lee BT. Alternative metrics of scholarly output: the relationship among altmetric score, mendeley reader score, citations, and downloads in plastic and reconstructive surgery. Plast Reconstr Surg 2018 Mar;141(3):801-809. [doi: 10.1097/PRS.0000000000004128] [Medline: 29481413]
78. Scotti V, de Silvestri A, Scudeller L, Abele P, Topuz F, Curti M. Novel bibliometric scores for evaluating research quality and output: a correlation study with established indexes. Int J Biol Markers 2016 Dec 23;31(4):e451-e455. [doi: 10.5301/jbm.5000217] [Medline: 27312588]
79. Shirazi M, Goltaji M. An altmetric study on scientific articles of 'health literacy' in social media. Payesh Health Monit 2018;17(3):249-256 [FREE Full text]
80. Smith ZL, Chiang AL, Bowman D, Wallace MB. Longitudinal relationship between social media activity and article citations in the journal gastrointestinal endoscopy. Gastrointest Endosc 2019 Jul;90(1):77-83. [doi: 10.1016/j.gie.2019.03.028] [Medline: 30935934]
81. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One 2013;8(5):e64841 [FREE Full text] [doi: 10.1371/journal.pone.0064841] [Medline: 23724101]
82. Thelwall M, Wilson P. Mendeley readership altmetrics for medical articles: an analysis of 45 fields. J Assn Inf Sci Tec 2015 May 5;67(8):1962-1972. [doi: 10.1002/asi.23501]
83. Wiehn J, Southam E. Monitoring the social-media footprint of newly published articles to assess short-term impact. Curr Med Res Opin 2017 May;33(sup1):i-iv. [doi: 10.1080/03007995.2017.1296267]
84. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
85. Mukaka M. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
86. Evaniew N, Adili AF, Ghert M, Khan M, Madden K, Smith C, et al. The scholarly influence of orthopaedic research according to conventional and alternative metrics: a systematic review. JBJS Rev 2017 May;5(5):e5. [doi: 10.2106/JBJS.RVW.16.00059] [Medline: 28557819]
87. Fenner M. What can article-level metrics do for you? PLoS Biol 2013 Oct;11(10):e1001687 [FREE Full text] [doi: 10.1371/journal.pbio.1001687] [Medline: 24167445]
88. Featherstone R. Scholarly tweets: measuring research impact via altmetrics. J Can Health Libr Assoc 2014 Aug 1;35(2):60. [doi: 10.5596/c14-015]

Abbreviations

AAS: Altmetrics attention score
DOI: digital object identifier
IF: impact factor
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RIF: research impact framework

Edited by G Eysenbach; submitted 24.07.19; peer-reviewed by EM Messner, A Barke, J McCall, A Ketchum, E Yan; comments to author 28.10.19; revised version received 30.12.19; accepted 29.03.20; published 06.07.20.

Please cite as:
Bardus M, El Rassi R, Chahrour M, Akl EW, Raslan AS, Meho LI, Akl EA
The Use of Social Media to Increase the Impact of Health Research: Systematic Review
J Med Internet Res 2020;22(7):e15607
URL: https://www.jmir.org/2020/7/e15607
doi:10.2196/15607
PMID:32628113

©Marco Bardus, Rola El Rassi, Mohamad Chahrour, Elie W Akl, Abdul Sattar Raslan, Lokman I Meho, Elie A Akl. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Information Loss in Harmonizing Granular Race and Ethnicity Data: Descriptive Study of Standards

Karen Wang[1,2], MD, MS; Holly Grossetta Nardini[3], MLS; Lori Post[4], PhD; Todd Edwards[5], MS, PhD; Marcella Nunez-Smith[1], MD, MHS; Cynthia Brandt[2,6], MD, MPH

[1]Equity Research and Innovation Center, General Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[2]Center for Medical Informatics, Yale School of Medicine, New Haven, CT, United States

[3]Harvey Cushing/John Hay Whitney Medical Library, Yale School of Medicine, New Haven, CT, United States

[4]Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Chicago, IL, United States

[5]Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, United States

[6]Veteran Affairs Connecticut Healthcare System, US Department of Veteran Affairs, West Haven, CT, United States

Corresponding Author:
Karen Wang, MD, MS
Equity Research and Innovation Center
General Internal Medicine
Yale School of Medicine
100 Church Street South, A200
New Haven, CT, 06520
United States
Phone: 1 203 785 5233
Email: karen.wang@yale.edu

Abstract

Background: Data standards for race and ethnicity have significant implications for health equity research.

Objective: We aim to describe a challenge encountered when working with a multiple–race and ethnicity assessment in the Eastern Caribbean Health Outcomes Research Network (ECHORN), a research collaborative of Barbados, Puerto Rico, Trinidad and Tobago, and the US Virgin Islands.

Methods: We examined the data standards guiding harmonization of race and ethnicity data for multiracial and multiethnic populations, using the Office of Management and Budget (OMB) Statistical Policy Directive No. 15.

Results: Of 1211 participants in the ECHORN cohort study, 901 (74.40%) selected 1 racial category. Of those that selected 1 category, 13.0% (117/901) selected Caribbean; 6.4% (58/901), Puerto Rican or Boricua; and 13.5% (122/901), the mixed or multiracial category. A total of 17.84% (216/1211) of participants selected 2 or more categories, with 15.19% (184/1211) selecting 2 categories and 2.64% (32/1211) selecting 3 or more categories. With aggregation of ECHORN data into OMB categories, 27.91% (338/1211) of the participants can be placed in the “more than one race” category.

Conclusions: This analysis exposes the fundamental informatics challenges that current race and ethnicity data standards present to meaningful collection, organization, and dissemination of granular data about subgroup populations in diverse and marginalized communities. Current standards should reflect the science of measuring race and ethnicity and the need for multidisciplinary teams to improve evolving standards throughout the data life cycle.

(J Med Internet Res 2020;22(7):e14591) doi:10.2196/14591

KEYWORDS

continental population groups; multiracial populations; multiethnic groups; data standards; health status disparities; race factors; demography

Introduction

The cutting edge of precision medicine, such as the integration of contextual data about the social determinants of health with individual health data and the leveraging of data from across different studies, is seen as a mechanism to innovate and solve health problems for all populations [1,2]. These solutions can only be realized by grounding them in the concept of health

XSL•FO
RenderX

equity [3,4]. According to the Robert Wood Johnson Foundation [5],

> *Health equity means that everyone has a fair and just opportunity to be healthier. This requires removing obstacles to health such as poverty, discrimination, and their consequences, including powerlessness and lack of access to good jobs with fair pay, quality education and housing, safe environments, and health care.*

To attain health equity, the research community must identify where health disparities exist but may inadvertently exacerbate health disparities by failing to identify invisible or at-risk populations. Without this information, we may reinforce inequities rather than identify policies, laws, systems, environments, and practices that can improve opportunities for health in these communities [6-9].

There is a significant body of literature about the collection of and proposed definitions for more granular data for race, ethnicity, and other demographic data in the health sciences that would better identify at-risk populations. In some areas, more granular data collection is used to be more inclusive of the ways that individuals self-identify and to examine how racism and discrimination affect health in these populations [10-13]. However, the research community has not sufficiently considered how to meaningfully organize granular data about subgroup populations in diverse communities so that these data can be used across multiple studies to address the challenges in these communities [5,14]. In this paper, we used our experience with a cohort study in the eastern Caribbean to illustrate the need for an updated and comprehensive data standard to enable future data integration and sharing of data sources for diverse multiracial populations.

Methods

The Eastern Caribbean Health Outcomes Research Network (ECHORN) cohort study follows community-dwelling adults older than 40 years residing in Barbados, Trinidad, Tobago, and the United States' territories of the US Virgin Islands and Puerto Rico [15]. Baseline participants were enrolled between 2013 and 2016 and completed a questionnaire to capture self-reported sociodemographics and health-related information. In the questionnaire, participants had the option to self-identify race and ethnicity by selecting any of the items listed: mixed or multiracial, white, black or African, Caribbean, Asian, East Indian, Hispanic or Latino, Puerto Rican or Boricua, other, or prefer not to answer. This list was developed with stakeholder input to reflect the current scientific literature on the measurement of race and ethnicity in epidemiologic and health outcomes research. Excluding the choice of "other", there were 256 potential answers.

The National Institutes of Health (NIH) requires investigators to collect and report race and ethnicity based on Office of Management and Budget (OMB) Statistical Policy Directive No. 15 data standards, developed in 1977 and updated in 2007 [16]. The standards include 7 race and ethnicity categories, with instructions to use 2 questions. The first question collects ethnicity information (Hispanic/Latino or not Hispanic/Latino), and the second provides the option to select more than one of the 5 racial categories (American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or other Pacific Islander, or white). Data reported are required to include (1) the number of respondents in each ethnic category, (2) the number of respondents who selected only 1 of 5 racial categories, (3) the number of respondents who selected multiple racial categories, and (4) the number of respondents in each racial category who identified as Hispanic or Latino. Reporting detailed distributions of multiple responses should be aggregated into the required categories.

Results

In an analysis of ECHORN data from 1211 participants, 901 (74.40%) selected 1 racial category. Of those that selected only 1 category, 13.0% (117/901) selected Caribbean; 6.4% (58/901), Puerto Rican or Boricua; and 13.5% (122/901), mixed or multiracial. A total of 17.84% (216/1211) of participants selected 2 or more categories, with 15.19% (184/1211) selecting 2 categories and 2.64% (32/1211) selecting 3 or more categories. The participants who selected more than one category included but were not limited to white Hispanic, black Hispanic, black Asian, black East Indian, and Caribbean Hispanic individuals. Based on OMB data standards, by collapsing those who selected multiple groups and those who selected only the multiracial and mixed category, 27.91% (338/1211) of the ECHORN participants are placed in the "more than one race" category. The nuanced race and ethnicity data become simply "more than one race, black or African American, white, Hispanic or Latino, Asian, or not reported/unknown."

Discussion

Principal Findings

Data standards for race and ethnicity need to reflect and evolve with the scientific advances around the measurement and mechanisms by which race and ethnicity can affect health [17,18]. The stated goals of the OMB standards are to harmonize the collection and presentation of population race and ethnicity information across federal databases and to facilitate comparisons and analyses [16]. To achieve the objectives of precision medicine, ECHORN data on those who selected multiple categories, such as the white Hispanic, black Hispanic, black Asian, black East Indian, or Caribbean Hispanic individuals, could be aggregated with other data sources with defined granular race and ethnicity data on this population. For example, the US Census Bureau has developed granular data collection standards on race and ethnicity, quantifying the growth of these multiracial and multiethnic populations. The State of Massachusetts, along with other states, has for several decades mandated the collection of more granular race and ethnicity information, with over 30 categories on many data collection forms, including birth certificates and clinical data [19-21]. However, it is not specified how multiracial and multiethnic populations' granular data can be harmonized to facilitate comparisons and analyses across any datasets collected using different standards. Current research collaborations, such

XSL•FO
RenderX

as Observational Health Data Sciences and Informatics, that are focused on facilitating harmonization and sharing data by using standard ontologies such as Systematized Nomenclature of Medicine - Clinical Terms and Logical Observation Identifiers Names and Codes, have no agreed-upon standard for granular race and ethnicity data or a mechanism to map across data standards [22].

Standards have not been refined or expanded adequately to accommodate studies collecting granular race and ethnicity data on mixed or multiracial individuals [23,24], and these imprecise standards have unmeasured effects on health. Current standards may contribute to health disparities by providing insights on persons privileged by these data structures and by further marginalizing persons whose racial and ethnic identity is obscured in suboptimal data standards for classification and harmonization. While the OMB standards clearly state that "the racial and ethnic categories set forth in the standards should not be interpreted as being primarily biological or genetic in reference," [16] current standards used in health sciences continue to support the conflation of the social categories of race and ethnicity with the biologic and genetic categories in day-to-day practice. This conflation has powerful implications for population-level biomedical discoveries and clinical care and treatment [25]. For example, inherent inequities in data standards are reflected in the continued use of monoracial guidelines as the standard of care in clinical practice, in spite of knowledge that suggests the fallacies of this framework [26,27]. It also has implications for population genetics and the translation of discoveries. These social constructs are used to develop standards for genetic population norms, wherein outlier data are discarded and study findings are associated with a singular racial group, influencing the development of biologic treatments directed at mutations that are predominantly in a particular racial group [28,29]. However, there is no biological origin for much of what we define as health inequities [30]. At the individual level, understanding how contextual factors, such as racism and discrimination, affect how a multiracial and multiethnic person understands and addresses their own health risks is significantly limited because of the lack of data and data standards that organize and share granular data in a viable way [31]. Importantly, multiracial and multiethnic populations are only one example of diverse subgroups affected by this lack of evolved standards; we risk the loss of rich information on the individuals who select more than one category for their race and ethnicity.

Several incremental changes can be instituted to improve these data standards on race and ethnicity. In addition to standardizing the definitions and collection of more granular race and ethnicity data across research studies, as advocated by numerous prior researchers [10-13], the standards need to be developed and adopted by funding agencies, such as the NIH, so that researchers from different groups who are collecting granular data can share and aggregate data [32,33]. These comprehensive data standards need to specify how to systematically categorize self-identified granular race and ethnicity data for those who have selected 2 or more categories. The standards can provide mechanisms to map race and ethnicity data collection standards to other standards over time, such as the Centers for Disease Control and Prevention Race and Ethnicity Code Set, Version 1.0, or the United States Census Bureau [32-34]. As data are collected and shared between regional and national entities, this mapping between standards will enable the aggregation of smaller data about underrepresented populations across studies and global contexts [32,35,36] and will provide a systematic method for individuals who have collected granular data to organize and map their data to less-evolved data collection standards.

Conclusion

Comprehensive data standards for race and ethnicity throughout the life course of health science research studies are critical to identifying and achieving heath equity for populations differentially affected by discrimination [23,24,37]. Partnerships with relevant stakeholders in the development and mapping of standards are essential and will likely increase availability of meaningful and usable data [38]. Multidisciplinary research teams that focus on health equity (eg, informaticists, data scientists, information scientists, geneticists, sociologists, and public health researchers) need to adapt standards and advance the theory and systems for organizing, storing, and analyzing complex data. These standards are urgently needed to ensure the sustained value of data collected about diverse populations and their health. With the status quo, we will continue to privilege one group's data over another's, lack meaningful and useable information of diverse racial and ethnic population groups, and likely further perpetuate health inequities.

Acknowledgments

We acknowledge Ms Hannah Friedman, who was a research assistant on this project, and Ms Mary Miller and Ms Natasha Wenner for their editorial assistance.

Conflicts of Interest

Drs KW, CB, and MNS are supported in part by U54MD010711 from the National Institute on Minority Health and Health Disparities. The funder had no role in the design or conduct of the manuscript. Views expressed are those of the authors and do not represent those of the funding source.

References

1. Fuchs VR. Social Determinants of Health: Caveats and Nuances. JAMA 2017 Jan 03;317(1):25-26. [doi: 10.1001/jama.2016.17335] [Medline: 28030707]

XSL•FO
RenderX

2. Breen N, Jackson JS, Wood F, Wong DW, Zhang X. Translational Health Disparities Research in a Data-Rich World. Am J Public Health 2019 Jan;109(S1):S41-S42. [doi: 10.2105/AJPH.2019.304951] [Medline: 30699034]
3. Schinasi LH, Auchincloss AH, Forrest CB, Diez Roux AV. Using electronic health record data for environmental and place based population health research: a systematic review. Ann Epidemiol 2018 Jul;28(7):493-502. [doi: 10.1016/j.annepidem.2018.03.008] [Medline: 29628285]
4. Braveman P. Health disparities and health equity: concepts and measurement. Annu Rev Public Health 2006;27:167-194. [doi: 10.1146/annurev.publhealth.27.021405.102103] [Medline: 16533114]
5. Braveman P, Arkin E, Orleans T, Proctor D, Plough A. What is Health Equity? Robert Woods Johnson Foundation. URL: https://www.rwjf.org/en/library/research/2017/05/what-is-health-equity-.html [accessed 2020-05-02]
6. Bourgois P, Holmes SM, Sue K, Quesada J. Structural Vulnerability: Operationalizing the Concept to Address Health Disparities in Clinical Care. Acad Med 2017 Mar;92(3):299-307 [FREE Full text] [doi: 10.1097/ACM.0000000000001294] [Medline: 27415443]
7. Khoury MJ, Iademarco MF, Riley WT. Precision Public Health for the Era of Precision Medicine. Am J Prev Med 2016 Mar;50(3):398-401 [FREE Full text] [doi: 10.1016/j.amepre.2015.08.031] [Medline: 26547538]
8. Anderson LM, Adeney KL, Shinn C, Safranek S, Buckner-Brown J, Krause LK. Community coalition-driven interventions to reduce health disparities among racial and ethnic minority populations. Cochrane Database Syst Rev 2015 Jun 15(6):CD009905. [doi: 10.1002/14651858.CD009905.pub2] [Medline: 26075988]
9. Betancourt JR, Green AR, Carrillo J, Ananeh-Firempong O. Defining cultural competence: a practical framework for addressing racial/ethnic disparities in health and health care. Public Health Rep 2003;118(4):293-302 [FREE Full text] [doi: 10.1093/phr/118.4.293] [Medline: 12815076]
10. Ford CL, Harawa NT. A new conceptualization of ethnicity for social epidemiologic and health equity research. Soc Sci Med 2010 Jul;71(2):251-258 [FREE Full text] [doi: 10.1016/j.socscimed.2010.04.008] [Medline: 20488602]
11. Griffith DM, Moy E, Reischl TM, Dayton E. National data for monitoring and evaluating racial and ethnic health inequities: where do we go from here? Health Educ Behav 2006 Aug;33(4):470-487. [doi: 10.1177/1090198106287923] [Medline: 16769756]
12. Mays VM, Ponce NA, Washington DL, Cochran SD. Classification of race and ethnicity: implications for public health. Annu Rev Public Health 2003;24:83-110 [FREE Full text] [doi: 10.1146/annurev.publhealth.24.100901.140927] [Medline: 12668755]
13. Williams DR, Lavizzo-Mourey R, Warren RC. The concept of race and health status in America. Public Health Rep 1994;109(1):26-41 [FREE Full text] [Medline: 8303011]
14. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750. [doi: 10.1038/gim.2016.183] [Medline: 27929525]
15. Wang KH, Thompson TA, Galusha D, Friedman H, Nazario CM, Nunez M, ECHORN Writing Group. Non-communicable chronic diseases and timely breast cancer screening among women of the Eastern Caribbean Health Outcomes Research Network (ECHORN) Cohort Study. Cancer Causes Control 2018 Mar;29(3):315-324 [FREE Full text] [doi: 10.1007/s10552-018-1005-4] [Medline: 29423760]
16. NIH Policy on Reporting Race and Ethnicity Data: Subjects in Clinical Research. National Institutes of Health.: National Institutes of Health; 2001 Aug 08. URL: https://grants.nih.gov/grants/guide/notice-files/not-od-01-053.html [accessed 2020-05-07]
17. Goodman AH. Why genes don't count (for racial differences in health). Am J Public Health 2000 Nov;90(11):1699-1702. [doi: 10.2105/ajph.90.11.1699] [Medline: 11076233]
18. Roth WD. The multiple dimensions of race. Ethnic and Racial Studies 2016 Mar 21;39(8):1310-1338. [doi: 10.1080/01419870.2016.1140793]
19. Jorgensen S, Thorlby R, Weinick RM, Ayanian JZ. Responses of Massachusetts hospitals to a state mandate to collect race, ethnicity and language data from patients: a qualitative study. BMC Health Serv Res 2010 Dec 31;10:352 [FREE Full text] [doi: 10.1186/1472-6963-10-352] [Medline: 21194450]
20. Hawkins SS, Cohen BB. Affordable Care Act standards for race and ethnicity mask disparities in maternal smoking during pregnancy. Prev Med 2014 Aug;65:92-95 [FREE Full text] [doi: 10.1016/j.ypmed.2014.05.002] [Medline: 24823903]
21. Folkerts G, van Esch B, Janssen M, Nijkamp FP. Superoxide production by broncho-alveolar cells is diminished in parainfluenza-3 virus treated guinea pigs. Agents Actions Suppl 1990;31:139-142. [doi: 10.1007/978-3-0348-7379-6_18] [Medline: 1964360]
22. Race and Ethnicity in the OMOP CDM [discussion forum]. Observational Health Sciences Data and Informatics. URL: https://forums.ohdsi.org/t/race-and-ethnicity-in-the-omop-cdm/8700 [accessed 2020-05-02]
23. Efthimiadis E, Afifi M. Population groups: indexing, coverage, and retrieval effectiveness of ethnically related health care issues in health sciences databases. Bull Med Libr Assoc 1996 Jul;84(3):386-396 [FREE Full text] [Medline: 8883987]
24. Aspinall PJ. The operationalization of race and ethnicity concepts in medical classification systems: issues of validity and utility. Health Informatics J 2016 Jul 25;11(4):259-274. [doi: 10.1177/1460458205055688]
25. Kahn J. Genes, race, and population: avoiding a collision of categories. Am J Public Health 2006 Nov;96(11):1965-1970. [doi: 10.2105/AJPH.2005.067926] [Medline: 17018831]

XSL•FO
RenderX

26. Cortes-Bergoderi M, Thomas RJ, Albuquerque FN, Batsis JA, Burdiat G, Perez-Terzic C, et al. Validity of cardiovascular risk prediction models in Latin America and among Hispanics in the United States of America: a systematic review. Rev Panam Salud Publica 2012 Aug;32(2):131-139. [doi: 10.1590/s1020-49892012000800007] [Medline: 23099874]
27. DeFilippis AP, Young R, McEvoy JW, Michos ED, Sandfort V, Kronmal RA, et al. Risk score overestimation: the impact of individual cardiovascular risk factors and preventive therapies on the performance of the American Heart Association-American College of Cardiology-Atherosclerotic Cardiovascular Disease risk score in a modern multi-ethnic cohort. Eur Heart J 2017 Feb 21;38(8):598-608 [FREE Full text] [doi: 10.1093/eurheartj/ehw301] [Medline: 27436865]
28. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. N Engl J Med 2015 Jun 04;372(23):2229-2234. [doi: 10.1056/NEJMsb1503104] [Medline: 26014593]
29. Spratt DE, Chan T, Waldron L, Speers C, Feng FY, Ogunwobi OO, et al. Racial/Ethnic Disparities in Genomic Sequencing. JAMA Oncol 2016 Aug 01;2(8):1070-1074 [FREE Full text] [doi: 10.1001/jamaoncol.2016.1854] [Medline: 27366979]
30. Suzuki K, Von Vacano DA, editors. Reconsidering Race: Social Science Perspectives on Racial Categories in the Age of Genomics. Oxford, England: Oxford University Press; 2018.
31. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, et al. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 2003 Mar 20;348(12):1170-1175. [doi: 10.1056/NEJMsb025007] [Medline: 12646676]
32. Institute of Medicine, Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement. In: Ulmer C, McFadden B, Nerenz DR, editors. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: National Academies Press; 2009.
33. Strmic-Pawl HV, Jackson BA, Garner S. Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality. Sociology of Race and Ethnicity 2017 Dec 04;4(1):1-13. [doi: 10.1177/2332649217742869]
34. Race and Ethnicity Code Set Version 1. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/phin/resources/vocabulary/documents/cdc-race--ethnicity-background-and-purpose.pdf [accessed 2020-05-02]
35. Representing Patient Race and Ethnicity. HealthIT.gov. URL: https://www.healthit.gov/isa/representing-patient-race-and-ethnicity [accessed 2020-05-02]
36. Ethnic group, national identity, and religion. Office for National Statistics. URL: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion [accessed 2020-05-02]
37. Aspinall P. The categorization of African descent populations in Europe and the USA: should lexicons of recommended terminology be evidence-based? Public Health 2008 Jan;122(1):61-69. [doi: 10.1016/j.puhe.2007.02.018] [Medline: 17645900]
38. Ward M, Schulz AJ, Israel BA, Rice K, Martenies SE, Markarian E. A conceptual framework for evaluating health equity promotion within community-based participatory research partnerships. Eval Program Plann 2018 Oct;70:25-34 [FREE Full text] [doi: 10.1016/j.evalprogplan.2018.04.014] [Medline: 29894902]

Abbreviations

ECHORN: Eastern Caribbean Health Outcomes Research Network
NIH: National Institutes of Health
OMB: Office of Management and Budget

Edited by G Eysenbach; submitted 03.05.19; peer-reviewed by P Aspinall, D Griffith; comments to author 29.09.19; revised version received 24.02.20; accepted 12.03.20; published 20.07.20.

Please cite as:
Wang K, Grossetta Nardini H, Post L, Edwards T, Nunez-Smith M, Brandt C
Information Loss in Harmonizing Granular Race and Ethnicity Data: Descriptive Study of Standards
J Med Internet Res 2020;22(7):e14591
URL: http://www.jmir.org/2020/7/e14591/
doi:10.2196/14591
PMID:32706693

©Karen Wang, Holly Grossetta Nardini, Lori Post, Todd Edwards, Marcella Nunez-Smith, Cynthia Brandt. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Guided Web-Based Self-Management Intervention for Depression or Dysthymia: Randomized Controlled Trial With a 12-Month Follow-Up Using an Active Control Condition

Caroline Oehler[1], MSc; Frauke Görges[2], PhD; Mandy Rogalla[2], MSc; Christine Rummel-Kluge[2*], Prof Dr; Ulrich Hegerl[3*], Prof Dr

[1]German Depression Foundation, Leipzig, Germany

[2]Department of Psychiatry and Psychotherapy, University Leipzig, Leipzig, Germany

[3]Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe-Universität Frankfurt, Frankfurt, Germany

[*]these authors contributed equally

Corresponding Author:
Caroline Oehler, MSc
German Depression Foundation
Goerdelerring 9
Leipzig, 04109
Germany
Phone: 49 03412238 ext 20
Email: caroline.oehler@deutsche-depressionshilfe.de

Abstract

Background: An increasing number of studies suggest that web-based interventions for patients with depression can reduce their symptoms and are expected to fill currently existing treatment gaps. However, evidence for their efficacy has mainly been derived from comparisons with wait-list or treatment as usual controls. In particular, designs using wait-list controls are unlikely to induce hope and may even have nocebo effects, making it difficult to draw conclusions about the intervention's efficacy. Studies using active controls are rare and have not yielded conclusive results.

Objective: The main objective of this study is to assess the acute and long-term antidepressant efficacy of a 6-week, guided, web-based self-management intervention building on the principles of cognitive behavioral therapy (iFightDepression tool) for patients with depression compared with web-based progressive muscle relaxation as an active control condition.

Methods: A total of 348 patients with mild-to-moderate depressive symptoms or dysthymia (according to the Mini International Neuropsychiatric Interview) were recruited online and randomly assigned to 1 of the 2 intervention arms. Acute antidepressant effects after 6 weeks and long-term effects at 3-, 6-, and 12-month follow-up were studied using the Inventory of Depressive Symptomatology–self-rating as a primary outcome parameter and change in quality of life (Short Form 12) and user satisfaction (client satisfaction questionnaire) as secondary outcome parameters. Treatment effects were assessed using mixed model analyses.

Results: Over the entire observation period, a greater reduction in symptoms of depression (P=.01) and a greater improvement of life quality (P<.001) was found in the intervention group compared with the active control group. Separate tests for each time point revealed significant effects on depressive symptoms at the 3-month follow-up (d=0.281; 95% CI 0.069 to 0.493), but not after 6 weeks (*main outcome:*d=0.192; 95% CI −0.020 to 0.404) and 6 and 12 months. The intervention was significantly superior to the control condition with respect to user satisfaction (25.31 vs 21.97; t_{259}=5.804; P<.01).

Conclusions: The fact that antidepressant effects have been found for a guided self-management tool in comparison with an active control strengthens the evidence base for the efficacy of web-based interventions. The antidepressant effect became most prominent at the 3-month follow-up. After 6 weeks of intervention, significant positive effects were observed on life quality but not on depressive symptoms. Although the effect size of such web-based interventions on symptoms of depression might be smaller than that suggested by earlier studies using wait-list control conditions, they can be a cost-effective addition to antidepressants and face-to-face psychotherapy.

Trial Registration: International Clinical Trials Registry Platform ICTRP080-15-09032015; https://apps.who.int/trialsearch/Trial2.aspx?TrialID=DRKS00009323

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e15361) doi:10.2196/15361

KEYWORDS

depression; dysthymic disorder; randomized controlled trial; cognitive behavioral therapy; internet-based intervention; active control; iCBT; self-management; iFightDepression; web-based intervention

Introduction

Web-Based Interventions in the Treatment of Depression

Web-based interventions for people with depression have been evaluated positively in numerous randomized controlled trials (RCTs). They raise the hope of offering a cost-effective and easily disseminated intervention via the internet [1] for one of the most common disorders worldwide [2]. Cognitive behavioral therapy (CBT) is an evidence-based treatment for depression recommended in national and international treatment guidelines, but access to this treatment is limited [3]. Especially in primary care, where the majority of patients with diagnosed depression are treated [4], access to psychotherapy is problematic. Web-based interventions are promising for closing this treatment.

The majority of web-based interventions for depression are based on techniques derived from CBT, consist of 5 to 15 modules, and incorporate psychoeducational material as well as interactive elements or tasks [5]. Their efficacy seems to have been confirmed by several reviews and meta-analyses finding statistically significant, moderate effect sizes (d=0.56 [6] and d=0.59 [7]; Hedges g=0.50 [8]; d=.67 [9]) when comparing internet-based interventions with treatment as usual (TAU) or with wait-list controls. A consistent finding is that interventions that incorporate some kind of guidance (through personal contact or via email support) turn out to have better retention rates and antidepressant effects than self-guided interventions [8,10]. Web-based interventions across several disorders also have been found to produce stable effects for up to 3 years in a review incorporating 14 trials with 902 participants, 3 of which dealt with depression and found positive effects after 2-3.5 years [11].

Although these studies and meta-analyses appear to provide consistent evidence for the efficacy of web-based interventions, an important limitation sheds doubts on this area of research: Effect sizes observed in trials investigating web-based interventions are dependent on the control condition used [12]. Studies with wait-list controls produce larger effect sizes (Hedges g=0.9) than those using care as usual or other control conditions (Hedges g=0.38) [10]. One possible explanation for this phenomenon is the influence of patients' expectations regarding the success of the intervention. Especially for patients with depression (a condition in which hopelessness and a negative view of the future are part of the symptomatology), becoming aware of *only* being *in the control condition* does not induce hope but might instead produce nocebo effects [12].

A Critical Review of the Evidence

So far, most trials on web-based interventions have relied on wait-list and care as usual controls [13], with comparisons to wait-list control likely overestimating the real efficacy. To date, only a few studies have compared web-based interventions for depression with active or placebo control interventions and even fewer report on follow-up data.

Mackinnon et al [14] tested 2 active treatments (a web-based intervention and an information website) against an attention control condition, in which participants discussed certain aspects of their lifestyle with the study team. Although varying in content (depression-specific content vs aspects about lifestyle), all 3 groups received the same amount of telephone contact with the study team [15]. The observed effect sizes for the web-based intervention (d=0.38) and the information website (d=0.29) were statistically significant but smaller than those in wait-list–controlled studies.

Johansson et al [16] tested a tailored and standardized version of the same web-based intervention against an active control condition (an online discussion forum) and found moderate to large effects with regard to symptoms of depression (d=0.84 and d=0.57). However, patients randomized into the control arm first received the invitation to join an online discussion forum and, after the intervention period, received the standardized treatment. In this design, patients knew when they were randomized into a control and wait-list condition and might have been disappointed or felt set back by this, potentially leading to less hope induction or even nocebo effects.

Glozier et al [17] conducted a double-blind study on an internet-based cognitive behavioral therapy (iCBT) program. The iCBT intervention was compared with an internet-delivered health intervention for depression in patients with cardiovascular disease and both, participants and study assistants assessing the outcome measure, were unaware of which condition was the active intervention. The iCBT program led to a significantly greater reduction in symptoms of depression, but the effect size in this well-controlled design was small (Cohen d=0.16).

Taken together, when using more valid control conditions, the evidence for the antidepressant efficacy of web-based interventions relies on a limited number of studies, and reported effect sizes appear to be small. The best available evidence for antidepressant efficacy thus far stems from a meta-analysis comparing internet-based interventions for depression with face-to-face psychotherapy. Andersson et al [18] combined 5 studies that directly compared guided internet-based interventions with face-to-face psychotherapy (often in a group setting) and found a small effect size in favor of web-based intervention (Hedges g=0.12), which was not significantly different from zero.

However, not only the direct intervention effects should be subjected to critical examination. The existing results on long-term efficacy might also be influenced by the choice of controls. Psychotherapeutic interventions are claiming long-term positive effects resulting from the learning of new behavioral and cognitive patterns. Although the first results thus far seem positive, the number of studies incorporating longer follow-up

XSL•FO
RenderX

periods is limited and often stems from wait-list–controlled trials. From the 40 high-quality RCTs that were included in a recent meta-analysis for web-based interventions targeting depression [8], only 3 studies contained data on a follow-up of at least 12 months. All 3 studies used within-group comparisons of symptom severity following the intervention to later time points and reported stable effects over the respective follow-up periods [19-21]. This statistical comparison is mainly used in wait-list–controlled trials, which have the disadvantage of not enabling between-group comparisons at follow-up. As symptoms of depression usually fluctuate spontaneously and episodes of depression are usually remitting after several months even when untreated, it is unclear if the results found at follow-up are due to a successful treatment, an initial placebo effect plus spontaneous remission, further treatment options participants took, or other external factors. So far, only a few studies with smaller samples have reported between-group comparisons at follow-up. For example, in one study on 69 participants, using a face-to-face intervention as a comparator, data from a long-term follow-up confirmed that after 3.5 years both groups still did not differ in a statistically significant way [20]. Through the design of this study, it will be possible to expand our understanding of the long-term effects of web-based interventions for depression.

This Study

The general objective of this study is to strengthen the evidence base for web-based interventions and to close the described gaps in the previous results. To this end, we implemented a web-based active control condition. This control condition was designed to be as similar as possible to the intervention concerning credibility, hope induction, and contact with the study assistants. Furthermore, a 12-month follow-up was implemented.

The main objective was to compare changes in self-rated symptom severity occurring during the trial period up to 12 months for patients with mild-to-moderate depression, who either used a CBT-based, web-based self-management tool (iFightDepression [iFD]) or took part in an active control condition (progressive muscle relaxation [PMR]). We expected the iFD tool to be superior to an active control in terms of symptom reduction. Our main outcome of intervention effects after the 6-week intervention period (as predefined in the study protocol) was supplemented by long-term data to extend the relevance of our conclusions.

Further objectives focused on a more in-depth analysis of the effects of both interventions by doing the following:

- Considering possible covariates such as age, gender, or amount of guidance that might influence the intervention effect.
- Examining the differences between both interventions with respect to changes in self-rated quality of life.
- Examining possible differences between the 2 conditions concerning user experience as well as the amount of usage and the duration and content of guidance in a descriptive and explorative manner.

Methods

Trial Design

This study is an RCT assessing the efficacy and usability of a guided web-based self-management intervention (iFD) compared with an active control condition (PMR) after 3 and 6 weeks of intervention as well as after 3, 6, and 12 months postintervention. The trial was conducted in accordance with the Declaration of Helsinki. The complete study protocol was published elsewhere [22] and is in line with the Consolidated Standards of Reporting Trials statement [23].

Recruitment and Selection of Participants

The study participants were recruited throughout Germany via the website, social media channels, appearances in other media, and newsletters of the German Depression Foundation (DF). Furthermore, newsletters of associated organizations were used for distribution. Interventions were offered free of charge, and no reimbursement was offered to participants.

Individuals interested in taking part in the study were directed to a website providing general information on the study procedures and a web-based questionnaire assessing several inclusion and exclusion criteria. In the study information provided to the participants, both interventions were described as equivalent offers to not induce a bias in expectations. After successfully passing the screening questionnaire, contact details could be left for the main screening that took place via telephone. This procedure led to the preselection of individuals with sufficient internet literacy to meet the minimal study requirements. All screening procedures and guidance during the trial were carried out by psychologists or psychotherapists. If the screening was successful, participants were asked to provide written informed consent for participating in the study and to provide the telephone number of a confidant, whom the study team could contact in case of a suspected crisis (for further details on the screening and inclusion procedures see Oehler et al [22]).

Inclusion criteria were outpatient status, a diagnosis of depressive disorder with presently mild or moderate severity (F32.0, F32.1, F33.0, and F33.1) or dysthymia (F34.1) according to the Mini International Neuropsychiatric Interview (MINI) and patient health questionnaire-9 (PHQ-9; score 5-14, indicating mild-to-moderate symptoms), aged 18 years and above, sufficient language skills to meet the study requirements, and internet access. Outpatient status was taken as one of the inclusion criteria, so that patients could be referred to their local care provider in case of a crisis. Exclusion criteria were dementia, drug or alcohol abuse within the last 6 months, drug or alcohol addiction, schizophrenia, manic episodes or bipolar disorder, obsessive-compulsive disorder (all according to the MINI), known personality disorders (F60.2 and F60.31), acute suicidal tendencies, severe somatic disorders requiring immediate treatment, pregnancy, and participation in another clinical trial within the past 4 weeks.

All participants who provided written informed consent and matched the inclusion and exclusion criteria were randomized using the minimization algorithm by Pocock [24] and stratified

XSL•FO
RenderX

for gender (male/female), depression severity (mild/moderate according to PHQ-9), and CBT experience (present/absent) with an 80% chance of using the algorithm's recommendation.

To the best of our knowledge, no comparable studies were available for power calculation at the time. Therefore, power calculation was based on the results of a study [25], which compared face-to-face CBT with an active control condition (guided self-help group) in depressed primary care patients and found a difference of 5.3 points on the Inventory of Depressive Symptomatology–clinician rated after 10 weeks. For this study, the difference was estimated to be 4.0 points on the Inventory of Depressive Symptomatology–self-rating (IDS-SR), as we expected the difference to be slightly smaller in a web-based trial compared with face-to-face interventions. On the basis of this estimation, 122 patients per group are needed to detect a difference with a power of 80% (α=.05). It was planned to include 360 participants to obtain at least 250 complete data sets after an expected dropout of approximately 30%.

Ethics and Trial Registration

The protocol for this study was reviewed and approved by the Ethics Committee at the Faculty of Medicine, University of Leipzig, on February 11, 2015.

The trial was registered under the identification code DRKS00009323 at the German Register for Clinical Trials, with the title *Efficacy of an Internet-Based Self-Management Intervention for Adult Primary Care Patients With Mild and Moderate Depression or Dysthymia*. The international registration of the study was carried out via the *International Clinical Trials Registry Platform* with the secondary ID 080-15-09032015.

Assessments

There were 3 main assessment points during the intervention period (before the intervention and after 3 and 6 weeks of intervention) and 3 follow-up assessments (3, 6, and 12 months after the end of the intervention period). During this period, the primary and secondary outcome measures were collected.

Measures

Primary Outcome

The IDS-SR was used as the primary outcome measure in this study (range 0-84) to assess changes in depression severity. The scale has been shown to be useful in detecting symptom change as well as residual symptoms in patients with depression [26]. The concordant validity with the Beck Depression Inventory and the Hamilton Rating Scale for Depression has been shown to be appropriate ($r \geq .88$) [27], and the internal consistency at baseline was acceptable (Cronbach α=.78).

Secondary Outcomes

To assess changes in the perceived health-related quality of life, Short-Form 12 (SF-12) was used. It was developed as a practical short form of the Short-Form-36. A mental and a physical component score (both ranging from 0 to 100) can be calculated from the questionnaire answers for which moderate to high convergent validity has been shown in several studies [28,29].

The German version (Fragebogen zur Messung der Patientenzufriedenheit) of the client satisfaction questionnaire-8 (CSQ-8) was used to assess acceptance and feasibility of the interventions. As the questionnaire was originally developed for the evaluation of hospital stays, the wording was slightly adapted to fit web-based interventions. A similar adaptation yielded good internal consistency (omega=.95) [30].

The usage of the intervention was assessed each week during the intervention period using a self-report measure with 2 items. Participants were asked how often they had worked with the intervention during the last week and how much time they had spent on it.

Tracking the objective usage was possible only for the iFD group through the log files of the iFD tool website. Offline use in the form of printed worksheets cannot be tracked. For the PMR group, it was registered if participants downloaded the weekly changing intervention audios. The actual use could not be tracked. Owing to these limitations, both objective measures only served as an approximation and validation of the subjective measures.

Monitoring Instrument

To monitor changes and detect possible deteriorations in depressive symptoms over the course of the intervention, the 9-item (PHQ-9; range 0-27) was used. The PHQ-9 is a short, well-validated, and widely used measure [31,32]. The internal consistency at baseline was below the values usually reported for this scale (Cronbach α=.66). During the intervention, patients reporting symptoms indicating severe depression for 3 weeks in a row or acute suicidality were contacted by the study assistants via telephone or email and, if necessary, advised to seek appropriate clinical support. This was necessary in 12 cases. A protocol for managing acute suicidality was established. If patients reported severe symptoms of depression but wanted to continue using the intervention, their cases were discussed with the supervising physician.

Adverse Events

At the beginning of the study (T0), after 3 weeks (T1), and after 6 weeks (T2), adverse events were recorded. Every new event that led to the inability to work or that needed medical treatment was recorded, and a possible connection to the intervention was assessed. Events that led to unplanned inpatient treatment or were life-threatening or lethal were classified as serious adverse events. Serious adverse events were passed on to the supervising physician for review.

Documentation of Guidance

The duration of all planned calls made by the study assistants during the intervention period was recorded and added to provide a sum score for the overall guidance received by each patient. Additionally, the content and perceived quality were rated by the study assistants. Adverse events were recorded and topics relevant to the study, for example, date of next appointment, were discussed. Additional calls to ensure patient safety after they reported suicidal thoughts or severe symptoms of depression in the questionnaires as well as the follow-up calls at T3 to T5 were not counted as guidance.

XSL•FO
RenderX

Interventions

The iFightDepression Tool

The iFD tool is a guided web-based self-management tool based on the principles of CBT. It includes 6 core workshops, each comprising written information, worksheets, exercises, and a mood rating. For this study, participants were asked to use the tool for 6 weeks and to complete 1 workshop per week. Each week's workshop covered a different topic (eg, an activity diary, monitoring and adapting one's sleep, or challenging automated negative thoughts). The content and development are described in more detail elsewhere [22,33]. iFD offers the opportunity to complete worksheets on the web or to use a printed version. Patients were asked to try out each workshop and, if helpful, continue using the learned techniques. The iFD tool did not change during the study period, except for a news box on the landing page that was updated approximately once a month.

Progressive Muscle Relaxation

In this study, PMR was used as the control condition. During the 6 weeks of intervention, participants were encouraged to practice PMR and to learn how to deliberately induce physical relaxation to reduce stress and mental tension. Lessons range from 13 to 33 min and build on one another, adding more muscle groups every week. At the beginning of each week, participants received a link to download the next lesson. They were instructed to practice on a daily basis, if possible, but at least two or 3 times a week and to integrate the practice into their daily routine.

PMR was chosen as a credible control intervention and is widely used in therapeutic settings, for example, as part of CBT or in the treatment of sleep disorders. The method is also highly accepted by the public as a form of self-help for depression [34,35] and is rated to be helpful for clinically depressed patients (n=736; 38.6%) very or moderately effective and (n=749; 39.2%) slightly effective [36]).

In a systematic review of several relaxation techniques (PMR or similar methods), relaxation was recommended as the first-line treatment in a stepped care approach. Antidepressant effects were visible shortly after relaxation interventions, superior to wait-list and no treatment but inferior to psychotherapy [37], making PMR a suitable choice as a control condition.

Guidance

Guidance was provided during 5 telephone calls by the study assistants (psychologists and psychotherapists) from the Research Center of the DF, supervised by a senior psychiatrist who was involved in the development of the iFD tool. Comparable with iFD guides outside of the study setting, all study assistants were qualified using the standard web-based seminar and used a guideline for the calls based on the webinar content. The focus of the guidance calls was to motivate the participants rather than discuss the intervention content.

To keep contact with the study staff comparable across both intervention groups, the same guideline was used in the guidance calls for both the iFD and PMR groups, and calls were carried out by the same study assistants.

Statistical Analysis

To pursue the main objective, changes over time in the primary outcome measure were investigated using a mixed model analysis, which included a random intercept and random slope for each participant. The variance-covariance structure was set to unstructured to avoid any constraints. This approach was adopted to make the best use of incomplete data while minimizing the bias to the parameter estimations [38,39]. All analyses were performed on the intent-to-treat sample using data from all randomized participants. As a sensitivity check, the analysis of the main outcome was repeated for a per-protocol sample (only participants having finished at least four workshops in the iFD tool or downloaded 4 sessions of PMR) and using an imputed data set. The parameter of interest for each model was the time x group interaction, specifying the differential change of symptoms over time attributable to the group assignment. A quadratic term and its interaction with the group variable was added (time x time x group) to allow for parabolic trends over time. Within- and between-group effect sizes (Cohen *d*) were calculated using the difference in means between intervention groups at posttreatment and follow-up for imputed data (with 50 imputations) taking into account the dependence of data collected within participants to avoid the loss of power due to incomplete cases [40] and employing the pooled standard deviation as the standardizer. Confidence intervals of the between-group effect sizes were examined to check for intervention effects at each time point.

Several studies have shown that certain covariates at the participant level, such as adherence to the intervention, participant age, or gender [41-43], can influence dropout rates and intervention success. It is also of interest whether the use of the intervention or the amount of contact with the study team influenced the changes in the outcome measures. Therefore, to meet our second objective, possible covariates (chosen based on the literature on covariates influencing the outcome of interventions for depressed patients) were added stepwise to the mixed model (base model), which contained fixed effects for time, time x time, and a dichotomous variable for the intervention group and their interactions, and were kept in the model if they improved the model fit as measured by the restricted maximum likelihood (REML) criterion at convergence. The following covariates were tested: the amount of guidance received (sum of minutes spent on guidance call per participant), sex (female vs male), age, self-reported amount of time spent working with the intervention (sum score of hours spent on intervention), and self-reported frequency of use (sum of self-reported times worked with the intervention). To control for possible effects of other treatments, a dichotomous variable for *taking antidepressants or receiving psychotherapy at screening* vs *not taking antidepressants or receiving psychotherapy at screening* was added as a possible covariate. Covariates and interactions that did not improve the model fit were excluded from the model.

For the third objective, an equivalent analysis was repeated for quality of life using the SF-12 data.

For all models described, model assumptions were checked using graphical inspections of the plotted residuals and the

XSL•FO
RenderX

normal Q-Q plot. The assumptions of normality and homoscedasticity were not violated in any of the models. The inspection of Cook distance estimates yielded several influential data points, but the model results remained unchanged after excluding these. Therefore, in this paper, mixed models using the full sample are reported.

To assess our final objective, descriptive analyses were performed on key features of intervention use and the results of the expectation questionnaire. Group differences in the contents of the guidance calls were tested for significance using a two-sample *Z* test for proportions to compare the frequency of each topic. A two tailed *t* test was performed on the CSQ-8 sum scores (user satisfaction) to assess the group difference statistically.

All analyses were performed using R [44]. For the mixed model calculations, the packages lme4 and lmerTest were used to estimate the model coefficients and corresponding *P* values. Results were calculated using REML estimation and the Kenward-Roger approximation to calculate the denominator degrees of freedom for the performed *t* statistics, as this has been reported to be the most robust way to determine the statistical significance of parameters in mixed models [45].

Results

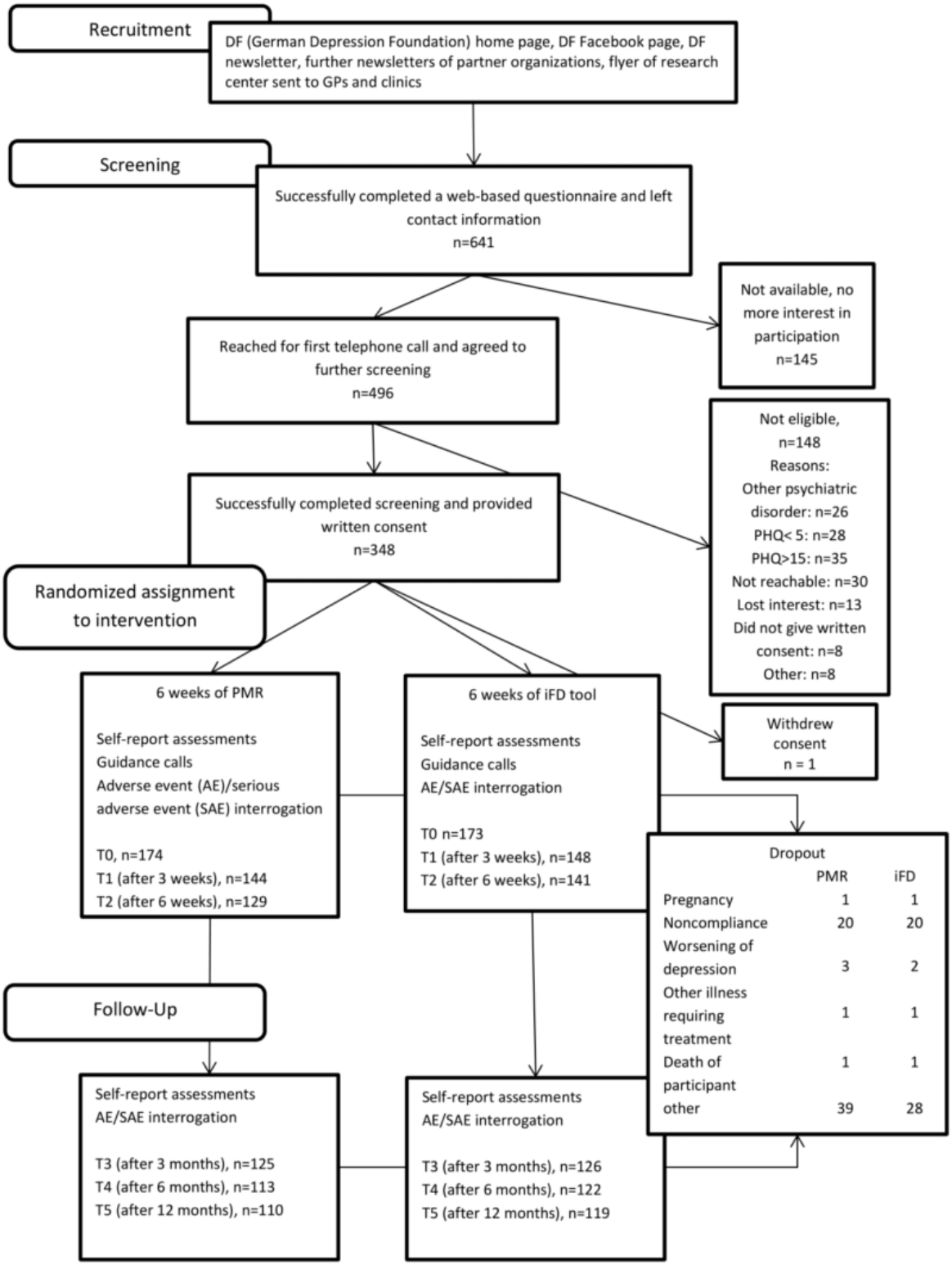
Enrollment and Baseline Characteristics

Recruitment for the trial commenced in June 2016 and was completed in August 2018. Follow-up data were collected until August 2019.

The participant flow chart (Figure 1) provides an overview of the screening and enrollment numbers of the patients (n=347) who were included in the study and randomized into 1 of the 2 treatment arms. The current sample can be described as treatment experienced and, in most cases, with recurrent depression. The levels of both education and internet literacy were high (Table 1).

XSL•FO
RenderX

Figure 1. Participant flow chart; a participant was counted as having completed 1 measurement point when either the telephone interview had taken place or the questionnaire had been filled out. Single measures might be reported with slightly differing sample sizes. PHQ: patient health questionnaire; PMR: progressive muscle relaxation; iFD: iFightDepression.



XSL•FO
RenderX

Table 1. Baseline characteristics.

Variable	Intervention group (n=173)	Control group (n=174)
Characteristics		
Age (years), mean (SD)	42.9 (12.4)	41.7 (12.4)
Female, n (%)	137 (79.2)	136 (78.3)
Acquired a university degree, n (%)	64 (37.0)	68 (39.1)
Internet usage <10 years, n (%)	157 (90.8)	155 (89.1)
Using the internet daily, n (%)	164 (94.8)	166 (95.4)
Inclusion diagnosis		
Major depression, n (%)	2 (1.2)	4 (2.3)
Recurrent major depression, n (%)	164 (94.8)	157 (90.2)
Dysthymia, n (%)	42 (24.3)	41 (23.6)
Currently under antidepressants, n (%)	115 (66.5)	108 (62.1)
Currently receiving psychotherapy, n (%)	97 (56.1)	93 (53.1)
Received psychotherapy in the past, n (%)	169 (97.7)	161 (92.6)
Psychiatric admission in the past, n (%)	120 (69.4)	117 (67.2)
Median number of self-reported episodes in the past, n (range)	6 (1-120)	5 (1-150)
Comorbidities		
Social phobia, n (%)	25 (14.5)	23 (13.2)
Agoraphobia or panic disorder, n (%)	33 (19.1)	27 (15.5)
Generalized anxiety disorder, n (%)	8 (4.6)	6 (3.4)

Adherence

Of the 347 patients included, 288 filled out the T1 measure (after 3 weeks) and 262 completed T2 (after 6 weeks). The follow-up measures were completed by 251 participants 3 months after the treatment ended, 235 participants after 6 months, and 229 participants after 12 months. Some of the measures reported have deviating sample sizes (n) due to a small number of participants omitting one or more of the measures.

According to self-report, participants in the intervention group used the iFD tool 23.6 (SD 12.6) times on average and participants in the PMR group practiced 22.4 (SD 9.5) times over the course of the 6-week intervention period. They reported to have spent an average of 6.2 (SD 4.7) hours using the iFD program and an average of 6.5 (SD 5.4) hours using PMR. Neither of these differences reached significance ($t_{303.26}$=0.606, *P*=.55 and $t_{334.91}$=0.620, *P*=.54, respectively).

Objective data, taken from the back end of the iFD tool and the download page of the PMR website, confirmed regular use and downloads of both interventions. The iFD users completed an average of 5.5 (SD 2.1) workshops and spent a mean of 3.8 (SD 3.0) hours using the tool on the web over the course of 18.7 sessions (the end of a session was defined by >30 min of idle time after the last click; time after the last click has not been included in average usage time). PMR users downloaded an average of 4.2 (SD 1.86) of the 6 relaxation lessons (n=165; for 10 participants, downloads had to be enabled differently due to technical problems and were not trackable).

Main Objective

Considering intervention effects on the IDS-SR as the primary outcome measure over the entire observation period (6 weeks of intervention + 12 months of follow-up), a significant difference in the symptom change over time was found, favoring the iFD group (an overview of the main and secondary outcomes over time is shown in Table 2). In the base model specified without covariates, the estimated fixed effect of interest (group x time interaction) differed statistically significant from zero ($t_{1157.2}$=−2.519; *P*=.01). A significant main effect of time ($t_{1196.2}$=−3.934; *P*<.001) was also observed, indicating a significant symptom reduction over time in both groups, as well as a significant interaction of time² x group caused by the greater curvature of the trajectory in the iFD group ($t_{1099.0}$=2.686; *P*=.007). Figure 2 depicts the changes in the main outcome variable as well as the values predicted by the base model.

XSL•FO
RenderX

Table 2. Mean values for symptoms of depression and quality of life over the course of the intervention, within-group effect sizes.

Intervention	T0 (baseline)		T1 (approximately after 3 weeks)		T2 (approximately after 6 weeks)		T3 (approximately after 3 months)		T4 (approximately after 6 months)		T5 (approximately after 12 months)		Within-group ES[a] (T0-T2)		Within-group ES[a] (T2-T5)	
	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Cohen *d*	95% CI	Cohen *d*	95% CI
IDS-SR[b]																
iFD[c]	27.5 (8.9)	173	24.3 (9.1)	148	20.8 (9.4)	135	19.3 (11.4)	126	19.8 (11.2)	122	19.8 (10.8)	119	−0.718	−0.937 to −0.501	−0.182	−0.394 to 0.032
PMR[d]	27.9 (8.8)	174	24.8 (10.4)	144	23.1 (10.3)	129	22.0 (11.7)	125	21.2 (11.5)	113	19.5 (11.0)	110	−0.619	−0.835 to −0.403	−0.372	−0.584 to −0.159
SF-12 MCS[e]																
iFD	33.6 (8.3)	173	35.6 (9.0)	146	38.9 (9.7)	133	40.3 (10.8)	126	40.2 (11.5)	122	39.8 (11.1)	119	0.519	0.304 to 0.734	0.056	−0.156 to 0.267
PMR	33.3 (8.1)	174	34.8 (9.7)	142	36.1 (10.2)	129	37.6 (10.1)	125	38.7 (10.5)	113	41.2 (10.0)	110	0.223	0.012 to 0.435	0.446	0.233 to 0.660
SF-12 PCS[f]																
iFD	46.9 (9.0)	173	48.1 (9.1)	146	47.3 (9.9)	133	48.8 (9.7)	125	48.7 (8.6)	122	47.8 (9.3)	119	0.093	−0.118 to 0.305	−0.040	−0.252 to 0.171
PMR	47.2 (9.9)	174	46.8 (9.8)	142	47.3 (9.7)	129	47.9 (8.9)	126	46.4 (9.9)	113	47.0 (9.8)	110	−0.038	−0.249 to 0.173	−0.092	−0.303 to 0.119
PHQ-9[g]																
iFD	9.1 (3.6)	173	7.9 (3.7)	146	6.9 (3.7)	133	7.0 (4.4)	126	7.4 (4.8)	122	6.7 (4.2)	119	−0.571	−0.898 to −0.356	0.012	−0.120 to 0.223
PMR	9.7 (3.3)	173	8.2 (3.8)	142	7.4 (3.7)	129	7.9 (4.4)	125	7.9 (4.1)	113	6.7 (4.6)	110	−0.800	−1.019 to −0.580	−0.066	−0.277 to 0.145

[a]ES: effect size, calculated based on imputed data sets.

[b]IDS-SR: Inventory of Depressive Symptomatology–self rating.

[c]iFD: iFightDepression.

[d]PMR: progressive muscle relaxation.

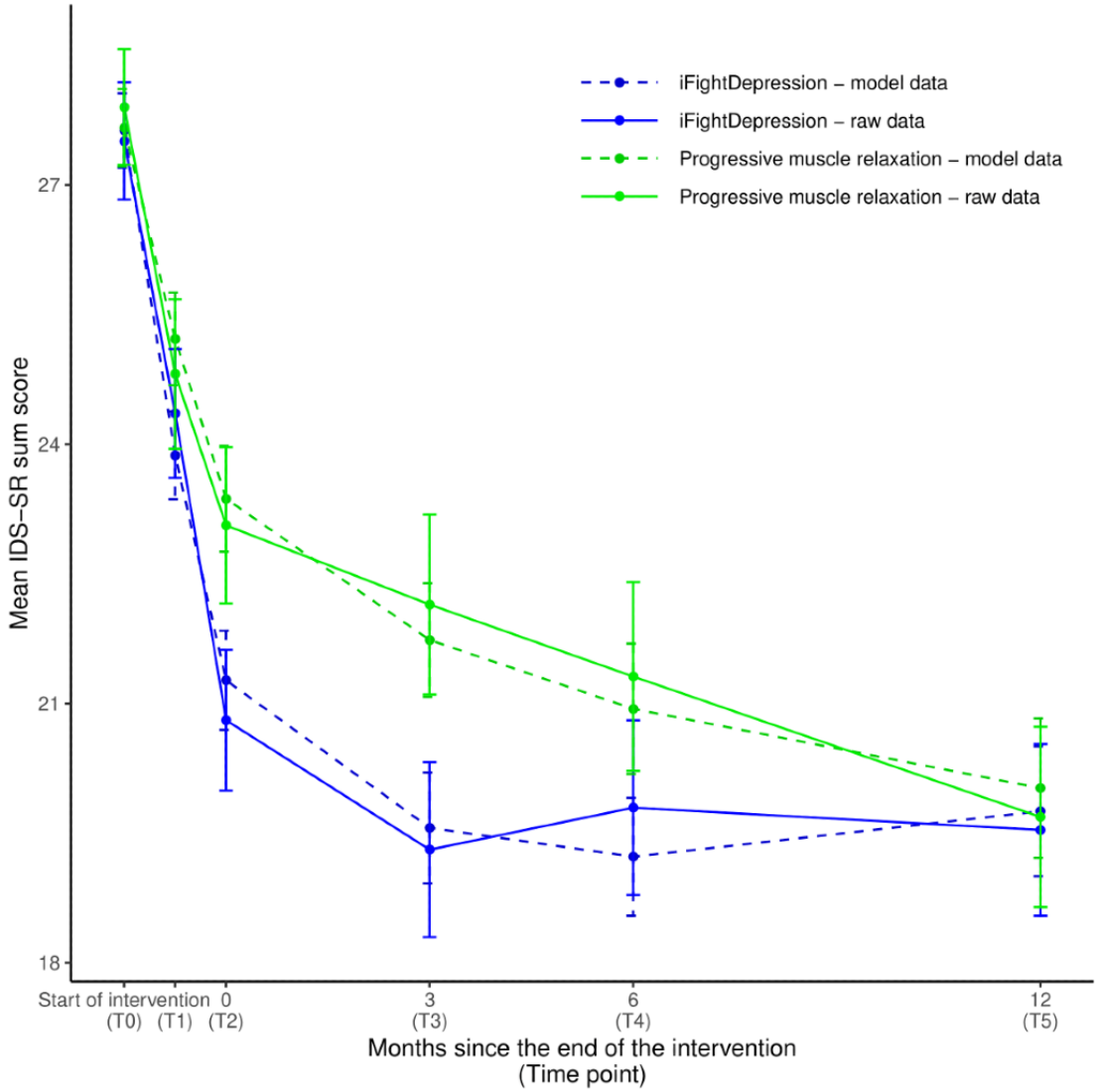
[e]SF-12 MCS: Short-Form 12 mental component score.

[f]SF-12 PCS: Short-Form 12 physical component score.

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

XSL•FO
RenderX

Figure 2. Changes over time in mean IDS-SR. Points of measurement: T0: before the intervention; T1: after 3 weeks of intervention; T2: after intervention period; T3 to T5: follow-up measurements after 3, 6, and 12 months; error bars: standard errors of the mean; and IDS-SR: Inventory of Depressive Symptomatology–self rating. Model results refer to the estimates of the base model including fixed effects for time x group and time x time + group.



As a measure of the model's ability to describe the data, R^2 was calculated. The base model yielded a conditional R^2 of 0.59 and a marginal R^2 of 0.08, indicating that 59% of the variance in the dependent variable was described by the model and 8% of variance can be explained by the fixed effects alone. The results did not change when using a per-protocol sample or the imputed data set and are therefore not reported separately (results for all models can be found in the Multimedia Appendix 1).

The between-group effect sizes and their confidence intervals, calculated from imputed data sets, are shown in Table 3 and provide an estimate of significance for the group differences at every assessment point. Group differences in the IDS-SR were not statistically significant following the 6-week intervention period (T2) but were so at the 3-month follow-up (T3). Within-group effect sizes comparing T0 and T2 measures (Table 2) can be described as medium according to the rough categorization proposed by Cohen [46] (iFD: d=−0.718; PMR: d=−0.619). After the intervention period, the IDS-SR scores remained stable in the iFD group and decreased further in the PMR group.

XSL•FO
RenderX

Table 3. Between-group effect sizes, results of the mixed models with and without covariates.

Measure	Between-group ES[a]–T2		Between-group ES[a]–T3		Between-group ES[a]–T4		Between-group ES[a]–T5		Fixed effect time x group			Fixed effect time x group		
	Cohen *d*	95% CI	Cohen *d*	95% CI	Cohen *d*	95% CI	Cohen *d*	95% CI	Base model estimate (SE)	*t* value (df)	*P* value	Covariate model estimate (SE)	*t* value (df)	*P* value
IDS-SR[b]	0.192	−0.020 to 0.404	0.281	0.069-0.493	0.030	−0.182 to 0.241	−0.025	−0.236 to 0.186	−2.486 (0.987)	−2.519 (1186.2)	.01	−2.975 (1.098)	−2.710 (941.3)	.007
SF-12 MCS[c]	−0.343	−0.555 to −0.130	−0.249	−0.461 to −0.037	−0.260	−0.472 to −0.048	−0.070	−0.141 to 0.281	3.553 (0.985)	3.608 (1198.4)	<.001	3.531 (1.111)	3.180 (945.7)	.002
SF-12 PCS[d]	−0.075	−0.286 to 0.136	−0.205	−0.416 to 0.007	−0.287	−0.499 to −0.075	−0.123	−0.334 to 0.088	0.794 (0.835)	0.950 (1161.1)	.34	0.837 (0.938)	0.892 (932.5)	.37

[a]ES: effect size; between-group effect sizes were corrected for unequal sample size and could therefore also be referred to as Hedges *g*. Effect sizes are calculated based on imputed values to make use of the full data set. Positive values indicate a higher score in the PMR group, and negative values indicate higher scores in the iFD group.

[b]IDS-SR: Inventory of Depressive Symptomatology–self-rating.

[c]SF-12 MCS: Short-Form 12 mental component score.

[d]SF-12 PCS: Short-Form 12 physical component score.

Secondary Objectives

Covariate Analysis

The final model built to predict the IDS-SR scores included the original fixed effects of time, time², and group as well as their interaction as predictors (the full model results are shown in the Multimedia Appendix 1). In addition, fixed effects for self-reported frequency of use, amount of guidance received over the course of the intervention, and the interaction of amount of guidance with the group variable were kept in the model. These covariates were chosen because they improved the model fit (as indicated by a smaller REML criterion at convergence). The significant effect of the amount of guidance ($t_{216.7}$=3.58; *P*<.001) showed that higher overall IDS-SR scores were associated with a greater amount of contact with the study team (model parameter for the fixed effect: 0.20 [SD 0.06]). In addition, the significant interaction of the group and amount of guidance ($t_{216.4}$=−2.13; *P*=.04) mirrors the fact that this was especially the case in the PMR condition.

The fixed effect for frequency of use did not reach statistical significance ($t_{216.2}$=0.58; *P*=.72) but improved the model fit and was therefore kept in the model. The fixed effects of sex, age, taking antidepressants or receiving psychotherapy at screening, and self-reported amount of time spent working with the intervention were not significantly different from zero and did not improve the overall model fit; thus, they were not added to the final model.

None of the other interactions between the covariates and the time or the group variable was significantly different from zero; therefore, the model was specified without additional interactions. This implies that, in this study, the intervention effect was not affected by the covariates named above in a way that was detectable with the current design. The final model yielded a conditional R^2 of 0.60 and a marginal R^2 of 0.10.

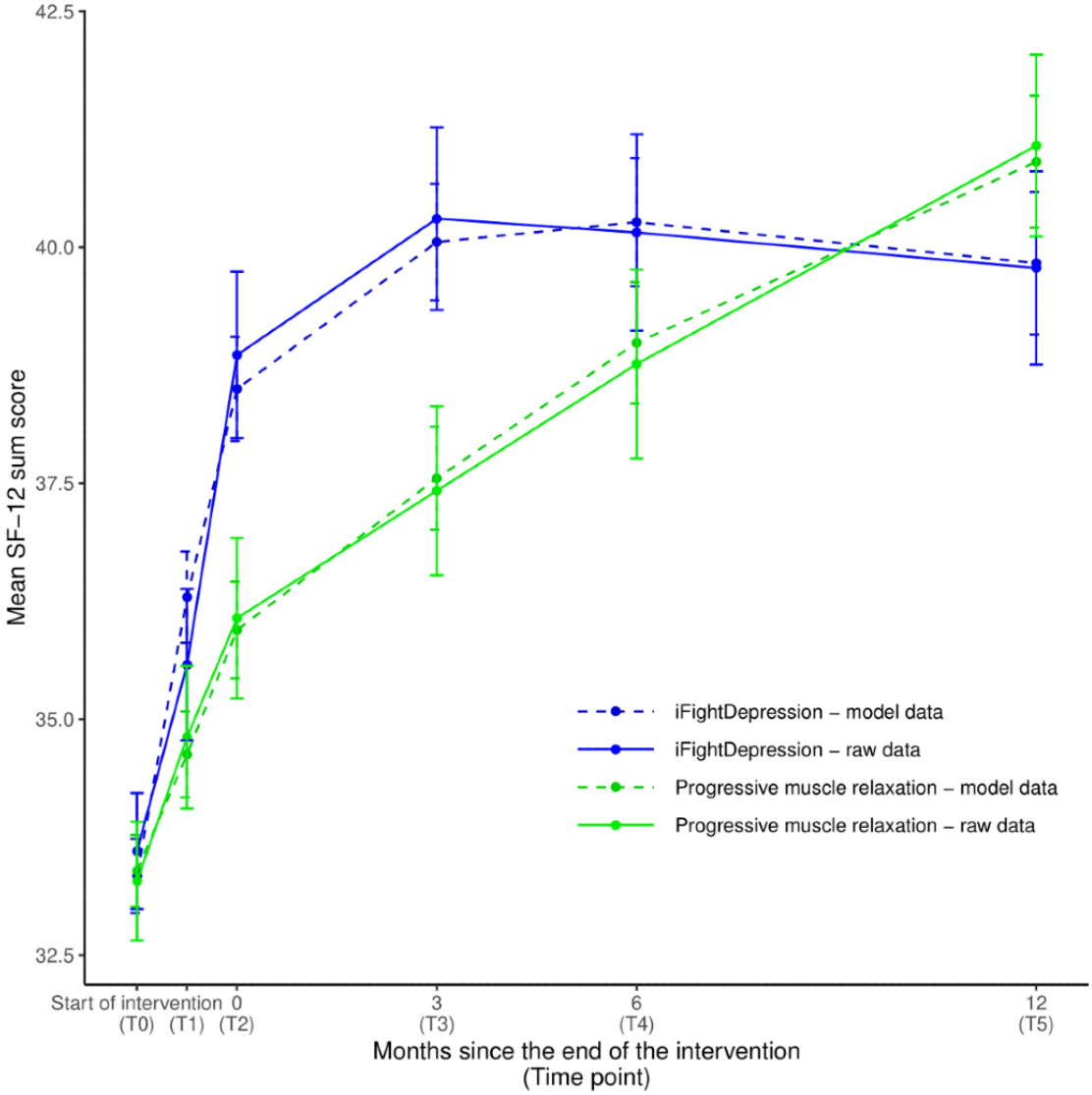
Quality of Life

In the mixed model describing the changes in quality of life (mental component score), the significant interaction of group x time ($t_{1198.4}$=−1.967; *P*=.049) reflects an intervention effect in favor of the iFD tool, indicating a greater improvement in the quality of life for iFD users compared with participants in the PMR group. A statistically significant time² x group interaction indicates greater curvature in the iFD group ($t_{904.8}$=−3.900; *P*<.001). The mental component score was significantly associated with the group variable ($t_{859.9}$=−2.274; *P*=.02; Figure 3), indicating an overall lower score on the SF-12 in the iFD group. The base model yielded a conditional R^2 of 0.55 and a marginal R^2 of 0.06.

The between-group effect sizes after the intervention (T2) and at the 3-month follow-up (T3) showed statistically significant intervention effects (Table 3).

XSL•FO
RenderX

Figure 3. Changes over time in mean SF-12. Points of measurement: T0: before the intervention; T1: after 3 weeks of intervention; T2: after intervention period; T3 to T5: follow-up measurements after 3, 6, and 12 months; error bars: standard errors of the mean; and SF-12: Short-Form 12 (mental component score). Model results refer to the estimates of the base model including fixed effects for time x group and time x time + group.



In an extended mixed model containing self-reported frequency of use and the amount of guidance, the covariates themselves did not have a significant effect on the SF-12 scores, but improved the model fit. In addition, the group variable was no longer assigned a significant fixed effect, meaning the model did not yield a significant main effect of intervention group. The final model yielded a conditional R^2 of 0.64 and a marginal R^2 of 0.06.

In the mixed model using the physical component score as a dependent variable, none of the relevant fixed effects significantly differed from zero. As expected, the physical component score was not affected by the intervention and remained unchanged over time (full model results are shown in the Multimedia Appendix 1).

Guidance

The mean time spent on the 5 guidance and study calls was 38.5 (SD 9.9) min per participant for the iFD group and 28.9 (SD 15.0) min per participant for the PMR group, with the difference being statistically significant ($t_{314.41}$=5.078; P<.001). The subjective quality as rated by the guiding study assistants was *positive* or *mostly positive* for 96.7% (736/761) of the calls in the iFD group and for 96.5% (691/716) in the PMR group. The relative occurrence of specific topics during the guidance calls, as recorded by the guides, are reported in Table 4.

XSL•FO
RenderX

Table 4. Frequency of specific topics that occurred in the telephone calls, rated by guide, raw values, and percentage (calculated with number of observations/interviews as 100%).

Topic[a]	iFD[b] (n=761 observations), n (%)	PMR[c] (n=716 observations), n (%)	Chi-square value (*df*)	*P* value[d]	Holm-corrected *P* value
Comprehension problems	54 (7.0)	4 (0.6)	40.1 (1)	<.001	<.001
Current state of health	658 (86.5)	626 (87.4)	0.2 (1)	.64	.76
Content support	247 (32.5)	126 (17.6)	42.4 (1)	<.001	<.001
Motivation of participant	121 (15.9)	137 (19.1)	2.5 (1)	.12	.59
Dissatisfaction with intervention	83 (10.9)	159 (22.2)	33.6 (1)	<.001	<.001
Positive feedback on the intervention	301 (39.6)	206 (28.8)	1.4 (1)	.23	.76
Questions about released content (only iFD)	27 (3.5)	N/A[e]	N/A	N/A	N/A
Study organizational topics	639 (84.0)	596 (83.2)	0.1 (1)	.76	.76
Irregular participation in intervention	78 (10.3)	61 (8.5)	1.1 (1)	.29	.76

[a]Multiple topics could be the subject of each call; therefore, percentages do not add up to 100.

[b]iFD: iFightDepression.

[c]PMR: progressive muscle relaxation.

[d]*P* values corrected for multiple testing using Hochberg correction.

[e]Not applicable.

User Satisfaction

The mean sum score of the CSQ-8, a measure of user satisfaction, was 25.31 in the iFD group and 21.97 in the PMR group out of a possible 32, the difference being statistically significant (t_{259}=5.8044; *P*<.01). Satisfaction with the iFD tool seems to be in the expected range, with a sum score similar to those of other studies on internet interventions (26.26 major depression [MD] prevention [30], 26.05 stress management, 24.51 MD treatment in routine psychiatric care [47], 24.96 guided internet-based intervention for depression [48], 22.88 unguided internet-based intervention for depression, and 22.4 happiness training [49]).

Adverse Events

During the course of the intervention, 159 adverse events and 4 serious adverse events were recorded. Three adverse events were rated as *possibly related* to the intervention, 2 in the iFD group, and 1 in the PMR group. The *possibly related* adverse events were all deteriorations of the patient's mood. One patient in each group described the feeling of trying one more intervention that did not help as a contributing factor to the deterioration, and 1 participant in the iFD group reported worse symptoms of depression after reducing her antidepressants without consulting her physician, as she expected to get better by using iFD. Of the serious adverse events, 2 occurred in each intervention group and none were rated as related or possibly related to the intervention. None of the adverse or serious adverse events recorded during the follow-up period were related or possibly related to the intervention.

Discussion

Principal Findings

This study is one of the first studies on the antidepressant effects of iCBT using a control condition that is both credible and comparable with the iCBT group in the form of delivery, terms of duration, and contact with the study team. To this end, we assessed between-group differences concerning changes in symptom severity over time as the main objective and possible covariates influencing the intervention effect as well as differences in health-related quality of life as a further objective.

The results concerning our main objective (decrease in symptoms of depression as measured with the IDS-SR) differed significantly between the intervention and control groups according to the mixed model results, favoring the iFD group. This effect was especially apparent at the 3-month follow-up (*d*=0.281, reduction on the IDS-SR was 8.2 points for iFD and 4.8 for PMR), whereas the group difference was only approaching significance after 6 weeks (following the intervention, the main outcome as predefined in the study protocol, *d*=0.192). These effect sizes are considerably smaller than those reported in a meta-analysis for trials using wait-lists as controls (*d*=0.56) but more similar to studies using TAU controls (*d*=0.23) [10]. Finding smaller effects or even no effect, when using an active control condition instead of a wait-list control, is to be expected as placebo effects should affect both groups in a similar manner and nocebo effects should be reduced to a minimum, thereby (mostly) ruling out expectation effects and greater hope induction driving the effect. As studies on depressed patients have been shown to be very susceptible to placebo [50], our finding of significant effects compared with a possible active control that is known to be helpful for depressed patients is noteworthy. The effect sizes are in accordance with the studies by Mackinnon et al [14] and

XSL•FO
RenderX

Johansson et al [15], who observed small but significant effects when comparing internet-based interventions to active and attention control conditions.

For further interpretation of the current results, the composition of the sample should be considered. The current sample was a self-selected community sample with a high rate of very educated, internet-affine patients who had experienced several episodes of depression in the past (IFD: median 6; PMR: median 5), who had already gained experience with psychotherapy (past: 330/347, 95.1%; present: 190/347, 54.8%), and who were receiving some kind of treatment at the beginning of the study. The mean age at the onset of depression was 21.8 (SD 12.1) years, indicating that, on average, participants had been experiencing episodes of depression for about 20 years before participating in this study.

This implies that the participants in this study could have already gained a lot of experience in therapeutic techniques and in managing their symptoms, which might have attenuated the treatment effects in comparison to studies treating patients with a first episode of depression or less treatment experience. This might indicate that even experienced patient groups can benefit from this type of intervention. On the other hand, some patients might assume that PMR is less likely to be effective than iCBT and will perceive less hope induction. iFD might be perceived as a new and promising treatment option. This can result in overestimation of the true antidepressant effect.

Patient characteristics such as gender and age as well as the amount of use and guidance did not significantly influence the intervention effect in this study. This is in line with the results on sociodemographic data having no predictive value in an individual patient data meta-analysis by Karyotaki et al [51]. However, Karyotaki et al [51] and Donkin et al [42] reported a positive association between the intervention effect and the amount of intervention received, which could not be replicated in this study. A possible explanation for the lack of this association is that patients might regulate uptake according to their needs, that is, one person might feel that they have received sufficient help after 2 modules, whereas a different person might feel that they need all 6 modules. If this holds true, the total amount of intervention received might not predict the outcome, as it did not in this study. In contrast to previous findings [8,15], the amount of guidance provided in this study did not predict treatment success in the mixed model analysis, possibly due to the fact that the mean amount of guidance (33.7 min) was at the lower end of the 30 to 180 min, which Baumeister et al [52] had previously proposed as an optimal amount of guidance in a review of 14 studies on the impact of guidance in web-based interventions. Larger variation in the amount of guidance, for example, in meta-analysis incorporating several studies with differing designs, might reestablish this effect.

In this study, the 1-year follow-up allowed the analysis of the long-term effects of the interventions. The current results show that although the improvement in the intervention group remained stable, the control condition caught up after 6 months and was not significantly different after 12 months. Our sample had a high proportion of patients with recurrent depressive disorders; therefore, spontaneous remission over the course of the 1-year follow-up is plausible.

For our third objective, assessing changes in quality of life, a statistically significant effect for intervention was found. Although both groups reported improved quality of life concerning their mental health, this increase was significantly more pronounced in the group assigned to the iFD tool with a small but statistically significant between-group effect size after 6 weeks of intervention (d=−0.343) and 3 months (d=−0.359) postintervention. In their overview of several studies on internet-based interventions for depression, Andrews and Williams [53] reported several trials showing positive effects on both quality of life and disability with moderate to large effect sizes. Our results are in line with this, the smaller magnitude of the effect possibly being caused by the stronger control condition.

To address our fourth objective, this study explored key features of the provision of a web-based intervention with guidance according to a practice-oriented guideline. The average time spent on guidance was 9.6 min longer in the iFD group, and content-related support and comprehension problems were significantly more often part of the guidance calls in the iFD group than in the PMR group. These differences in the amount and content of the guidance calls might reflect a differential need within the groups. Although the PMR training was extended a bit each week, the iFD tool offered new topics and new tasks each week and might have been intellectually more challenging. However, it was also perceived as more positive by the participants, leading to significantly fewer complaints about the intervention during the telephone calls and to a higher score on the satisfaction rating.

In addition to improving treatment outcomes, guidance and contact with the study team are often referred to in the literature as important factors to improve adherence to the intervention. In this study, dropout was acceptable and subjective, and objective measures indicated that usage was high in both groups, with most participants completing the majority of weekly assignments. This implies that this rather small amount of guidance was sufficient, raising hopes that iFD might also prove effective in routine care where guidance will be offered by therapists and physicians. However, the highly structured study environment with an extensive screening interview (not counted as guidance time in this study) and regular questionnaires might have further improved adherence. This should be kept in mind when implementing iFD in routine care.

Limitations

Several limitations must be considered. First, the active control condition cannot be considered a placebo because there are studies pointing toward an antidepressant effect of PMR [37]. This can lead to an underestimation of the true antidepressant effect of the iCBT intervention. As Hart et al [54] argued, it is almost impossible to construct a psychosocial or behavioral intervention that is a true placebo, that is, an intervention that is harmless, completely inactive or inert, and comparable with the intervention tested. Therefore, we decided to choose an active control that might have affected the outcome measures, but that is thought to be less effective in treating depression

XSL•FO
RenderX

while still resembling the web-based intervention in its form and the amount of time patients spend on it. However, this design has the disadvantage of not disentangling spontaneous remission from the effects produced by the control condition.

Second, the amount of guidance was significantly longer in the iFD group despite our best efforts to keep it parallel. Although the amount of guidance was not related to outcome, it cannot be ruled out that this difference had a certain impact on our results.

Third, the generalizability of the results is limited by the fact that the results were obtained within the context of an RCT, which differs from implementation in routine practice. Johansson and Andersson [15] showed that even adding a structured screening interview to otherwise unguided web-based interventions led to increased effects of the intervention, so it is plausible that for the current trial, the study procedures had an impact on the effects reported here. Furthermore, our sample was self-selected and is not representative of all patients with depression. This might have led to more motivated participants, who were more interested in web-based interventions than the average patient. In addition, the trial was conducted by the institution implementing iFD in Germany; thus, the presence of allegiance bias cannot be excluded.

Finally, we did not collect data on additional treatment options that patients might have utilized during the follow-up period. Therefore, we cannot control for possible treatment differences between the iFD and PMR groups during follow-up.

Conclusions

This study is one of the few studies that used a valid control condition (PMR) as well as a 12-month follow-up. The results confirm that with active controls, the effect sizes are smaller than those in wait-list–controlled designs. Nevertheless, the results strengthen the evidence base for the efficacy of web-based CBT interventions in patients with depressive disorders. Over the entire observation period, the iFD tool was superior to an active control in the reduction of symptoms of depression (P=.01) and in the improvement of quality of life (P<.001). Although the predefined primary outcome (reduction of symptoms of depression on the IDS-SR after 6 weeks) was not statistically significant, the secondary outcome (quality of life, measured through the mental component score on the SF-12 after 6 weeks) was. Patients in the intervention group reported an accelerated symptom improvement most prominent 3 months after the intervention and, on average, transitioned from moderate to mild symptoms of depression (according to the cutoff values of the IDS-SR [55]). The intervention effects remained stable for up to 12 months, with the control participants continuing to improve and catch up with the intervention group from month 6 onward. A total guidance time of 38.5 min with an acceptable adherence indicates that iFD has the potential to be an efficient complement to treatment and to help reduce treatment gaps in psychosocial interventions. The fact that iFD is free of charge and available in 12 languages makes it an option in many countries worldwide.

Acknowledgments

The authors would like to thank the project team of Preventing Depression and Improving Awareness through Networking in the EU for the development of the iFD tool and the European Alliance Against Depression, especially Juliane Hug, Elisabeth Kohls, and Pia Driessen for their support concerning the iFD tool and their scientific contribution.

Finally, the authors would like to express their gratitude to all participants taking part in the study.

The trial was funded within the framework of cooperation between the DF and the *Deutsche Bahn Stiftung gGmbH*.

Authors' Contributions

CR-K and FG developed the study design, wrote the ethics proposal, developed the quality control plan (inspection plan), and revised the manuscript. FG and CO performed the final testing and adaptation of the study procedures. FG, MR, and CO conducted the study and provided guidance for web-based interventions. CO performed the data analysis, wrote the first draft of the manuscript, and supervised the editing process. UH and CR-K supervised the data analysis and, together with MR and FG, revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Conflicts of Interest

UH is a member of the advisory board for Janssen Pharmaceutica, received travel costs and honorarium as speaker and research grant for an investigator-initiated trial from Medice, and travel costs and honorarium as speaker for Servier. CR received lecture honoraria from Servier and was part of an advisory board for RECORDATI. UH, CR, MR, FG, and CO work or have worked for the DF, implementing the iFD tool in Germany. UH and CR have been part of the project in which the iFD tool was developed.

Multimedia Appendix 1
Efficacy of a Guided Web-Based Self-Management Intervention for Depression or Dysthymia: Randomized Controlled Trial With a 12-Month Follow-Up Using an Active Control Condition.
[PDF File (Adobe PDF File), 770 KB - jmir_v22i7e15361_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 4365 KB - jmir_v22i7e15361_app2.pdf]

References

1. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 1;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
2. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the global burden of disease study 2015. Lancet 2016 Oct 8;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
3. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression Langfassung. AWMF Sktuell. 2009. URL: https://www.awmf.org/uploads/tx_szleitlinien/nvl-005l_S3_Unipolare_Depression_2017-05.pdf [accessed 2020-06-13]
4. Wittchen H, Pittrow D. Prevalence, recognition and management of depression in primary care in Germany: the depression 2000 study. Hum Psychopharmacol 2002 Jun;17(Suppl 1):S1-11. [doi: 10.1002/hup.398] [Medline: 12404663]
5. Andersson G, Carlbring P, Hadjistavropoulos H. Internet-based cognitive behavior therapy. In: Hofmann SG, Asmundson GJG, editors. The Science of Cognitive Behavioral Therapy. Cambridge, Massachusetts: Elsevier Academic Press; 2017:535-549.
6. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
7. Twomey C, O'Reilly G, Byrne M. Effectiveness of cognitive behavioural therapy for anxiety and depression in primary care: a meta-analysis. Fam Pract 2015 Feb;32(1):3-15. [doi: 10.1093/fampra/cmu060] [Medline: 25248976]
8. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):- [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
9. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
10. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
11. Andersson G, Rozental A, Shafran R, Carlbring P. Long-term effects of internet-supported cognitive behaviour therapy. Expert Rev Neurother 2018 Jan;18(1):21-28. [doi: 10.1080/14737175.2018.1400381] [Medline: 29094622]
12. Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, et al. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. Lancet Psychiatry 2017 Sep;4(9):725-732. [doi: 10.1016/S2215-0366(17)30153-0] [Medline: 28396067]
13. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 1;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
14. Mackinnon A, Griffiths KM, Christensen H. Comparative randomised trial of online cognitive-behavioural therapy and an information website for depression: 12-month outcomes. Br J Psychiatry 2008 Feb;192(2):130-134. [doi: 10.1192/bjp.bp.106.032078] [Medline: 18245031]
15. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
16. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]
17. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
18. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
19. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
20. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]

XSL•FO
RenderX

21. Buhrman M, Syk M, Burvall O, Hartig T, Gordh T, Andersson G. Individualized guided internet-delivered cognitive-behavior therapy for chronic pain patients with comorbid depression and anxiety: a randomized controlled trial. Clin J Pain 2015 Jun;31(6):504-516. [doi: 10.1097/AJP.0000000000000176] [Medline: 25380222]
22. Oehler C, Görges F, Böttger D, Hug J, Koburger N, Kohls E, et al. Efficacy of an internet-based self-management intervention for depression or dysthymia - a study protocol of an RCT using an active control condition. BMC Psychiatry 2019 Mar 14;19(1):90 [FREE Full text] [doi: 10.1186/s12888-019-2063-1] [Medline: 30871544]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 Mar;31(1):103-115. [Medline: 1100130]
25. Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schütze M, Scheunemann W, et al. Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. Int J Neuropsychopharmacol 2010 Feb;13(1):31-44 [FREE Full text] [doi: 10.1017/S1461145709000224] [Medline: 19341510]
26. Rush AJ, Carmody T, Reimitz P. The inventory of depressive symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int J Method Psychiat Res 2006 Jun;9(2):45-59. [doi: 10.1002/mpr.79]
27. Drieling T, Schärer LO, Langosch JM. The inventory of depressive symptomatology: German translation and psychometric validation. Int J Methods Psychiatr Res 2007;16(4):230-236 [FREE Full text] [doi: 10.1002/mpr.226] [Medline: 18200596]
28. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
29. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Qual Life Res 2009 Aug;18(6):727-735. [doi: 10.1007/s11136-009-9483-1] [Medline: 19424821]
30. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
31. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the patient health questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
32. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
33. Arensman E, Koburger N, Larkin C, Karwig G, Coffey C, Maxwell M, et al. Depression awareness and self-management through the internet: protocol for an internationally standardized approach. JMIR Res Protoc 2015 Aug 6;4(3):e99 [FREE Full text] [doi: 10.2196/resprot.4358] [Medline: 26251104]
34. Jorm AF, Christensen H, Griffiths KM. The public's ability to recognize mental disorders and their beliefs about treatment: changes in Australia over 8 years. Aust N Z J Psychiatry 2006 Jan;40(1):36-41. [doi: 10.1080/j.1440-1614.2006.01738.x] [Medline: 16403035]
35. Mnich E, Makowski AC, Lambert M, Angermeyer MC, Knesebeck OV. Beliefs about depression--do affliction and treatment experience matter? Results of a population survey from Germany. J Affect Disord 2014 Aug;164:28-32. [doi: 10.1016/j.jad.2014.04.001] [Medline: 24856549]
36. Parker G, Crawford J. Judged effectiveness of differing antidepressant strategies by those with clinical depression. Aust N Z J Psychiatry 2007 Jan;41(1):32-37. [doi: 10.1080/00048670601057726] [Medline: 17464679]
37. Jorm AF, Morgan AJ, Hetrick SE. Relaxation for depression. Cochrane Database Syst Rev 2008 Oct 8(4):CD007142. [doi: 10.1002/14651858.CD007142.pub2] [Medline: 18843744]
38. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
39. Krueger C, Tian L. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biol Res Nurs 2004 Oct;6(2):151-157. [doi: 10.1177/1099800404267682] [Medline: 15388912]
40. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Br Med J 2009 Jun 29;338:b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
41. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
42. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
43. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the

XSL•FO
RenderX

U-CARE heart randomized controlled trial. J Med Internet Res 2018 Oct 10;20(10):e10754 [FREE Full text] [doi: 10.2196/10754] [Medline: 30305255]
44. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.
45. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
46. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
47. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
48. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
49. Görges F, Oehler C, von Hirschhausen E, Hegerl U, Rummel-Kluge C. GET.HAPPY - acceptance of an internet-based self-management positive psychology intervention for adult primary care patients with mild and moderate depression or dysthymia: a pilot study. Internet Interv 2018 Jun;12:26-35 [FREE Full text] [doi: 10.1016/j.invent.2018.03.001] [Medline: 30135766]
50. Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG. Meta-analysis of the placebo response in antidepressant trials. J Affect Disord 2009 Nov;118(1-3):1-8. [doi: 10.1016/j.jad.2009.01.029] [Medline: 19246102]
51. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
52. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
53. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: a model for dissemination into primary care. Clin Psychol Rev 2015 Nov;41:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]
54. Hart T, Fann JR, Novack TA. The dilemma of the control condition in experience-based cognitive and behavioural treatment research. Neuropsychol Rehabil 2008 Jan;18(1):1-21. [doi: 10.1080/09602010601082359] [Medline: 17852761]
55. IDS/QIDS Interpretation. URL: http://www.ids-qids.org/interpretation.html [accessed 2020-03-27]

Abbreviations

CBT: cognitive behavioral therapy
CSQ-8: client satisfaction questionnaire 8
DF: German Depression Foundation
iCBT: internet-based cognitive behavioral therapy
IDS-SR: Inventory of Depressive Symptomatology–self-rating
iFD: iFightDepression
MD: major depression
MINI: Mini International Neuropsychiatric Interview
PHQ-9: patient health questionnaire-9
PMR: progressive muscle relaxation
RCT: randomized controlled trial
REML: restricted maximum likelihood
SF-12: Short Form 12
TAU: treatment as usual

XSL•FO
RenderX

Edited by G Eysenbach; submitted 04.07.19; peer-reviewed by J Zhou, M Wolf; comments to author 16.12.19; revised version received 01.04.20; accepted 26.04.20; published 14.07.20.

Please cite as:
Oehler C, Görges F, Rogalla M, Rummel-Kluge C, Hegerl U
Efficacy of a Guided Web-Based Self-Management Intervention for Depression or Dysthymia: Randomized Controlled Trial With a 12-Month Follow-Up Using an Active Control Condition
J Med Internet Res 2020;22(7):e15361
URL: http://www.jmir.org/2020/7/e15361/
doi:10.2196/15361
PMID:32673233

©Caroline Oehler, Frauke Görges, Mandy Rogalla, Christine Rummel-Kluge, Ulrich Hegerl. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation

Yan Zhuang[1], BSc; Yin-Wu Chen[2], BSc, MSc, PhD; Zon-Yin Shae[2], BSc, MSc, PhD; Chi-Ren Shyu[1], PhD

[1]Institute for Data Science and Informatics, University of Missouri - Columbia, Columbia, MO, United States

[2]Artifical Intelligence Research Lab, Asia University, Taichung, Taiwan

Corresponding Author:
Chi-Ren Shyu, PhD
Institute for Data Science and Informatics
University of Missouri - Columbia
241 Robert and Patricia Naka Hall
University of Missouri
Columbia, MO, 65211
United States
Phone: 1 5738823884
Email: shyuc@missouri.edu

Abstract

Background: Data coordination across multiple health care facilities has become increasingly important for many emerging health care applications. Distrust has been recognized as a key barrier to the success of such applications. Leveraging blockchain technology could provide potential solutions tobuild trust between data providers and receivers by taking advantage of blockchain properties such as security, immutability, anonymity, decentralization, and smart contracts. Many health technologies have empirically proven that blockchain designs fit well with the needs of health care applications with certain degrees of success. However, there is a lack of robust architecture to provide a practical framework for developers to implement applications and test the performance of stability, efficiency, and scalability using standard blockchain designs. A generalized blockchain model is needed for the health care community to adopt blockchain technology and develop applications in a timely fashion.

Objective: This study aimed at building a generalized blockchain architecture that provides data coordination functions, including data requests, permission granting, data exchange, and usage tracking, for a wide spectrum of health care application developments.

Methods: An augmented, 3-layered blockchain architecture was built on a private blockchain network. The 3 layers, from bottom to top, are as follows: (1) incorporation of fundamental blockchain settings and smart contract design for data collection; (2) interactions between the blockchain and health care application development environment using Node.js and web3.js; and (3) a flexible development platform that supports web technologies such as HTML, https, and various programing languages. Two example applications, health information exchange (HIE) and clinical trial recruitment, were developed in our design to demonstrate the feasibility of the layered architecture. Case studies were conducted to test the performance in terms of stability, efficiency, and scalability of the blockchain system.

Results: A total of 331,142 simulated HIE requests from accounts of 40,000 patients were successfully validated through this layered blockchain architecture with an average exchange time of 11.271 (SD 2.208) seconds. We also simulated a clinical trial recruitment scenario with the same set of patients and various recruitment criteria to match potential subjects using the same architecture. Potential subjects successfully received the clinical trial recruitment information and granted permission to the trial sponsors to access their health records with an average time of 3.07 seconds.

Conclusions: This study proposes a generalized layered blockchain architecture that offers health technology community blockchain features for application development without requiring developers to have extensive experience with blockchain technology. The case studies tested the performance of our design and empirically proved the feasibility of the architecture in 2 relevant health application domains.

(J Med Internet Res 2020;22(7):e19029) doi:10.2196/19029

KEYWORDS

blockchain; smart contract; health information exchange; electronic health records; health care application

XSL•FO
RenderX

Introduction

Background

The health care industry generates abundant health data from various sources [1]. The meaningful use of health data can improve the decisions of health care providers and patient outcomes [2]. The increasing adoption of digitalized health care records such as electronic health records (EHRs) provides the opportunity for health care data analytics and the coordination of results with care of patients [3,4]. Many health care applications are designed to maximize the potential benefits of EHR usage, such as analyzing epidemiological disease patterns to improve public health across the nation or timely health information exchange (HIE) to provide patients with coordinated and efficient care across health care facilities. Data sharing is needed when required data are distributed and stored in different sources [5,6]. However, multiple barriers to data coordination exist: (1) data privacy and security concerns during HIE [7,8], (2) the limitations of institutional privacy rules [3,9], and (3) the time-consuming process of generating agreements on data exchange between institutions [10,11]. There are security and privacy concerns about the exchange of sensitive health data [8]. Due to the Health Insurance Portability and Accountability Act, legislation limits EHR access without patient authorization [3,12]. Therefore, there needs to be a sustainable and secure data collection mechanism by which each data owner can maintain control of their data and only if the owner of the data allows it to occur [13].

Blockchain Approaches

Facing the challenges of data coordination, blockchain is considered to be a disruptive technology that fits the needs of many health care applications [14,15]. Blockchain is a distributed ledger technology that was first applied in the financial sector [16]. Bitcoin is one of the most popular applications of blockchain that shows its security, durability, and robustness. All the users in the blockchain are anonymous and represented by unique pairs of randomly generated 256-bit public and private keys [16]. This feature of protecting user privacy is one of the reasons for the success of Bitcoin and is also why blockchain is considered for potential health care applications without concerns for patient privacy issues [17]. Similar to the public blockchain used for Bitcoin, blockchain can be implemented privately, also known as a *permissioned chain* for different applications [18]. Users need to gain permission to join the private chain, which limits the data access of the blockchain to only authorized users. Blockchain is a chain of blocks that contain current and former block numbers and validated transactions occurring in a short period [19]. Generation of blocks follows a certain consensus protocol, such as Proof of Work (PoW), which requires miners to provide computing power to validate the legitimacy of the transactions and generate new blocks. Data owners can use blocks to track the timestamps and requesters of their data. All transactions are publicly auditable. Any malicious transaction is expected to be detected by the users in the blockchain and will be discarded thereafter [20]. Once a transaction is made, all the users will validate the identity of the sender and the legitimacy of the transaction. There is no trusted third party to perform the validation process. Rather, the legitimacy of the transaction data accuracy and data provenance is ensured by all the users in a transparent manner [21]. Verified transactions will be written into the subsequent block, and the contents of the transaction can never be erased or changed [16]. The Ethereum blockchain keeps all the original blockchain features and adds a function of programmable self-executing computer protocols to the blockchain system with agreements between the requester and receiver called *smart contracts* [19]. Smart contracts are coded in Turing-complete languages such as Solidity in the Ethereum blockchain that can solve any computational problems [22]. For example, using smart contracts can regulate transactions such as enforcing the interoperability standard of exchanged data [23].

As the features of decentralized transaction validation—ensured data provenance, data sharing, data integration, and flexibility of smart contracts—fit well with the needs of many health care applications, there have been many efforts to apply blockchain to areas of health care, such as HIE, pharmaceutical supply chains, and clinical trial management [24,25]. However, most blockchain applications in the health care area are still in the early stages of implementation [15].

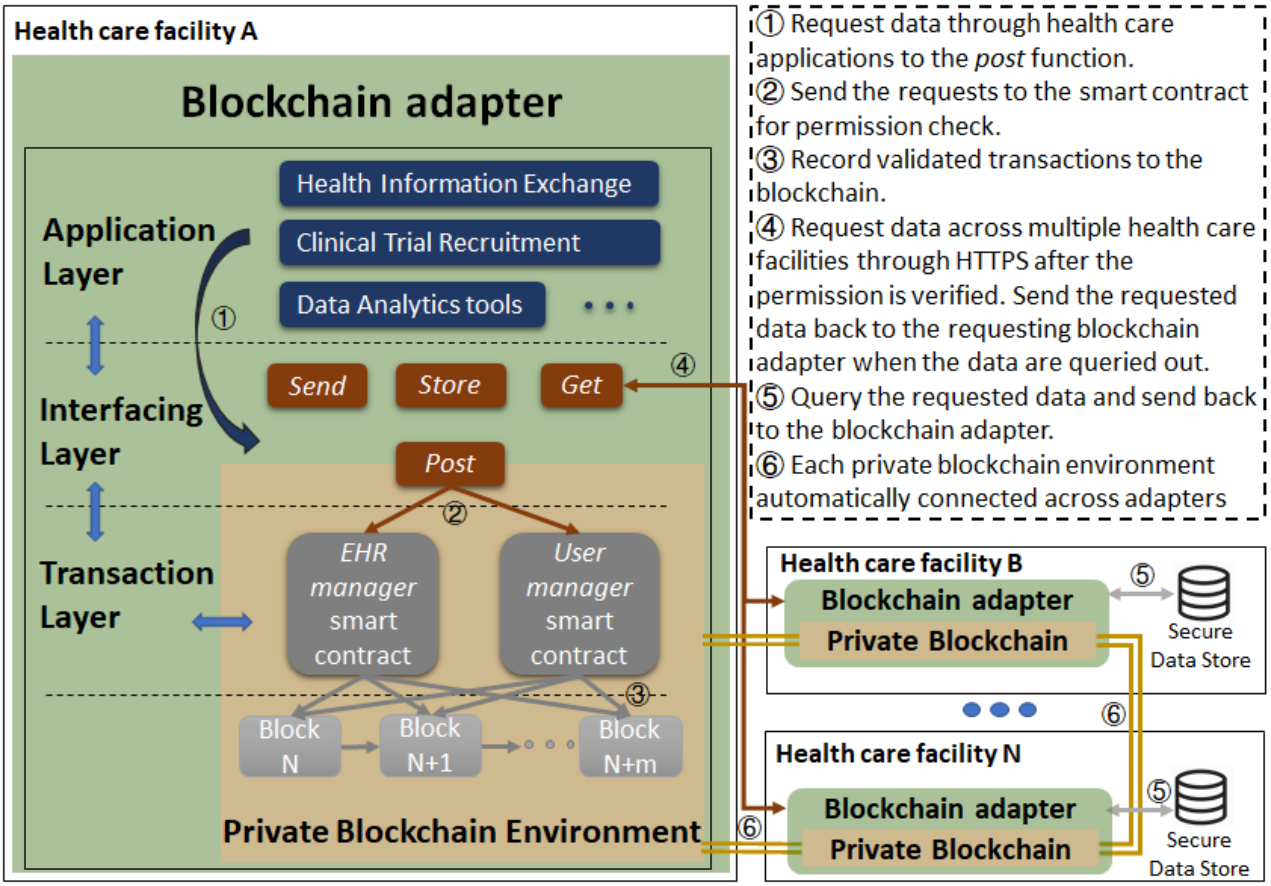
Objectives

This was an extended study of previous blockchain application designs and an exploration of a new generalized layered architecture, as shown in Figure 1. We built a private blockchain environment, implemented the layered architecture, and built 2 prototype applications based on this architecture: HIE and subject recruitment for clinical trials.

Compared with traditional blockchain architecture, this study proposed a generalized blockchain system that fits a wide spectrum of health care applications for cross-site data coordination. The layered architecture provides a blockchain platform with predefined functions for data collection for developers to implement health care applications without an extensive experience of blockchain. We explained the blockchain environment setup and the new layered architecture in the Methods section, followed by case studies and simulation results in the Results section to demonstrate the feasibility, scalability, and compatibility of the architecture for health care application development.

XSL•FO
RenderX

Figure 1. Overall layered blockchain architecture. The transaction layer consists of 2 smart contracts to manage data access tasks. The interfacing layer interacts with the blockchain environment, graphical user interfaces, and other blockchain adapters. The application layer provides a flexible platform for health care application development. The figure shows the general process of data requests using the architecture. EHR: electronic health record.



Methods

Augmented Three-Layered Blockchain Architecture

The blockchain network offers advantages in managing digital assets [26,27]. A well-known digital asset managed by blockchain is Bitcoin or, in general, cryptocurrency [25]. In health care applications, EHR access is a digital asset in management. The health data for patients who opt to participate in the blockchain are encrypted and stored in secured off-blockchain databases located in health care facilities protected by their own firewalls. The foundation private blockchain is used to store all transactions referring to EHR requests and exchanges and the metadata that contain pertinent health care data with the identifications of patients and health care facilities. In addition, the following components are captured in the metadata: time of creation, data set location, access permission and control, data decryption, and data authenticity.

As blockchain is a fully distributed system, we have built 3 layers on the top of the private blockchain network: *transaction layer*, *interfacing layer*, and *application layer*. As shown in Figure 1, the *transaction layer* consists of 2 types of smart contracts coded in Solidity, namely, *EHR manager* smart contracts and *user manager* smart contracts, to manage the storage of and access to metadata that are encrypted with decentralized validity and authenticity checks using blockchain security [28]. These 2 smart contracts are fixed in the system and are not permitted to change.

Data requests from one health facility to another trigger certain functions of smart contracts through the *interfacing layer*. Only through the information in the trustable metadata can original data be retrieved and verified for authenticity by the *interfacing layer*. Applications such as subject recruitment for clinical trials, EHR management, and artificial intelligence (AI)–based data analytics tools can be built on the *application layer*. This layered architecture has the following benefits compared with the previous blockchain systems used for health care applications: (1) compatibility of most health care applications that require data exchange, (2) semipublicity to fix the blockchain settings and smart contract functions but retain most blockchain features, (3) security settings of each layer to protect the identities and data of patients during an exchange, and (4) traceability of who have accessed the data and how they used the data.

Environment Setup

To build the blockchain system, each health care facility is required to provide at least one blockchain node, which is an electronic device that can install the blockchain system. In our setting, each node runs the official Go implementation of the Ethereum protocol on an Ubuntu Linux server. To connect the blockchain with EHR systems, we have developed a blockchain adapter, which is a blockchain node designed to abide by the

XSL•FO
RenderX

local data access policies set by individual health care facilities [29].

To ensure the security of health care data and meet the needs of current EHR operations, our blockchain system does not store patient data. There are two main reasons for this. First, it is not practical to store a large chunk of health care data in a blockchain because of the health care facilities' policies of sharing health information and blockchain storage constraints [25]. Second, the health care industry is still unreceptive to allowing patient data to move across the blockchain network [12]. A metadata set containing pertinent information of the original EHR data is created and submitted to the blockchain platform. The creation and updating of metadata are recorded into a chain of data blocks in the blockchain. These transactions executed via smart contracts are immutable and traceable, thus creating a trustable metadata transaction. As there are different interoperability standards, such as Fast Healthcare Interoperability Resources and Health Level 7 version 3, there will be different metadata points of different data stores for the receiver to choose from. The receiver can choose the compatible interoperability standard of their home department's standard during the HIE process [30,31]. The metadata owner, who is the same as the data set owner, can grant, reject, or revoke access permission automatically via smart contract or interactively by means of electronic notification and confirmation. For example, smart contracts can be programmed to grant or reject access permission based on time or data type or to delegate access permission to a specific user. In all cases, because a third-party intervention is not necessary for granting or rejecting permissions, the time efficiency of data sharing can be greatly improved. This is the decentralized feature of the blockchain network that enables peer-to-peer HIE.

Each blockchain user, including the blockchain adapter, is represented by a hash value (account address) derived from the public key generated by the blockchain [32]. The private key is kept private by the user, and the public key is used for internal and external transactions and communications. Any transaction related to the account address must be signed by the signature, which is the private key. All transactions need to ensure the public key and the private key matched before transactions are recorded in the blockchain. The patients need to go to the health care facilities to opt into the system so that they can claim ownership of their data. A user can create a username and password or use biometric information that is mapped to the public and private key instead of memorizing the real key's value. The metadata permission control carried out by smart contract is anonymous, which ensures privacy. The metadata used for locating encrypted data are communicated with the secured data stores via the https protocol, and the result is communicated back to the user via the same protocol; thus, it is considered to be a secure data transfer.

Foundation Private Blockchain Network

The foundation of our layered architecture is a private Ethereum blockchain, which involves an immutable chain of data blocks consisting of committed ledgers and multiple blockchain nodes synchronously maintaining the same chain of data blocks. In the overall architecture, this layer ensures data immutability, decentralized consensus, data transparency, and traceability. The private blockchain is initiated from a starting node with special settings to make the blockchain unique. The smart contracts are deployed through the starting node when the private blockchain is built. All the participating nodes from health care facilities must obtain permission from the starting node to join the system. This procedure will disallow unauthorized parties from joining the system. As the participating nodes joined into the system, the blockchain will automatically generate accounts for their blockchain adapters. All other users, such as patients and health care providers, need to opt into the system through health care facilities. The blockchain accounts will be generated for the users from each health care facility as soon as the applicants' identities are proven.

The private blockchain stores all the transactions for (1) patients and health care facilities granting, revoking, and denying access to their EHR data; (2) authentication of patients and health care providers to retrieve the EHR data; and (3) health care facilities to store metadata for patient visits. The transactions will record the receiver, sender, contained data, and the timestamp into the blocks through blockchain adapters. Users can also make transactions in the backend blockchain console through the blockchain node. These transactions still need to pass smart contract regulations to become effective. Most users will interact with the graphical user interfaces (GUIs) built on the *application layer* to execute functions in the blockchain system.

Transaction Layer

The *transaction layer* consists of 2 smart contracts that specify a metadata model for health care records and methods that regulate data access rights, permission policies, and data encryption. Two smart contracts, the *EHR manager* smart contract and the *user manager* smart contract, are deployed to the blockchain to securely accomplish the basic EHR management tasks. The *EHR manager* smart contract can only be used by health care facilities to submit the EHR metadata to the blockchain. The *user manager* smart contract is used by patients or facilities to manage access to their data. Once a patient has opted into the system from a health care facility, the health care facility's blockchain adapter will automatically encrypt his or her patient ID and public key using the patient's private key and input to the *user manager* smart contract. Health care facilities will have adapter IDs stored in the *user manager* smart contract. The following scenarios demonstrate the use of smart contracts for the EHR metadata input and HIE.

Electronic Health Record Manager Smart Contract for Submitting Metadata to the Blockchain

The *EHR manager* smart contract (as shown in Figure 2) defines several structures to record patient information: EHRDataID and EHRdata define the metadata components, PatientID stores the patient ID and health care facility ID for the registered patients, and patientData maps the different health care facility visit records of patients with the patients' IDs. Once the blockchain adapter receives a record from the EHR system, the blockchain adapter automatically performs the following steps to submit the metadata to the blockchain through an addEHR function:

XSL•FO
RenderX

Figure 2. Main part of the electronic health record manager smart contract code that defines the metadata structure. Blockchain adapters must extract the information and calculate the encrypt keys and then store them into the smart contract. The record will automatically associate with the adapter's blockchain ID.

```
pragma solidity ^0.4.13;
contract EHRMgrContract {

struct EHRDataID {
    bytes32 datasetID;      // Dataset ID
    string datasetDesp;     // Dataset description
    bytes32 timeOfCreation;     // dataset creation time
}

struct patientID {
    bytes32 patient_ID;     // patient ID in a hospital
    bytes32 hospital;       // hospital ID
    bytes32 patientName;        // patient name
}

struct EHRData {
    EHRDataID   ehrdataID;  // dataset ID struct
    string datasetLoc;      // an URL or a data object hash
    bytes32 keycipher;      // cipher of the encryption key for EHR
    address bcAdapter;      // blockchain adapter address
}
struct patientData {
    patientID patientid;        // patientID in a hospital
    mapping (bytes32 => EHRData) ehrCollection;    // all EHRs for a patient in a hospital
}

mapping (address => patientData ) patientDataSet; // user info for the EHR datasets

function addPatient(bytes32 _patient_ID, bytes32 _hospital, bytes32 _patientName, address _patient) public {
    patientDataSet[_patient].patientid.patient_ID=_patient_ID;
    patientDataSet[_patient].patientid.hospital=_hospital;
    patientDataSet[_patient].patientid.patientName=_patientName;
}// to add a patient to blockchain

function addEHR(string _loc, bytes32 _keycipher, bytes32 _datasetID, string _datasetDesp, bytes32 _timeOfCreation, address _patient) public{
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.datasetID=_datasetID;
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.datasetDesp=_datasetDesp;
     patientDataSet[_patient].ehrCollection[_datasetID].ehrdataID.timeOfCreation=_timeOfCreation;
     patientDataSet[_patient].ehrCollection[_datasetID].datasetLoc=_loc;
     patientDataSet[_patient].ehrCollection[_datasetID].keycipher=_keycipher;
     patientDataSet[_patient].ehrCollection[_datasetID].bcAdapter=msg.sender;
}// to submit an EHR metadata to blockchain

}
```

1. Extract the patient ID from the EHR data set.
2. Find a public key associated with the patient from the *user manager* smart contract.
3. Generate a random *data key* for encrypting the EHR data set.
4. Encrypt the EHR data set using data key and store the encrypted data to an *off-chain* secured data store.
5. Use the patient's public key to encrypt the data key. Call the encrypted data key "key cipher."
6. Submit the following metadata to the blockchain:
    - Patient ID.
    - Encrypted data set location as a URL.
    - Key cipher.
    - Associated blockchain adapter ID.

The blockchain adapters will generate public and private key pairs following the Diffie-Hellman protocol using Node.js. The data encryption and decryption settings ensure that the data belonging to the patient is only decrypted by the health care facility that produced the data. The key cipher makes the data key more secure and can only be computed when the data owner is known. Figure 3 shows one patient's decoded metadata retrieved through blockchain using the Remix integrated development environment, which is an open-source visualization tool used for interacting with blockchain nodes and smart contract development and deployment.

XSL•FO
RenderX

Figure 3. Example of a patient's metadata retrieved through blockchain. The information is decoded by the Remix web-based integrated development environment, which is connected to the retrieving blockchain node. The patient's metadata contain data location, key cipher, data set ID, data set description, and creation time.

transaction hash	call0x9a095c6a3c41390172232fa6e2441bb07120fd7a0x21966efeaa8e307a aa22d819639abf8b8fd1a4520x0a79309b0000000000000000000000009a095c 6a3c41390172232fa6e2441bb07120fd7a
from	0x9a095c6a3c41390172232fa6e2441bb07120fd7a
to	EHRMgrContract.retrieve(address) 0x21966efeaa8e307aaa22d819639ab f8b8fd1a452
hash	call0x9a095c6a3c41390172232fa6e2441bb07120fd7a0x21966efeaa8e307a aa22d819639abf8b8fd1a4520x0a79309b0000000000000000000000009a095c 6a3c41390172232fa6e2441bb07120fd7a
input	0x0a7...0fd7a
decoded input	{ "address _patient": "0x9A095C6a3c41390172232Fa6e2441BB07 120fD7A" }
decoded output	{ "0": "string: mongo --host 128.206.20.167 -d tmp --port 27017 --username 39791 "1": "string: e0697b0c5c0dcebc95", "2": "string: 08838", "3": "string: Diagnosis:C508", "4": "string: 2019-07-28" }

User Manager Smart Contract for the Health Information Exchange Process

A scenario is described in this section to show the process of health care provider X retrieving patient A's EHR. Through a mobile application or web browser with biometric authentication, Patient A can grant access privileges to health care provider X using the application program interface (API) to set data permission on the *interface layer*. In some cases, more than one clinician is involved in the patient's care. The health care facility needs to create a shared blockchain account for the provider's department so that all involved clinicians can access the patient's data with one-time authentication [33]. Blockchain adapters will record who has accessed the data and submit it to the blockchain. The blockchain adapters from the receiver's home health care facility will perform the following steps for the data retrieval process:

1. Verify health care provider X's permission to access patient A's records through the *EHR manager* smart contracts.
2. Retrieve patient A's metadata from the *EHR manager* smart contract.
3. Request the encrypted EHR data set from the remote health care facility via an https service provided by blockchain adapter.
4. Retrieve the encrypted data using encrypted data set location information in EHR metadata.
5. Decrypt the EHR data set. This step involves decrypting the key cipher using the patient's private key to obtain the data key that decrypts the EHR data set.

Similar processes will be used for sharing data between health care facilities and for patients retrieving their own EHR records. The entire process will be performed automatically through the blockchain adapter.

Interfacing Layer

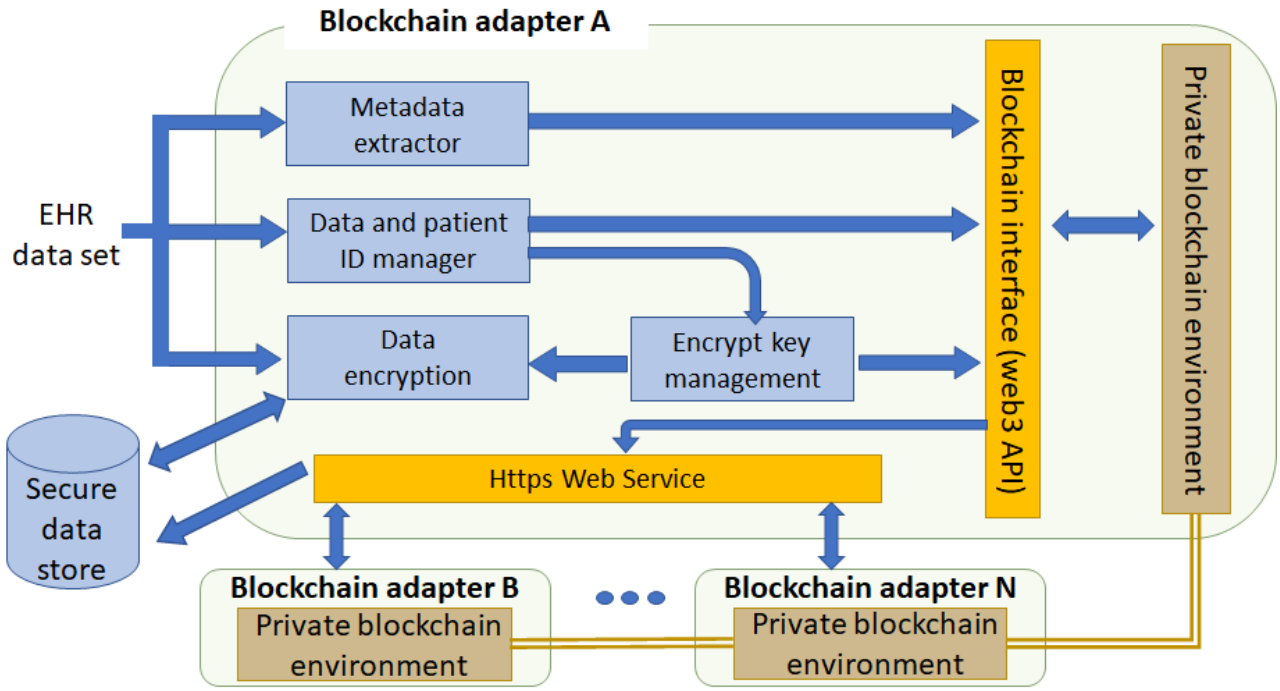
The *interfacing layer* provides 4 high-level methods: *get* the health care data from different facilities, *store* the encrypted data securely, *post* metadata or data request to the blockchain via smart contracts in the *transaction layer*, and *send* the encrypted data to the receiver who has gained permission from the data owner. Using the functions in this layer, application developers can implement distributed data applications (DApps) without the knowledge of smart contracts and the underlying blockchain network. This layer consists of APIs and https web services to define a set of primitive coordinate functions: (1) submit data, (2) set data permission, and (3) retrieve data. The data submission API will extract metadata from the original data and call the *transaction layer*'s smart contracts to record it to the blockchain. It will also encrypt the original data and store the encrypted version to a secure off-blockchain data store. The data retrieval API will call smart contracts to retrieve metadata from blockchain, verify encrypted data authenticity with metadata, and decrypt the encrypted data in the off-blockchain data store to obtain the original data. The data

XSL•FO
RenderX

permission setting API will call the *transaction layer*'s smart contracts to set access policies and methods for a piece of metadata. Information contained in metadata is used to retrieve and decrypt data. The https web services provide secure data transport when data are to be transported through a channel and can potentially be eavesdropped. Using the blockchain adapter to serve as a gateway to the EHR system minimizes the concerns of data exchange security.

We implemented blockchain adapters as a Node.js application and used the web3.js package for interfacing with a blockchain node and https.js package for https secure web services. web3.js is also available in the Python library as web3.py. The https-based web services are mainly used for communication among blockchain adapters. The blockchain adapter is embedded as software that will install the missing component automatically, such as node.js and web3.js. Figure 4 shows a high-level block diagram of a blockchain adapter. The metadata extractor extracts metadata such as patient ID and data set ID from the EHR data set for data identification purposes in the blockchain. The data and patient ID manager maps the patient ID to the data set ID and records the information in both the *user manager* smart contract as well as the *EHR manager* smart contract. The data set encryption block in Figure 4 encrypts the EHR data set and stores away in the secure data store with a URL or a data object hash for future access.

Figure 4. Blockchain adapter components and functions. Blockchain adapter extracts the metadata from the electronic health record, encrypts the electronic health record, stores the encrypted electronic health record into the secured data store, and maps the patient ID and data ID to the blockchain account. Blockchain adapters use https to interact with other adapters and communicate with the foundation private blockchain network through a blockchain interface. API: application program interface; EHR: electronic health record.



From the security and reliability point of view, the following design guidelines are strictly followed for the blockchain adapter:

1. Blockchain adapter is modeled as a nonhuman blockchain user and has its own private and public key pair when the blockchain account or address is established. The public key is made public via the *user manager* smart contract. The private key is kept in a blockchain adapter.
2. The https service uses a separate key and certificate file.
3. The Diffie-Hellman key agreement protocol is used for data set encryption.
4. When one blockchain adapter fails, transactions (data set exchange) with the associated organization will be interrupted, but transactions among all other organizations will not be affected.

Application Layer

With the above *interfacing layer* architecture and smart contract setup, many health care applications involving data exchange can be developed in the *application layer*. This layer will rely on the *interfacing layer* to securely collect the data and then perform data analytics. Applications will not change the existing blockchain settings. These applications can allow researchers or data owners to have better use of the EHR data. For example, personal health records management can be developed on the *application layer*. A patient's identity will be verified through smart contracts in the blockchain. All patient records can be retrieved through the *interfacing layer*. In addition to the HIE application, subject recruitment for clinical trials could also be developed in this layer. Clinical trial sponsors need to obtain permission from the patients through blockchain adapters from clinical trial sites before the matching process [34]. After the patients grant the sponsors permission to gather their data, clinical trial sponsors can use the data analytics tools developed in the *application layer* to match the patients with their recruitment criteria automatically. We implemented these 2 sample applications on our private blockchain system. The

XSL•FO
RenderX

Results section shows the interactions of the *application layer* and the blockchain system and the simulation results.

Results

Case Studies

To test the feasibility of our layered architecture, we built a blockchain environment that contains 1 starting node and 4 health care facility nodes. All the nodes have installed an Ubuntu 18.04.2 operating system and Go Ethereum with the default PoW setting. A blockchain adapter has been installed to each node to communicate with the blockchain and its own secured data store, which was implemented using MongoDB. We created 100 accounts for health care providers and 10,000 patient accounts on each health care facility node of patients' records from the Surveillance, Epidemiology, and End Results database [35]. A total of 2431, 2587, and 2505 patients have multiple records distributivity stored in 2, 3, and 4 health care facility nodes, respectively. The remaining patients' records were stored in a single facility node. Selected records were stored in each secured data store using an automated script following the same procedure described in the *EHR manager* smart contract. The ownerships have been claimed when the metadata were pushed into the blockchain using the smart contract in the *transaction layer*.

After setting up the environment, we implemented 2 applications to interact with our blockchain system. The applications are built off-chain but can communicate with the blockchain system using node.js. These 2 applications are examples of the potential use of the layered blockchain architecture. We tested the accuracy, scalability, and speed of our system. We made the following assumptions to simulate the 2 processes: (1) each clinical site has provided at least one node to the system, (2) each clinical site agrees to connect the blockchain adapter to the secured data store, and (3) patients have authorized the blockchain system as well as the 2 applications to access their health records.

Health Information Exchange

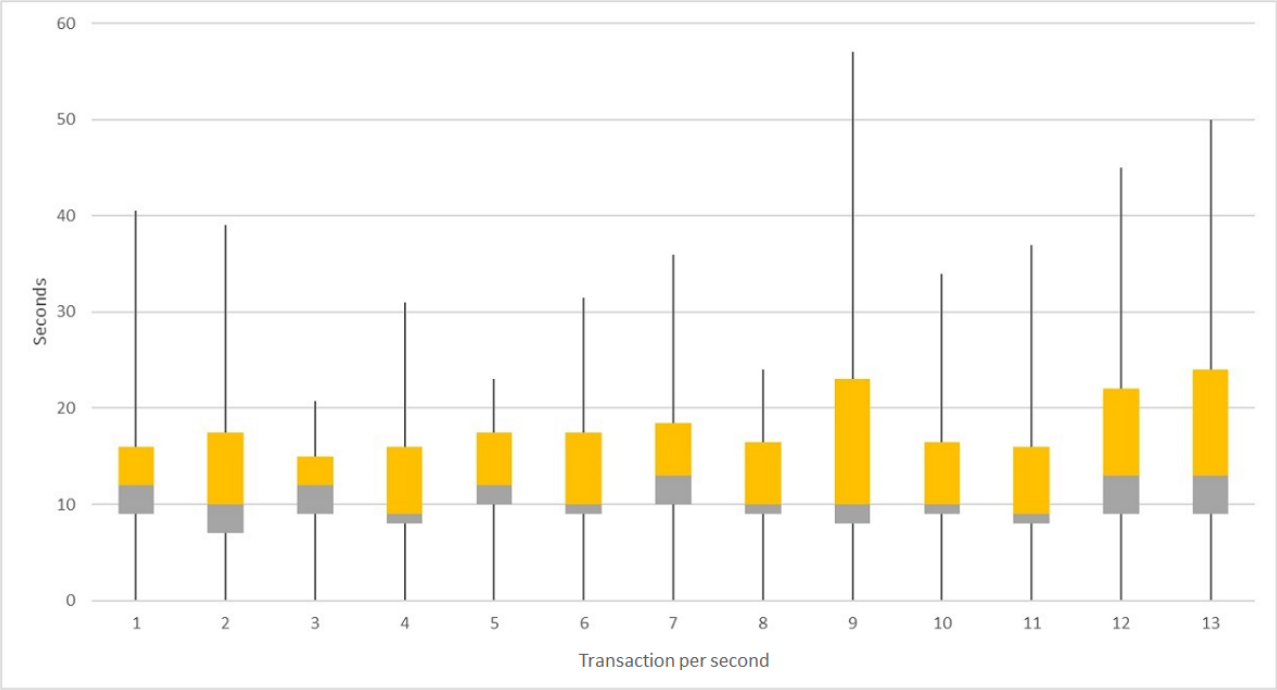
This application provides an interface for users to manage access to personal health care records through the *transaction layer*. Patients can use this application to grant and revoke access to their records. Patients can also track how many times their records have been accessed through this aplication. To test the accuracy, speed, and scalability of our system, we simulated the process of patients granting permission to health care providers of their EHR. We developed 5 scripts to automate the simulation process by (1) randomly selecting 1 patient to grant 1 health care provider permission to their EHR per second for an hour and recording the timestamp, (2) recording the timestamp when the health care providers received the permission, (3) recording the timestamp when the health care providers received the data, and (5) adding 1 patient per second to script 1 then repeating scripts 1 to 4 until reaching the system limitation because of the known scalability constraints of Ethereum [36].

The simulation only contains the period of interactions with the blockchain. Retrieving data is an off-chain process through the https portal and varies with different health care facilities. From our simulation results, the system breaks at a certain point when the scale reaches 14 transactions per second (TPS). We simulated 331,142 access-granting transactions. All the transactions have successfully retrieved the records except the last second's 14 transactions due to reaching the Ethereum scalability limitation. The average time of writing a transaction to a block is 11.271 (SD 2.208) seconds. We did not find a correlation between TPS and validation time. All health care providers received the metadata in an average of 1.73 seconds.

In this study, the scalability of the blockchain using various transaction frequencies from 1 to 14 TPS through blockchain adapters was tested. Figure 5 shows the time spent granting permission from different scales (the 14 TPS group was excluded because of incomplete results). Once the permission is granted by writing the transactions into blocks, the receiver can retrieve the metadata from the smart contract through the blockchain adapter without making another transaction for users to validate the legitimacy. This means that the average time of receiving metadata is much shorter than the grant permission. The script of the 9 TPS group runs slowly compared with the former groups. All the blockchain nodes were restarted separately, and the script was restarted with the 10 TPS group. The speed is affected by the processing speed of the blockchain nodes and Ethereum performance. The starting node's blockchain adapter was used to control the overall frequency. All transactions from the blockchain adapters of health care facility nodes will queue in the starting node's adapter until the earlier batches of transactions have been executed by each blockchain adapter. We controlled the overall frequency as 13 TPS, which avoids the Ethereum's scalability constraints by spacing the transactions.

XSL•FO
RenderX

Figure 5. Box plot of simulation results for 1 to 13 transactions per second groups. The boxes show the different quantiles of time cost related to different scales of transactions per second.



Subject Recruitment for Clinical Trials

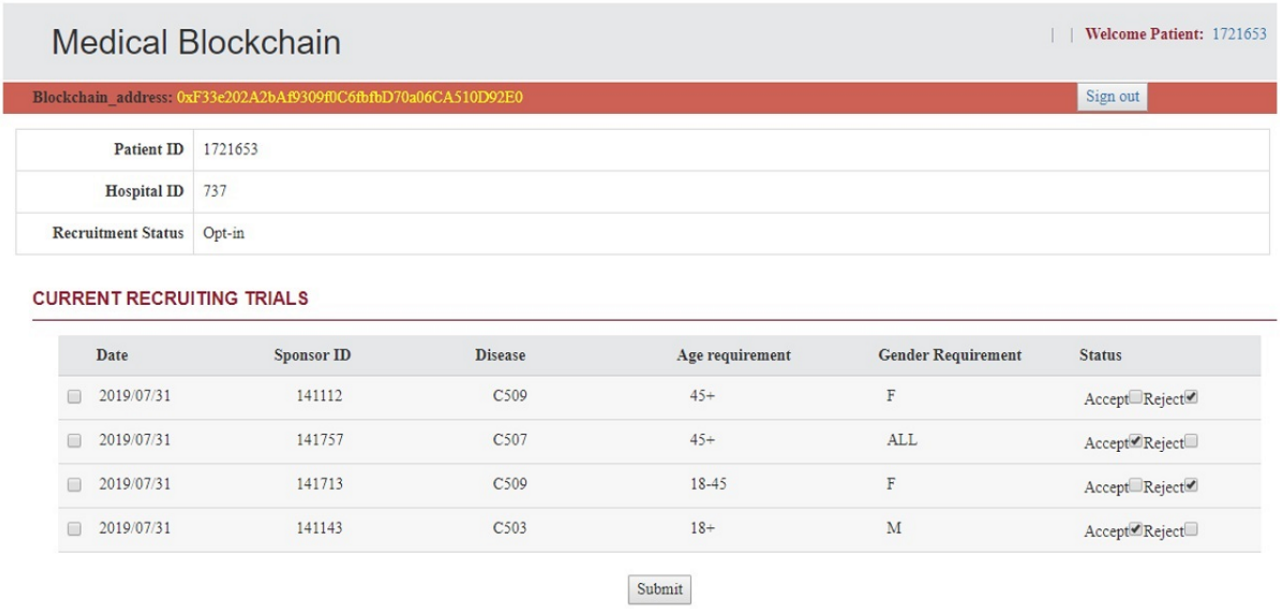
To prove the compatibility of our layered blockchain system, another application was implemented as subject recruitment for clinical trials. This application involves posting criteria, granting permission, and data exchange. To use this application, each clinical trial sponsor needs to provide at least one blockchain node with a blockchain adapter installed. All patients who want to participate in clinical research are required to opt into the application so that they can receive the recruitment criteria. This application involves the following procedures:

- Clinical research recruiters will send recruitment criteria to the opted-in patients in the *interfacing layer*. The criteria also contain the recruiters' blockchain address used for patient authentication.
- Patients who received the criteria can check whether they are matched. If they are interested in clinical research, they can authenticate the clinical research recruiter to access their EHR.
- Recruiters will be notified by the blockchain of whether any patients have granted their permissions to access their data. Then, they will retrieve patient data for precise matching.
- The recruiters will notify the patients if they are matched and send further instructions.

We have implemented a GUI to demonstrate the usage of this application. Patients need to opt for this application to receive the current clinical trial recruitment information. We randomly selected 4 sponsors to post their recruitment criteria using their blockchain adapters through the blockchain *interfacing layer*. Only patients who opted will receive recruiting clinical trial information. Figure 6 shows the GUI for a simulated patient with ID 1721653. It will provide basic information for the patient, such as blockchain address and registered health care facility ID. The recruiting criteria include the sponsor ID, basic inclusion criteria such as age and gender, and study disease (in this case, we used the primary site for the study disease in our simulation). Patients can use the GUI to grant the sponsors access to receive their data to have a precise match with their clinical trial criteria. This action will automatically send a transaction to the *user manager* smart contract to add the sponsor's blockchain address to their access list. Through our simulation, 2 sponsors successfully received permission in an average of 3.07 seconds.

XSL•FO
RenderX

Figure 6. Graphical user interface for patients to permit clinical trial sponsors to access their data. The graphical user interface shows the patient's blockchain address, opt-in status, current recruiting trials' basic recruitment criteria, and permission options. When the patients select the recruiting trial and click submit, the sponsor will be added to their access list in the USER manager smart contract through the blockchain adapter.



Discussion

Conclusions and Future Work

This paper described an augmented layered blockchain system in development for most health care applications involving data coordination across multiple health care facilities. The design of this layered architecture provides generic functions and methods for application developers to securely collect data from different sources without requiring extensive experience of the blockchain technology. The layered architecture allows users the ability to audit the legitimacy of previously occurring transactions but prevents users from modifying any components in the blockchain. The features of blockchain provide a solution to current data coordination challenges. The blockchain-based approach extends the ownership of the EHR data set to each patient. On the basis of decentralized features of blockchain technology for peer-to-peer transactions, this approach can greatly reduce the health care data set sign-off and release. Data security and authenticity are also guaranteed by the immutability of the blockchain and smart contract–regulated data exchange.

Through our simulation process, our system empirically proved the feasibility of the architecture for health care applications. We also tested the scalability of our blockchain system and provided an optimal solution to avoid blockchain scalability constraints. Our future work will continue to evaluate the validation mechanism to improve the blockchain performance and add an AI component on the *application layer* for data analytics to maximize the use of EHRs based on the layered blockchain system.

Limitation

The main limitation of our approach is the setup requirement from each participating site. Each health care facility is required to provide at least one blockchain node to the system and keep an encrypted EHR outside of the operation EHR for registered patients into the secured data stores that can communicate with blockchain adapters. Patients will potentially also need to provide blockchain nodes such as mobile devices to exchange and store their personal health records generated by their personal medical devices. The performance of the model can be affected by the properties of the blockchain node. If a single node creates a mass of transactions at the same time through the blockchain adapter, this action will use up all the memory and break the node before it is sent to the blockchain.

Acknowledgments

This research was supported in part by the University of Missouri Informatics and Data Science Research Initiatives (YZ and CS), the Ministry of Science and Technology in Taiwan (MOST) under grant numbers MOST 108-2634-F-468-001, and MOST 108-2511-H-468 -002 from through Pervasive Artificial Intelligence Research Laboratories (YC and ZS).

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in US hospitals. N Engl J Med 2009 Apr 16;360(16):1628-1638. [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
2. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
3. Weng C, Appelbaum P, Hripcsak G, Kronish I, Busacca L, Davidson KW, et al. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc 2012;19(5):684-687 [FREE Full text] [doi: 10.1136/amiajnl-2012-000878] [Medline: 22542813]
4. Jha AK. Meaningful use of electronic health records: the road ahead. J Am Med Assoc 2010 Oct 20;304(15):1709-1710. [doi: 10.1001/jama.2010.1497] [Medline: 20959581]
5. Dimitropoulos L, Patel V, Scheffler S, Posnack S. Public attitudes toward health information exchange: perceived benefits and concerns. Am J Manag Care 2011 Dec;17(12 Spec No):SP111-SP116 [FREE Full text] [Medline: 22216769]
6. Shapiro JS, Mostashari F, Hripcsak G, Soulakis N, Kuperman G. Using health information exchange to improve public health. Am J Public Health 2011 Apr;101(4):616-623. [doi: 10.2105/AJPH.2008.158980] [Medline: 21330598]
7. Fernandes LM, O'Connor M, Weaver V. Big data, bigger outcomes: healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis. J AHIMA 2012 Oct;83(10):38-43; quiz 44. [Medline: 23061351]
8. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
9. Sharyl J, Laura A, Lawrence O. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: The National Academies Press; 2009.
10. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010;17(3):288-294 [FREE Full text] [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
11. Buterin V. A Next Generation Smart Contract & Decentralized Application Platform. The Blockchain Lab. 2014. URL: https://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf [accessed 2020-06-03]
12. Fernández-Alemán JL, Señor IC, Lozoya PO, Toval A. Security and privacy in electronic health records: a systematic literature review. J Biomed Inform 2013 Jun;46(3):541-562 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.003] [Medline: 23305810]
13. Ray P, Wimalasiri J. The Need for Technical Solutions for Maintaining the Privacy of EHR. In: Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society. 2006 Presented at: IEMBS'06; August 30-September 3, 2006; New York, NY, USA. [doi: 10.1109/iembs.2006.260862]
14. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 1;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
15. Kuo T, Rojas HZ, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 1;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
16. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. New York, USA: BN Publishing; 2008.
17. Zhang P, Schmidt D, White J, Lenz G. Blockchain technology use cases in healthcare. Adv Comput 2018;111:1-41. [doi: 10.1016/bs.adcom.2018.03.006]
18. Shrier D, Wu W, Pentland A. Blockchain & Infrastructure (Identity, Data Security). GetSmarter: Online Courses with the World's Top Universities. 2016. URL: https://www.getsmarter.com/blog/wp-content/uploads/2017/07/mit_blockchain_and_infrastructure_report.pdf [accessed 2020-06-03]
19. Dannen C. Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners. Berkeley, CA: Apress; 2014.
20. Atzei N, Bartoletti M, Cimoli T. A Survey of Attacks on Ethereum Smart Contracts (SoK). In: Proceedings of the International Conference on Principles of Security and Trust. 2017 Presented at: POST'17; April 24-25, 2017; Uppsala, Sweden. [doi: 10.1007/978-3-662-54455-6_8]
21. Zyskind G, Nathan O. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In: IEEE Security and Privacy Workshops. 2015 Presented at: SPW'15; May 21-22, 2015; San Jose, CA, USA. [doi: 10.1109/SPW.2015.27]
22. Dannen C. Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners. New York, USA: Springer; 2017.
23. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRchain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
24. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]

XSL•FO
RenderX

25. Zhuang Y, Sheets L, Shae Z, Tsai JJ, Shyu C. Applying blockchain technology for health information exchange and persistent monitoring for clinical trials. AMIA Annu Symp Proc 2018;2018:1167-1175 [FREE Full text] [Medline: 30815159]
26. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: cochrane systematic review and meta-analysis. BMJ Open 2013;3(2):e002360 [FREE Full text] [doi: 10.1136/bmjopen-2012-002360] [Medline: 23396504]
27. Karafiloski E, Mishev A. Blockchain Solutions for Big Data Challenges: a Literature Review. In: 17th International Conference on Smart Technologies. 2017 Presented at: EUROCON'17; July 6-8, 2017; Ohrid, Macedonia. [doi: 10.1109/eurocon.2017.8011213]
28. Crosby M, Pattanayak P, Verma S, Kalyanaraman V. Blockchain Beyond Bitcoin: Blockchain Technology Challenges and Real-World Applications. In: International Conference on Computing, Electronics & Communications Engineering. 2018 Presented at: iCCECE'18; August 16-17, 2018; Southend, United Kingdom. [doi: 10.1109/iCCECOME.2018.8658518]
29. Hripcsak G, Bloomrosen M, FlatelyBrennan P, Chute C, Cimino J, Detmer D, et al. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 health policy meeting. J Am Med Inform Assoc 2014;21(2):204-211 [FREE Full text] [doi: 10.1136/amiajnl-2013-002117] [Medline: 24169275]
30. Gordon WJ, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
31. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful Approach to Healthcare Information Exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013 Presented at: CBMS'13; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627810]
32. Zhuang Y, Sheets L, Chen YW, Shae Z, Tsai JJ, Shyu CR. A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform 2020 May 8;152(2014):-. [doi: 10.1109/JBHI.2020.2993072] [Medline: 32396110]
33. Gummadi S, Housri N, Zimmers TA, Koniaris LG. Electronic medical record: a balancing act of patient safety, privacy and health care delivery. Am J Med Sci 2014 Sep;348(3):238-243. [doi: 10.1097/MAJ.0000000000000287] [Medline: 24879530]
34. Zhuang Y, Sheets LR, Shae Z, Chen Y, Tsai JJ, Shyu C. Applying blockchain technology to enhance clinical trial recruitment. AMIA Annu Symp Proc 2019;2019:1276-1285 [FREE Full text] [Medline: 32308925]
35. SEER Cancer Statistics Review (CSR) 1975-2017. Surveillance, Epidemiology, and End Results Program. 2020. URL: https://seer.cancer.gov/csr/1975_2017/ [accessed 2020-06-03]
36. Linn L, Koo M. Blockchain For Health Data and Its Potential Use in Health IT and Health Care Related Research. Office of the National Coordinator for Health Information. 2016. URL: https://www.healthit.gov/sites/default/files/11-74-ablockchainforhealthcare.pdf [accessed 2020-06-03]

Abbreviations

AI: artificial intelligence
API: application program interface
EHR: electronic health record
GUI: graphical user interface
HIE: health information exchange
MOST: Ministry of Science and Technology in Taiwan
PoW: Proof of Work
TPS: transactions per second

Edited by K Clauson, P Zhang; submitted 31.03.20; peer-reviewed by TT Kuo; comments to author 26.04.20; revised version received 03.05.20; accepted 08.05.20; published 27.07.20.

Please cite as:
Zhuang Y, Chen YW, Shae ZY, Shyu CR
Generalizable Layered Blockchain Architecture for Health Care Applications: Development, Case Studies, and Evaluation
J Med Internet Res 2020;22(7):e19029
URL: https://www.jmir.org/2020/7/e19029
doi:10.2196/19029
PMID:32716300

©Yan Zhuang, Yin-Wu Chen, Zon-Yin Shae, Chi-Ren Shyu. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution

XSL•FO
RenderX

License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment

Amanda L Graham[1,2], PhD; George D Papandonatos[3], PhD; Megan A Jacobs[1], MPH; Michael S Amato[1,2], PhD; Sarah Cha[1], MSPH; Amy M Cohn[4], PhD; Lorien C Abroms[5], PhD; Robyn Whittaker[6], PhD

[1]Innovations Center, Truth Initiative, Washington, DC, United States

[2]Mayo Clinic College of Medicine and Science, Rochester, MN, United States

[3]Center for Statistical Sciences, Brown University, Providence, RI, United States

[4]Oklahoma Tobacco Research Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[5]Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

[6]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

Corresponding Author:
Amanda L Graham, PhD
Innovations Center
Truth Initiative
900 G Street, NW
4th Floor
Washington, DC, 20001
United States
Phone: 1 12024545938
Email: agraham@truthinitiative.org

Related Article:

Correction of: https://www.jmir.org/2020/4/e17734/

(J Med Internet Res 2020;22(7):e21027) doi:10.2196/21027

The authors of "Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment" (J Med Internet Res 2020;22(4):e17734) noticed several errors in their published manuscript which had been introduced after proofreading. The following corrections have been implemented:

The symbols μ and ϕ were presented as m and f, respectively, in the following sentence of the Methods section; additionally, in this section, the negative sign was incorrectly subscripted:

> *SMDs for frequency counts were calculated as* $(m_{1-}m_2)/[f(m_1+m_2)]^{1/2}$*, where* m_1 *and* m_2 *were the sample means of each comparison group.*

This has been revised to:

> *SMDs for frequency counts were calculated as* $(\mu_1-\mu_2)/[\phi(\mu_1+\mu_2)]^{1/2}$*, where* μ_1 *and* μ_2 *were the sample means of each comparison group.*

In the same paragraph, the negative sign was again incorrectly subscripted in the following sentence:

> *SMDs for binary outcomes were calculated as* $(p_{1-}p_2)/[p_1\times q_1+p_2\times q_2]^{1/2}$*, where* $p_1=1-q_1$ *and* $p_2=1-q_2$ *were the sample outcome prevalence of each comparison group.*

This has been revised to:

> *SMDs for binary outcomes were calculated as* $(p_1-p_2)/[p_1\times q_1+p_2\times q_2]^{1/2}$*, where* $p_1=1-q_1$ *and* $p_2=1-q_2$ *were the sample outcome prevalence of each comparison group.*

Additionally, due to a technical error the following sentence was published in the Results section of the Abstract:

> *As no SMD >0.30 was observed for main effects on any outcome, results suggest that for some outcomes, the combined intervention was stronger than individual factors alone.*

This has been revised to:

> *As no SMD >0.30 was observed for main effects on any outcome, results suggest that for some outcomes,*

XSL•FO
RenderX

the combined intervention was stronger than individual factors alone.

The correction will appear in the online version of the paper on the JMIR website on July 28, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 03.06.20; this is a non–peer-reviewed article;accepted 03.06.20; published 28.07.20.

Please cite as:
Graham AL, Papandonatos GD, Jacobs MA, Amato MS, Cha S, Cohn AM, Abroms LC, Whittaker R
Correction: Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment
J Med Internet Res 2020;22(7):e21027
URL: https://www.jmir.org/2020/7/e21027
doi:10.2196/21027
PMID:32721924

©Amanda L Graham, George D Papandonatos, Megan A Jacobs, Michael S Amato, Sarah Cha, Amy M Cohn, Lorien C Abroms, Robyn Whittaker. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework

Claire Reidy[1,2], PhD; Claire Foster[3], PhD; Anne Rogers[1], PhD

[1]National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care, School of Health Sciences, Faculty of Environmental & Life Sciences, University of Southampton, Southampton, United Kingdom

[2]School of Primary Care, Population Health & Medical Education, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[3]Macmillan Survivorship Research Group, School of Health Sciences, Faculty of Environmental & Life Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Claire Reidy, PhD
National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care
School of Health Sciences, Faculty of Environmental & Life Sciences
University of Southampton
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 44 23 8059 7628
Email: c.reidy@soton.ac.uk

Related Article:

Correction of: https://www.jmir.org/2020/5/e13980/

(J Med Internet Res 2020;22(7):e21381) doi:10.2196/21381

In "A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework" (J Med Internet Res 2020;22(5):e13980) the authors noted several errors.

In the Results section of the Abstract, the text was revised from:

> *"(4) professional responsibility and associated risks and dangers, whereas HCPs are fearful of the consequences of promoting non-NHSSM support, and they question whether SM support fits into their role."*

to

> *"(4) professional responsibility and associated risks and dangers, whereas HCPs are fearful of the consequences of promoting non-NHS SM support, and they question whether SM support fits into their role."*

In the Discussion section, one sentence has been updated to correct reference citations. The text was revised from:

> *"In addition, there has been a recent drive for the integration of psychosocial support into routine diabetes care [19,22], and this study provides an initial engagement with the factors that would impact how psychosocial support is taken up with HCPs and the priorities for patients."*

to

> *"In addition, there has been a recent drive for the integration of psychosocial support into routine diabetes care [18,21], and this study provides an initial engagement with the factors that would impact how psychosocial support is taken up with HCPs and the priorities for patients."*

Multimedia Appendices 1 and 2 contained tracked changes comments. These files have now been replaced and no longer contain the comments.

The corrections will appear in the online version of the paper on the JMIR Publications website on July 30, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories

XSL•FO
RenderX

Submitted 12.06.20; this is a non–peer-reviewed article;accepted 12.06.20; published 30.07.20.

Please cite as:
Reidy C, Foster C, Rogers A
Correction: A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework
J Med Internet Res 2020;22(7):e21381
URL: http://www.jmir.org/2020/7/e21381/
doi:10.2196/21381
PMID:32730220

©Claire Reidy, Claire Foster, Anne Rogers. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Authorship Correction: The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology

Mario Alonso Puig[1], MD; Mercedes Alonso-Prieto[2], MD; Jordi Miró[3], PhD; Raquel Torres-Luna[2], MSN; Diego Plaza López de Sabando[4], MD; Francisco Reinoso-Barbero[2,5], PhD

[1]Juegaterapia Foundation, Madrid, Spain

[2]Pediatric Pain Unit, Anesthesiology-Critical Care Service, University La Paz Hospital, Madrid, Spain

[3]Department of Psychology, Unit for the Study and Treatment of Pain-ALGOS, Rovira i Virgili University, Tarragona, Spain

[4]Pediatric Hemato-oncology Service, University La Paz Hospital, Madrid, Spain

[5]Department of Anatomy-Histology and Neuroscience, School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain

Corresponding Author:
Mercedes Alonso-Prieto, MD
Pediatric Pain Unit
Anesthesiology-Critical Care Service
University La Paz Hospital
Paseo de la Castellana, 261
Madrid, 28046
Spain
Phone: 34 676687569
Email: meme27@hotmail.es

Related Article:

Correction of: https://www.jmir.org/2020/3/e16013/

(J Med Internet Res 2020;22(7):e19961) doi:10.2196/19961

In the paper "The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology" (J Med Internet Res 2020;22(3):e16013) the authors noticed that one of the authors was not listed on the original published manuscript. The missing author was Mario Alonso Puig; they are first author on the corrected manuscript.

Mario Alonso Puig's affiliation will be listed as "Juegaterapia Foundation, Madrid, Spain" and will appear as affiliation 1 in the corrected manuscript. Author affiliations 1-4 in the original published manuscript will be renumbered to affiliations 2-5 in the corrected manuscript.

The correction will appear in the online version of the paper on the JMIR website on July 7, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 17.05.20; this is a non–peer-reviewed article;accepted 21.05.20; published 07.07.20.

Please cite as:
Alonso Puig M, Alonso-Prieto M, Miró J, Torres-Luna R, Plaza López de Sabando D, Reinoso-Barbero F
Authorship Correction: The Association Between Pain Relief Using Video Games and an Increase in Vagal Tone in Children With Cancer: Analytic Observational Study With a Quasi-Experimental Pre/Posttest Methodology
J Med Internet Res 2020;22(7):e19961
URL: https://www.jmir.org/2020/7/e19961
doi:10.2196/19961
PMID:32673254

©Mario Alonso Puig, Mercedes Alonso-Prieto, Jordi Miró, Raquel Torres-Luna, Diego Plaza López de Sabando, Francisco Reinoso-Barbero. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach

Cheng-Sheng Yu[1,2], PhD; Yu-Jiun Lin[1,2], MD; Chang-Hsien Lin[1,2], MD; Shiyng-Yu Lin[1,2], MD; Jenny L Wu[1,2], BSc; Shy-Shin Chang[1,2], MD, PhD

[1]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[2]Department of Family Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Shy-Shin Chang, MD, PhD
Department of Family Medicine
School of Medicine
College of Medicine, Taipei Medical University
250 Wuxing St
Taipei, 11031
Taiwan
Phone: 886 2 23565926
Email: sschang0529@gmail.com

Related Article:

Correction of: https://www.jmir.org/2020/6/e18585/

(J Med Internet Res 2020;22(7):e21753) doi:10.2196/21753

In "Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach" (J Med Internet Res 2020;22(6):e18585) the authors noticed an error in the Institutional Review Board (IRB) approval number.

Under the Ethics heading in the Methods section, the IRB approval number was listed as "N201906023"

(TMUH TMU-JIRB number N201906023)

The correct number is "N202003088":

(TMUH TMU-JIRB number N202003088)

The correction will appear in the online version of the paper on the JMIR Publications website on July 27, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.06.20; this is a non–peer-reviewed article;accepted 24.06.20; published 27.07.20.

Please cite as:
Yu CS, Lin YJ, Lin CH, Lin SY, Wu JL, Chang SS
Correction: Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach
J Med Internet Res 2020;22(7):e21753
URL: http://www.jmir.org/2020/7/e21753/
doi:10.2196/21753
PMID:32716902

©Cheng-Sheng Yu, Yu-Jiun Lin, Chang-Hsien Lin, Shiyng-Yu Lin, Jenny L Wu, Shy-Shin Chang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of Invited Girls: Randomized Controlled Trial

Mirjam Pot[1,2], PhD; Theo GWM Paulussen[1], PhD; Robert AC Ruiter[2], Prof Dr; Iris Eekhout[1,3], PhD; Hester E de Melker[4], PhD; Maxine EA Spoelstra[5], MSc; Hilde M van Keulen[1], PhD

[1]Netherlands Organization for Applied Scientific Research (TNO), Child Health, Leiden, Netherlands

[2]Department of Work and Social Psychology, Maastricht University, Maastricht, Netherlands

[3]VU University Medical Center, Epidemiology & Biostatistics, Amsterdam, Netherlands

[4]National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, Netherlands

[5]Department of Psychology, Leiden University, Leiden, Netherlands

Corresponding Author:
Mirjam Pot, PhD
Netherlands Organization for Applied Scientific Research (TNO)
Child Health
Schipholweg 77
Leiden, 2316 ZL
Netherlands
Phone: 31 643234293
Email: mirjampot90@gmail.com

Related Article:

Correction of: https://www.jmir.org/2017/9/e312/

(J Med Internet Res 2020;22(7):e22565) doi:10.2196/22565

In "Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of Invited Girls: Randomized Controlled Trial" J Med Internet Res 2017;19(9):e312 the authors noted several errors.

In the Methods section, under the subheading "Statistical Analyses, the phrase:

> *Effect sizes of the linear regressions were calculated in R (R Development Core Team) [45] using Cohen]f^2 (R^2 including the outcome at baseline and condition/1− R^2 only including the outcome at baseline).*

Has been revised to:

> *Effect sizes for the linear regressions were calculated in R (R Development Core Team) [45] using Cohen f^2 statistic, ($R^2_{AB} – R^2_A$)/(1 – R^2_{AB}), in which B is the variable of interest (ie, condition), A is the set of all other variables (ie, the outcome at baseline), R^2_{AB} is the proportion of variance accounting for A and B together, and R^2_A is the proportion of variance accounted for by A.*

This change has not affected the results of the paper.

Additionally, the following degrees have been revised: Mirjam Pot's degree has been changed from MSc to PhD, Maxine ME Spoelstra's degree has been changed from BSc to MSc, and Robert AC Ruiter's degree has been changed from PhD to Prof Dr.

Finally the Corresponding Author section has been revised. Mirjam Pot's email address has been updated from mirjam.pot@tno.nl to mirjampot90@gmail.com. The phone number has been updated from "31 888662213" to "31 643234293", and the fax number (31 888660613) has been removed.

The correction will appear in the online version of the paper on the JMIR Publications website on July 28, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 16.07.20; this is a non–peer-reviewed article;accepted 17.07.20; published 28.07.20.

Please cite as:
Pot M, Paulussen TGWM, Ruiter RAC, Eekhout I, de Melker HE, Spoelstra MEA, van Keulen HM
Correction: Effectiveness of a Web-Based Tailored Intervention With Virtual Assistants Promoting the Acceptability of HPV Vaccination Among Mothers of Invited Girls: Randomized Controlled Trial
J Med Internet Res 2020;22(7):e22565
URL: http://www.jmir.org/2020/7/e22565/
doi:10.2196/22565
PMID:32721923

©Mirjam Pot, Theo GWM Paulussen, Robert AC Ruiter, Iris Eekhout, Hester E de Melker, Maxine EA Spoelstra, Hilde M van Keulen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mapping of Health Literacy and Social Panic Via Web Search Data During the COVID-19 Public Health Emergency: Infodemiological Study

Chenjie Xu[1*], MSc; Xinyu Zhang[1*], PhD; Yaogang Wang[1*], PhD

School of Public Health, Tianjin Medical University, Tianjin, China

*all authors contributed equally

Corresponding Author:
Yaogang Wang, PhD
School of Public Health
Tianjin Medical University
No 22, Qixiangtai Road
Heping District
Tianjin, 300070
China
Phone: 86 13820046130
Email: wangyg@tmu.edu.cn

Abstract

Background: Coronavirus disease (COVID-19) is a type of pneumonia caused by a novel coronavirus that was discovered in 2019. As of May 6, 2020, 84,407 cases and 4643 deaths have been confirmed in China. The Chinese population has expressed great concern since the COVID-19 outbreak. Meanwhile, an average of 1 billion people per day are using the Baidu search engine to find COVID-19–related health information.

Objective: The aim of this paper is to analyze web search data volumes related to COVID-19 in China.

Methods: We conducted an infodemiological study to analyze web search data volumes related to COVID-19. Using Baidu Index data, we assessed the search frequencies of specific search terms in Baidu to describe the impact of COVID-19 on public health, psychology, behaviors, lifestyles, and social policies (from February 11, 2020, to March 17, 2020).

Results: The search frequency related to COVID-19 has increased significantly since February 11th. Our heat maps demonstrate that citizens in Wuhan, Hubei Province, express more concern about COVID-19 than citizens from other cities since the outbreak first occurred in Wuhan. Wuhan citizens frequently searched for content related to “medical help,” “protective materials,” and “pandemic progress.” Web searches for “return to work” and “go back to school” have increased eight-fold compared to the previous month. Searches for content related to “closed community and remote office” have continued to rise, and searches for “remote office demand” have risen by 663% from the previous quarter. Employees who have returned to work have mainly engaged in the following web searches: “return to work and prevention measures,” “return to work guarantee policy,” and “time to return to work.” Provinces with large, educated populations (eg, Henan, Hebei, and Shandong) have been focusing on “online education” whereas medium-sized cities have been paying more attention to “online medical care.”

Conclusions: Our findings suggest that web search data may reflect changes in health literacy, social panic, and prevention and control policies in response to COVID-19.

(J Med Internet Res 2020;22(7):e18831) doi:10.2196/18831

KEYWORDS

COVID-19; China; Baidu; infodemiology; web search; internet; public health; emergency; outbreak; infectious disease; pandemic; health literacy

Introduction

In December 2019, a severe public health emergency was induced by the outbreak of a novel coronavirus, which has since been named coronavirus disease (COVID-19) by the World Health Organization (WHO) [1]. Since the first-level response by government officials to COVID-19 across China’s provinces and cities, governmentally imposed social isolation has provided

XSL•FO
RenderX

the Chinese population with ample time to search online for the latest COVID-19–related news [2]. Baidu, as the most widely used Chinese search engine, accounts for two-thirds of China's search engine market share [3]. At the time of the COVID-19 outbreak, we found that the number of searches for COVID-19 had increased exponentially, despite the fact that its incidence and mortality rates were much lower than those of some noncommunicable diseases, such as cancer. Hence, this phenomenon is worthy of further attention and discussion.

Methods

We obtained web search data from the Baidu Index [4]. As of May 2020, Baidu accounts for 71.23% of the search engine market share in China [5] and is the most widely used search engine in the country. It is a well-known and extensive platform for information/resource sharing that Chinese internet users rely on.

Search data, dating to as early as 2004, were derived from search frequencies on Baidu. Frequencies were calculated based on the search volumes of specific search terms entered by internet users, and data on a daily, monthly, and yearly basis were obtained for the search terms we chose from the Baidu Index.

We entered the formal Chinese names for "lung cancer," "liver cancer," "esophageal cancer," "colon and rectum cancer," and "breast cancer," respectively, and acquired their search data history from January 1, 2020, to March 17, 2020. The Baidu Index provides search term analysis; it also uses a process to scientifically determine related search terms based on the mode through which the searchers initiate a search request. This means that other relevant search terms will be provided automatically once the internet users enter a search term. Thus, we entered "COVID-19" into the Baidu Index and obtained its related search terms from February 11, 2020, to March 17, 2020 (search volumes for "COVID-19" can only be obtained from February 11, 2020 onward). The search terms mainly included health literacy, social panic, and prevention and control measures relating to COVID-19. All search data were downloaded free of charge on March 17, 2020.

We performed descriptive analyses to describe and compare the overall search situation under the context of the pandemic from different aspects. Statistical analysis was conducted using Excel 2016 (Microsoft Corporation) and IBM SPSS (version 22.0, IBM Corporation). We used Tableau (version 2018.3, Tableau Software) and Excel 2016 (Microsoft Corporation) to create heat maps.

Results

Mapping of Health Literacy and Social Panic Via Web Search Data During a Public Health Emergency

The COVID-19 pandemic necessitates requirements for public health literacy. Despite China's active response in combating this outbreak, a proportion of the Chinese population has remained afraid, which may seriously affect progress in controlling and preventing future outbreaks. This phenomenon is particularly striking on the internet and within social media, and investigating COVID-19–related health literacy and social panic via these platforms may yield important insights.

Taking the COVID-19 outbreak as an example, spurious information disseminated via social media by nonexpert individuals has included false claims on how to kill the COVID-19 virus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]), for example, by drinking alcohol, smoking, aromatherapy, essential balms, use of a hair dryer, or taking a hot bath [2]. The internet has figuratively become a double-edged sword in the context of the COVID-19 outbreak.

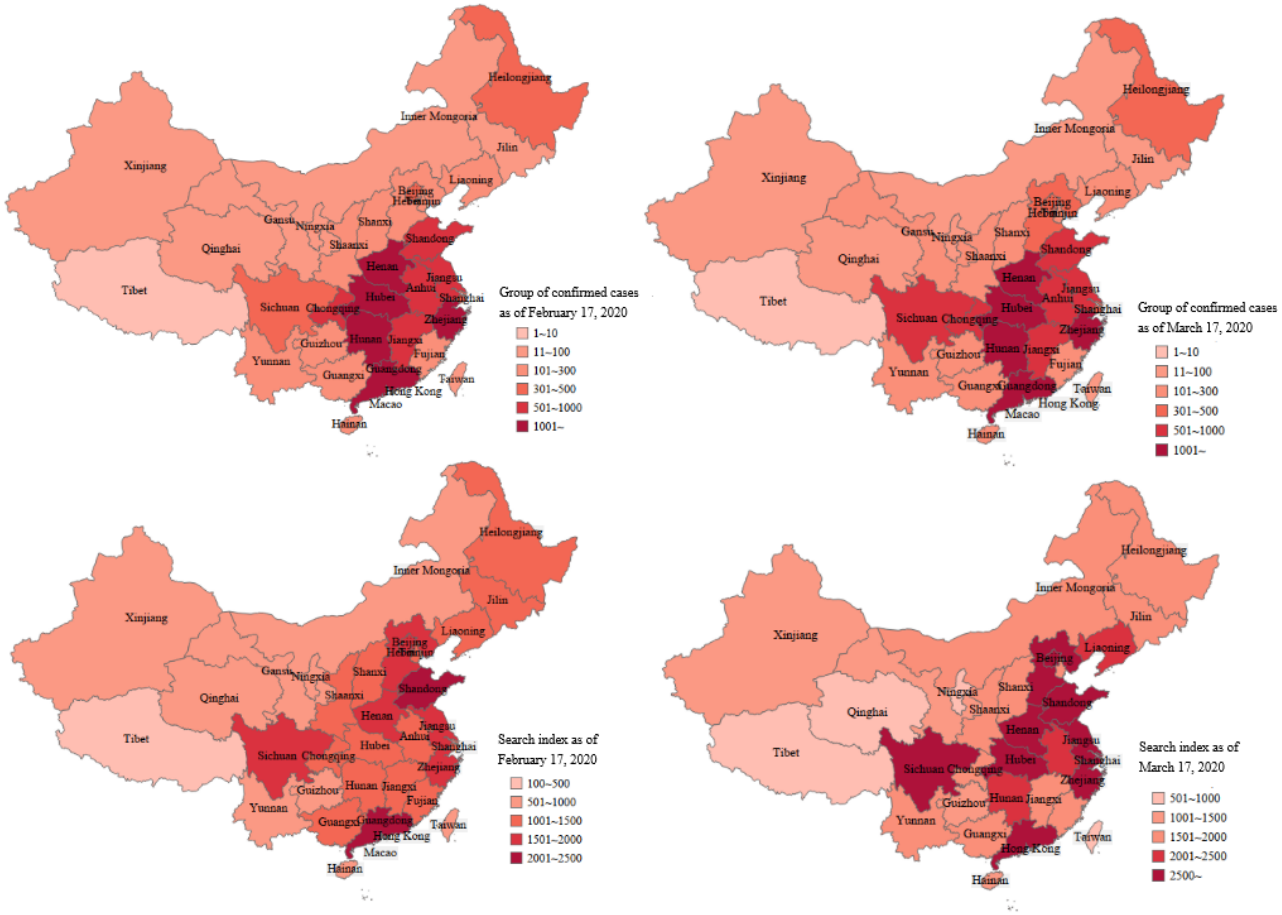
Concerns of Internet Users During the COVID-19 Outbreak

By summarizing web searches for COVID-19, the most searched content has focused on the progression of the pandemic and tips on how to protect oneself from infection [6]. Web searches on COVID-19 have mainly included the following: the latest news on the pandemic; relevant knowledge about COVID-19 infection prevention and control; basic, as well as more comprehensive, information describing SARS-CoV-2; and rumors and suggested policies from unqualified sources [6].

Citizens in Wuhan, Hubei Province, have been more concerned about COVID-19 compared to citizens from other cities during the same period since the outbreak first occurred in Wuhan (Figure 1). Wuhan citizens searched for content related to "medical help" 22% of the time, followed by searches for "protective materials" and "pandemic progress." Medical help refers to special assistance and support provided to the citizens of Wuhan during the pandemic, which included the deployment of 42,000 trained medical workers from across the country to Wuhan and other cities in Hubei Province, and donations of medical supplies (medical masks, medical protective clothing, ventilators, etc).

XSL•FO
RenderX

Figure 1. The cumulative number of confirmed cases of coronavirus disease (COVID-19) (top row) and the change in COVID-19's search index (bottom row) in China.



Since the COVID-19 outbreak, web searches for wild animals have reached a historical peak—the Baidu topic “refuse to eat wild animals” has reached nearly 100 million views [7]. Content related to “harm to wild animals” has received considerable attention. At the early stage of the pandemic, there were some speculation that wild animals may have spread the coronavirus to people in Wuhan, although there is no scientific conclusion regarding the origin of COVID-19 [8-10]. In this context, on February 24, 2020, on the basis of the *Law of the People's Republic of China on the Protection of Wildlife*, China established a comprehensive system to prohibit the consumption of wild animals [11].

With the pandemic now gradually under control in China, public concerns have changed accordingly [12]. Web searches for “when to return to work and start school” have increased eight-fold from those during the previous month based on the web search data in the Baidu Index. Searches for content related to “closed community and remote office” have continued to rise, and searches for “remote office demand” have risen by 663% from the previous quarter. Individuals from first-tier cities such as Beijing, Shanghai, Shenzhen, Guangzhou, and Chengdu have paid more attention to telecommuting. Employees who have returned to work have mainly engaged in web searches on “return to work and prevention measures,” “return to work guarantee policy,” and “time to return to work.” Men have been half as attentive as women regarding protective measures related to returning to work.

Previously, the four major industries most discussed by Chinese netizens have been online education, online medical care, online entertainment, and fresh electronic commerce (e-commerce) (Table 1) [6]. During the pandemic, provinces with large populations of educated individuals (eg, Henan, Hebei, and Shandong) have been focusing on “online education.” In contrast, fourth-tier cities have been paying more attention to “online medical care” (in China, cities are usually graded according to various levels such as city development, comprehensive economic strength, talent attractiveness, information exchange capability, international competitiveness, technological innovation capability, and transportation accessibility. First-tier cities refer to metropolises that are important for social, political, and economic activities [eg, Shanghai, Beijing]. Most fourth-tier cities are medium-sized cities.). Additionally, young and middle-aged groups exhibited greater search volumes for “online medical care” [10].

XSL•FO
RenderX

Table 1. Growth rate of search volumes of the four major industries discussed by netizens [6].

Industry	Search growth rate (%)
Online education	248
Online medical care	200
Online entertainment	170
Fresh electronic commerce	120

Impact of a Lack of Health Literacy on the COVID-19 Outbreak

COVID-19 is a novel and highly infectious disease [13,14], but in the early days of the outbreak, people had limited knowledge of it and did not know how to prevent it, which concomitantly caused panic. Despite the COVID-19 pandemic becoming gradually controlled in China, most people still remain under high alert. Some misinterpretations of discussions on the resumption of work on the internet have caused another wave of panic among some groups and cities since returning to work. This phenomenon can be observed on Weibo, a social media platform based on information sharing, dissemination, and acquisition in real time (similar to Facebook). Many people are worried about whether a large number of people returning to work will reinitiate the spread of COVID-19. However, individuals in areas with no outbreaks of the disease or low incidence of infections have also expressed this concern. However, if cities continue to stagnate, such stagnation may lead to a higher mortality rate than that attributed to COVID-19. Zhang Wenhong, head of a Shanghai medical treatment expert group, stated the following in an interview [15]:

> *If the hospital does not return to work, cancer patients cannot receive chemotherapy and surgery, and other infected people cannot be treated. Patients with trauma cannot get a good treatment. Under such circumstances, the number of patients dying from other diseases will far exceed the number of people dying from the new coronavirus.*

Why Are Public Health Emergencies of Greater Concern Than Highly Fatal Chronic Diseases Like Cancer?

Cancer-related deaths in China account for approximately 27% of all global cancer-related deaths [16,17]. As presented in Table 2, except for esophageal cancer, the search frequencies of all other cancers have decreased since the COVID-19 outbreak. Meanwhile, the search frequency of COVID-19 has increased significantly [4]. Dr Tedros Adhanom Ghebreyesus, Director-General of the WHO, reported that the global mortality rate of COVID-19 is approximately about 3.4% [18]. According to the latest data from the National Health Commission of the People's Republic of China on February 3, 2020, the mortality rate of COVID-19 in Hubei Province is 3.1%, while the national COVID-19 mortality rate is even lower at 0.2% [19].

Table 2. Changes in the search index of the top five cancers and coronavirus disease (COVID-19) in China from January 27, 2020, to March 17, 2020.

Disease	Incidence rate (per 100,000 persons)[a]	Mortality rate (per 100,000 persons)[a]	Daily mean value		Overall search index[b]		Mobile search index[b]	
			Overall search index	Mobile search index	Year-on-year change (%)	Month-on-month change (%)	Year-on-year change (%)	Month-on-month change (%)
Lung cancer	58	49	3849	3538	–23	–1	–19	__[c]
Liver cancer	37	30	1301	1128	–63	–22	–65	–18
Esophageal cancer	17	15	1108	957	11	3	14	3
Colon and rectum cancer	31	13	215	120	–12	–9	–21	–16
Breast cancer	26	6	2235	2034	–50	–15	–50	–13
COVID-19	—	—	25,256	19,614	—	4263	—	3784

[a]Incidence and mortality rates for the five cancers were obtained from the Global Burden of Diseases Database [20].

[b]Negative values represent decline.

[c]Not available.

The death rate associated with cancer is much higher than that of COVID-19. If we use the national data (excluding Hubei Province), the mortality rate of COVID-19 is comparable to that of the general influenza, and the total incidence rate is far lower than that of influenza [21,22]. There are several reasons that may explain the relatively high mortality rates of COVID-19 in Wuhan and in Hubei Province in general, including the stronger virulence of SARS-CoV-2 in Wuhan, more cross-infection, and the prevalence of patients with mild symptoms who did not see a doctor [23].

XSL•FO
RenderX

Discussion

Principal Findings

Health literacy includes two aspects: (1) knowledge, which comprises basic health knowledge and skills; and (2) ability, which refers to one's ability to acquire, understand, screen, and apply health information [24,25]. Health information literacy represents the core of health literacy—it can greatly improve the public's capacity for self-protection in order to improve the overall response to public health emergencies.

The health literacy rate of Chinese residents in 2018 was 17.06% [26]. It mainly covers basic health knowledge and concept literacy, healthy lifestyle and behavior literacy, and basic health skills literacy. Although the health literacy rate of residents has improved, the uneven distribution of health literacy levels between urban and rural areas, and across regions and populations, still exists. The health literacy levels of rural residents, residents in the Midwest, and the elderly are relatively low. As mentioned above, with the rapid development of internet technology, people can easily use the internet to search for health information. However, the new coronavirus that caused the recent pandemic was previously unknown. Since the COVID-19 outbreak, the related transmission characteristics, symptoms, transmission channels, and methods for protection have been gradually communicated to the public via recent publications on COVID-19–related research. Public health emergencies have the characteristics of urgency and paroxysm since they require the public to respond quickly. At such times, the ability to acquire, understand, and use health data will enable individuals to more quickly facilitate disease control and prevention in the face of a public health emergency.

People use the internet for almost everything they do nowadays. By uploading and downloading information, everyone can be a publisher and conveyor of the immense quantity of information available on the internet. Search engines facilitate the acquisition and learning of health information from a variety of sources, which provides the public with more diversified content, autonomy, and greater control over choices. However, search engines also lead to many problems. For example, they may disseminate data from unreliable sources, making it difficult for the public to distinguish between high-quality and low-quality health information [27-29]. In these unprecented times, people are more vulnerable and credulous to the impact of new information related to COVID-19. Hence, the vast amount of information available on the internet undoubtedly has a considerable impact on public health literacy and may influence the control and prevention of further outbreaks.

Improving health information literacy is the primary component of improving overall health literacy. First, it is necessary to raise awareness of the important role of health information literacy, and to realize how it can improve public health and promote the development of health services. Second, the state should take the lead in setting up a network containing national health information and support services from multiple sources (eg, the public, medical organizations, governmental sectors) to ensure that the public can receive assistance in obtaining health information and related skills. Because health information involves professional knowledge and also concerns the health of every citizen, the release of authoritative information from professional agencies during the pandemic is particularly important. During the COVID-19 pandemic, many users of social media often cited the words of academician Zhong Nanshan, a trusted public academic, to spread information from a credible source. This pattern demonstrates that individuals need authoritative health information. Timely and authoritative information may be one of the most effective ways to eliminate doubt and reduce panic, especially in the face of public health emergencies,. Third, education should be strengthened to improve the level of public health literacy throughout schools, communities, and villages. Individual education levels and education systems are important factors affecting public health information literacy, and health information literacy education through comprehensive linkage represents the most inclusive and cost-effective precautionary measure [30,31]. As recommended by the WHO, the following response strategies are required: rapidly establishing international coordination and operational support; scaling up of national readiness and responses to operations; and accelerating priority research and innovation [32].

In addition to face-to-face communication, the online role of health care providers in public health communications is also important for mitigating medical misinformation. Examples of such online public health communications are WebMD in the United States, AskDr in Singapore, HaoDF in China, etc. Through these outlets, health care authorities can communicate with the public via the internet and provide professional and reliable health information.

While classic public health measures still play very important roles in tackling the current COVID-19 pandemic, there are also many new potentially enabling technological domains that can be applied to help monitor, survey, detect, and prevent such pandemics, including the Internet of Things, big data analytics, artificial intelligence, and blockchain technology. Therefore, we should make full use of these emerging technologies to contain the current COVID-19 outbreak, as well as future pandemics in a timely and effective manner [33-39].

Conclusions

Since the Chinese New Year, the COVID-19 outbreak has become the most important issue in China. Fear of an unknown virus represents the beginnings of panic, and the internet and social media networks have since become incubators and catalysts of panic. Individual cases can be shared by tens of millions of people in a single day through social media dissemination. Hundreds of millions of people are eagerly absorbing information about the novel coronavirus, but reading about new cases is undoubtedly causing concerns among citizens about their own health status.

The pandemic has diverted attention from other health issues. China accounts for 19% of the global population, and its incidence of cancer accounts for 22% of the total global prevalence of cancer. Nevertheless, these diseases have never caused panic at the level of COVID-19. The mortality rate of COVID-19 cannot be compared with that of any other major noncommunicable disease. Despite this discrepancy, the Chinese

XSL•FO
RenderX

population continues to be horrified by COVID-19 but at ease with other known diseases. We need to collectively alter our minds about COVID-19 in order to manage such public health emergencies more rationally. This change requires a more logical use of information obtained from social media and the internet during public health emergencies, as well as improvements in health literacy and the ability to cope with social panic.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (#71910107004). We also want to acknowledge Baidu Corporation for its open-data support.

Authors' Contributions

All authors critically revised the manuscript, reviewed, and contributed to the final version and approved it. YW is responsible for the study design.

Conflicts of Interest

None declared.

References

1. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. World Health Organization. 2020 Feb 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020 [accessed 2020-03-01]
2. Mowbray H. Letter from China: covid-19 on the grapevine, on the internet, and in commerce. BMJ 2020 Feb 19;368:m643. [doi: 10.1136/bmj.m643] [Medline: 32075783]
3. Chinese Internet users search behavior study. China Internet Network Information Center. 2014. URL: http://www.cnnic.cn/hlwfzyj/hlwmrtj/201410/t20141017_49359.htm [accessed 2020-03-01]
4. Baidu Index. Baidu. URL: http://index.baidu.com [accessed 2020-06-06]
5. Search Engine Market Share China. Statcounter. URL: https://gs.statcounter.com/search-engine-market-share/all/china/#monthly-201905-202005 [accessed 2020-06-18]
6. The epidemic search big data is coming. Baidu. URL: https://baike.baidu.com/vbaike/%E7%96%AB%E6%83%85%E6%90%9C%E7%B4%A2%E5%A4%A7%E6%95%B0%E6%8D%AE%E6%9D%A5%E4%BA%86/56236 [accessed 2020-03-10]
7. Refuse to eat wild animals. Baidu. URL: https://baike.baidu.com/vbaike/%E7%96%AB%E6%83%85%E6%8A%A5%E5%91%8A%E6%8B%92%E7%BB%9D%E9%87%8E%E5%91%B3%E7%AF%87/56282 [accessed 2020-03-10]
8. Cyranoski D. Did pangolins spread the China coronavirus to people? Nature 2020 Feb 7;113(3):554-559 [FREE Full text] [doi: 10.1038/d41586-020-00364-2] [Medline: 26729863]
9. Mallapaty S. Animal source of the coronavirus continues to elude scientists. Nature 2020 May 18. [doi: 10.1038/d41586-020-01449-8] [Medline: 32427902]
10. Cyranoski D. Mystery deepens over animal source of coronavirus. Nature 2020 Mar;579(7797):18-19. [doi: 10.1038/d41586-020-00548-w] [Medline: 32127703]
11. The National People’s Congress of the People’s Republic of China. Remarks on the "Decision of the Standing Committee of the National People's Congress Regarding the Prohibition of Illegal Wildlife Trade, the Elimination of the Abuse of Wild Animal Eating, and the Effective Protection of People's Health and Safety (Draft)". URL: http://www.npc.gov.cn/npc/c30834/202003/40803ceecf044b619942133f66cfdc44.shtml [accessed 2020-06-04]
12. How to prevent and control resumption of production. Baidu. URL: https://baike.baidu.com/vbaike/%E5%A4%8D%E5%B7%A5%E5%A4%8D%E4%BA%A7%E6%80%8E%E4%B9%88%E5%81%9A%E9%98%B2%E6%8E%A7/56763 [accessed 2020-03-10]
13. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
14. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
15. If people do not return to work, the death rate will be much higher than COVID-19. Tencent News. URL: https://view.inews.qq.com/a/20200303V05O0Q00 [accessed 2020-06-18]
16. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: priorities for prevention. Carcinogenesis 2010 Jan;31(1):100-110 [FREE Full text] [doi: 10.1093/carcin/bgp263] [Medline: 19934210]

XSL•FO
RenderX

17. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015 Mar;65(2):87-108 [FREE Full text] [doi: 10.3322/caac.21262] [Medline: 25651787]
18. Global mortality rate of new coronary pneumonia is 3.4%. Changchun Daily. URL: http://epaper.1news.cc/ccrb/pc/paper/c/202003/06/content_1791833.html [accessed 2020-03-01]
19. National mortality rate of new coronary pneumonia was 2.1%, the highest in Wuhan, Hubei, and the National Health and Health Commission responded. CN Healthcare. URL: https://www.cn-healthcare.com/article/20200204/wap-content-529896.html [accessed 2020-05-05]
20. Global Burden of Diseases. GBD Results Tool. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2020-06-18]
21. Begley S. Lower death rate estimates for coronavirus, especially for non-elderly, provide glimmer of hope. STAT. URL: https://www.statnews.com/2020/03/16/lower-coronavirus-death-rate-estimates/ [accessed 2020-05-05]
22. Rettner R. How does the new coronavirus compare with the flu? Live Science. URL: https://www.livescience.com/new-coronavirus-compare-with-flu.html [accessed 2020-05-05]
23. Coronavirus disease 2019 (COVID-19) Situation Report 32. World Health Organization. 2020 Feb. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200221-sitrep-32-covid-19.pdf?sfvrsn=4802d089_2 [accessed 2020-05-05]
24. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
25. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
26. Regular press release on August 27, 2019. National Health Commission of the People's Republic of China. 2019. URL: http://www.nhc.gov.cn/xcs/s7847/201908/2a77e09aed9d427ab6d4e472095ba020.shtml [accessed 2020-06-04]
27. Choi D, Chun S, Oh H, Han J, Kwon TT. Rumor Propagation is Amplified by Echo Chambers in Social Media. Sci Rep 2020 Jan 15;10(1):310 [FREE Full text] [doi: 10.1038/s41598-019-57272-3] [Medline: 31941980]
28. Jones NM, Thompson RR, Dunkel Schetter C, Silver RC. Distress and rumor exposure on social media during a campus lockdown. Proc Natl Acad Sci U S A 2017 Oct 31;114(44):11663-11668 [FREE Full text] [doi: 10.1073/pnas.1708518114] [Medline: 29042513]
29. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]
30. Meng L, Lyu Y, Cao Y, Tu W, Hong Z, Li L, et al. Information obtained through Internet-based media surveillance regarding domestic public health emergencies in 2013. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2015 Jun;36(6):607-611. [Medline: 26564634]
31. Wang X, Zhang X, He J. Challenges to the system of reserve medical supplies for public health emergencies: reflections on the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic in China. Biosci Trends 2020 Mar 16;14(1):3-8 [FREE Full text] [doi: 10.5582/bst.2020.01043] [Medline: 32062645]
32. 2019 Novel Coronavirus (2019-nCoV): Strategic preparedness and response plan. World Health Organization. URL: https://www.who.int/publications/i/item/strategic-preparedness-and-response-plan-for-the-new-coronavirus [accessed 2020-05-05]
33. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
34. Perkel JM. The Internet of Things comes to the lab. Nature 2017 Jan 30;542(7639):125-126. [doi: 10.1038/542125a] [Medline: 28150787]
35. Ting D, Lin H, Ruamviboonsuk P, Wong T, Sim D. Artificial intelligence, the internet of things, and virtual clinics: ophthalmology at the digital translation forefront. The Lancet Digital Health 2020 Jan;2(1):e8-e9. [doi: 10.1016/S2589-7500(19)30217-1] [Medline: 30718888]
36. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 2020 Jan;26(1):29-38. [doi: 10.1038/s41591-019-0727-5] [Medline: 31932803]
37. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
38. Ting DSW, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. Nat Med 2018 May;24(5):539-540. [doi: 10.1038/s41591-018-0029-3] [Medline: 29736024]
39. Heaven D. Bitcoin for the biological literature. Nature 2019 Feb;566(7742):141-142. [doi: 10.1038/d41586-019-00447-9] [Medline: 30718888]

Abbreviations

COVID-19: coronavirus disease
e-commerce: electronic commerce
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Eysenbach; submitted 22.03.20; peer-reviewed by M Manning Hutson, D Gunasekeran, O Keiser; comments to author 29.04.20; revised version received 06.05.20; accepted 13.06.20; published 02.07.20.

Please cite as:
Xu C, Zhang X, Wang Y
Mapping of Health Literacy and Social Panic Via Web Search Data During the COVID-19 Public Health Emergency: Infodemiological Study
J Med Internet Res 2020;22(7):e18831
URL: https://www.jmir.org/2020/7/e18831
doi:10.2196/18831
PMID:32540844

©Chenjie Xu, Xinyu Zhang, Yaogang Wang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study

Hui Xu[1*], MD; Sufang Huang[2*], MSc; Chun Qiu[3], PhD; Shangkun Liu[1], MSc; Juan Deng[2], MSc; Bo Jiao[1], MSc; Xi Tan[1], MSc; Ling Ai[1], MD; Yaru Xiao[2], MSc; Mirko Belliato[4*], MD; Li Yan[2*], MD

[1]Department of Anesthesiology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Emergency Medicine, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, ON, Canada

[4]Anestesia e Rianimazione 1, Foundation IRCCS Policlinico San Matteo, Pavia, Italy

[*]these authors contributed equally

Corresponding Author:
Li Yan, MD
Department of Emergency Medicine
Tongji Hospital of Tongji Medical College
Huazhong University of Science and Technology
1095 Jie Fang Avenue
Qiaokou District
Wuhan, 430030
China
Phone: 86 13607124420
Email: yanli008@163.com

Abstract

Background: Most patients with coronavirus disease (COVID-19) who show mild symptoms are sent home by physicians to recover. However, the condition of some of these patients becomes severe or critical as the disease progresses.

Objective: The aim of this study was to evaluate a telemedicine model that was developed to address the challenges of treating patients with progressive COVID-19 who are home-quarantined and shortages in the medical workforce.

Methods: A telemedicine system was developed to continuously monitor the progression of home-quarantined patients with COVID-19. The system was built based on a popular social media smartphone app called WeChat; the app was used to establish two-way communication between a multidisciplinary team consisting of 7 medical workers and 188 home-quarantined individuals (including 74 confirmed patients with COVID-19). The system helped patients self-assess their conditions and update the multidisciplinary team through a telemedicine form stored on a cloud service, based on which the multidisciplinary team made treatment decisions. We evaluated this telemedicine system via a single-center retrospective study conducted at Tongji Hospital in Wuhan, China, in January 2020.

Results: Among 188 individuals using the telemedicine system, 114 (60.6%) were not infected with COVID-19 and were dismissed. Of the 74 confirmed patients with COVID-19, 26 (35%) recovered during the study period and voluntarily stopped using the system. The remaining 48/76 confirmed patients with COVID-19 (63%) used the system until the end of the study, including 6 patients whose conditions progressed to severe or critical. These 6 patients were admitted to hospital and were stabilized (one received extracorporeal membrane oxygenation support for 17 days). All 74 patients with COVID-19 eventually recovered. Through a comparison of the monitored symptoms between hospitalized and nonhospitalized patients, we found prolonged persistence and deterioration of fever, dyspnea, lack of strength, and muscle soreness to be diagnostic of need for hospitalization.

Conclusions: By continuously monitoring the changes in several key symptoms, the telemedicine system reduces the risks of delayed hospitalization due to disease progression for patients with COVID-19 quarantined at home. The system uses a set of scales for quarantine management assessment that enables patients to self-assess their conditions. The results are useful for medical staff to identify disease progression and, hence, make appropriate and timely treatment decisions. The system requires few staff

XSL•FO
RenderX

to manage a large cohort of patients. In addition, the system can solicit help from recovered but self-quarantined medical workers to alleviate shortages in the medical workforce and free healthy medical workers to fight COVID-19 on the front line. Thus, it optimizes the usage of local medical resources and prevents cross-infections among medical workers and patients.

(J Med Internet Res 2020;22(7):e19514) doi:10.2196/19514

KEYWORDS

telemedicine system; home quarantine; quarantine management assessment; progressive COVID-19 patients; COVID-19

Introduction

Since its outbreak in December of 2019, coronavirus disease (COVID-19) has spread worldwide, causing more than 5 million infections and tens of thousands of deaths in the course of three months [1]. In cities that were severely affected by COVID-19, such as Wuhan (China), Lombardy (Italy), and New York City (United States) [1], local medical capacities were quickly depleted by large numbers of patients who hurried to hospitals for treatment. Many medical workers were infected, and medical care supplies were further exhausted [2]. To prevent collapse of the global health care system, many countries have advocated for infected patients with mild symptoms to stay home and self-quarantine [3]. However, it has been observed that the condition of some home-quarantined patients becomes severe or critical as the disease progresses. Home quarantine can delay timely treatment and hospitalization of these patients, which may lead to their death.

In this paper, we report a telemedicine model that we developed to address the challenges outlined above. This telemedicine system enabled close monitoring of 74 home-quarantined COVID-19 patients from January 6 to 31, 2020. Of the 74 patients, 6 (8%) were admitted to hospital when signs of deterioration were detected by the system. One patient received extracorporeal membrane oxygenation (ECMO) treatment for 17 days. All 74 patients recovered.

The telemedicine model was built based on WeChat, a popular smartphone app for instant messaging and social media. The WeChat app established two-way communication between the home-quarantined patients and a multidisciplinary team. The multidisciplinary team contacted the patients regularly to receive information updates. The information was subsequently analyzed to determine the latest status of the patients. Home-quarantined patients could also initiate communication to report any abnormalities in addition to receiving feedback about their medical conditions. The system we developed is the first to quantify subjective symptom descriptions with objective numerical scales. In addition, this telemedicine system helps alleviate the workload of overwhelmed medical staff [4], as we found that excessive laboratory data and physical examinations were not strictly necessary to identify the prognoses of patients with mild symptoms. Furthermore, the telemedicine model minimizes the risk of infection among caregiving staff by reducing their direct physical contact with patients. In this paper, we summarize the procedures of the telemedicine model and present clinical evidence of its success.

Methods

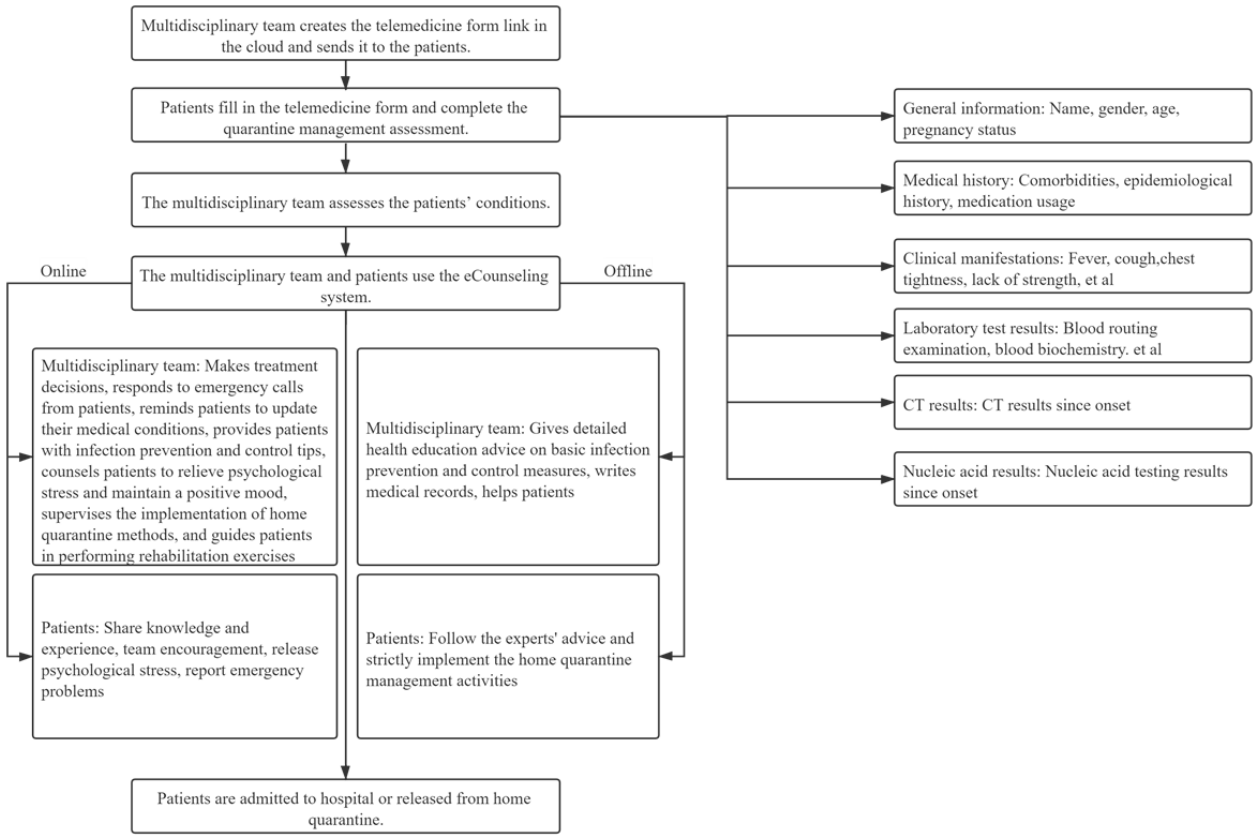
We evaluated the telemedicine system via a single-center retrospective study conducted at Tongji Hospital in Wuhan, China, between January 6 and January 31, 2020. The study was approved by the Tongji Hospital Ethics Committee before data were collected retrospectively.

Procedure and Information Flow of the Telemedicine System

Below, we outline the procedure of the telemedicine system. When a patient signed up for the telemedicine system, the multidisciplinary team created a patient-specific telemedicine form in the cloud. The telemedicine form was then sent to the patient through a link via the WeChat app to the patient's mobile phone or by email. The patient began the process by filling in the telemedicine form (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of the telemedicine system, including the steps to enroll a patient, conduct self-assessment via the quarantine management assessment, and update conditions using the eCounseling system. CRP: C-reactive protein; CT: computed tomography; eCounseling: electronic counseling; PCT: procalcitonin.



The patient first answered a set of standardized questions, which included general demographic information (name, gender, age, occupation, and pregnancy status), medical history (comorbidities, epidemiological history within the past 14 days, and medication usage), clinical manifestations (eg, fever, cough, chest tightness, lack of strength), various laboratory tests, computed tomography (CT) images, and nucleic acid test results for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) detection (Figure 1).

Next, the patient underwent quarantine management assessment. This assessment was based on a set of medical observation scales developed by the multidisciplinary team based on both a literature review [5-9] and input from 34 medical experts from our medical school or its affiliated hospitals (Multimedia Appendix 1). The quarantine management assessment helped the multidisciplinary team assess the patients' conditions.

Based on the assessment, the multidisciplinary team determined whether the patient should be hospitalized immediately or placed under medical observation at home. If the patient was placed under medical observation, the patient would then begin using the electronic counseling (eCounseling) system, which facilitated close observation by and efficient communication with the multidisciplinary team. Specifically, the patient was required to update their conditions on a daily basis using the telemedicine form. Due to the convenience provided by the cloud service, the information could be instantly accessed by pertinent multidisciplinary team members. They subsequently provided feedback and guidance on the telemedicine form, which then could also be accessed by the patient immediately.

In addition, the patient was invited to join a WeChat group consisting of patients participating in the telemedicine system and multidisciplinary team members. In this way, the patient could receive health tips from the multidisciplinary team and conduct a group chat with other patients and multidisciplinary team members. The patient could also initiate a one-to-one chat or telephone call with any multidisciplinary team member using built-in functions in the WeChat app.

The multidisciplinary team adjusted their observational attention based on the progression of each patient's condition. If a patient's condition continued to worsen (eg, repeated and persistent fever >38.5 degrees Celsius, cyanosis or CT image deterioration >50% within 48 hours) or their mental state scores continued to decline, the patient was flagged as "red." The multidisciplinary team then determined that the patient should be admitted to hospital to be treated. The detailed decision-making process regarding hospitalization is presented in Figure 2. Figures 3 and 4 provide screenshots of the interfaces in which the telemedicine form is created and used, respectively.

XSL•FO
RenderX

Figure 2. Decision tree of patient treatment for MDT based on patients' condition updates via the eCounseling system. ºC: degrees Celsius; BP: blood pressure; bpm: beats per minute; CT: computed tomography; mmHg: millimeters of mercury; SpO2: oxygen saturation.

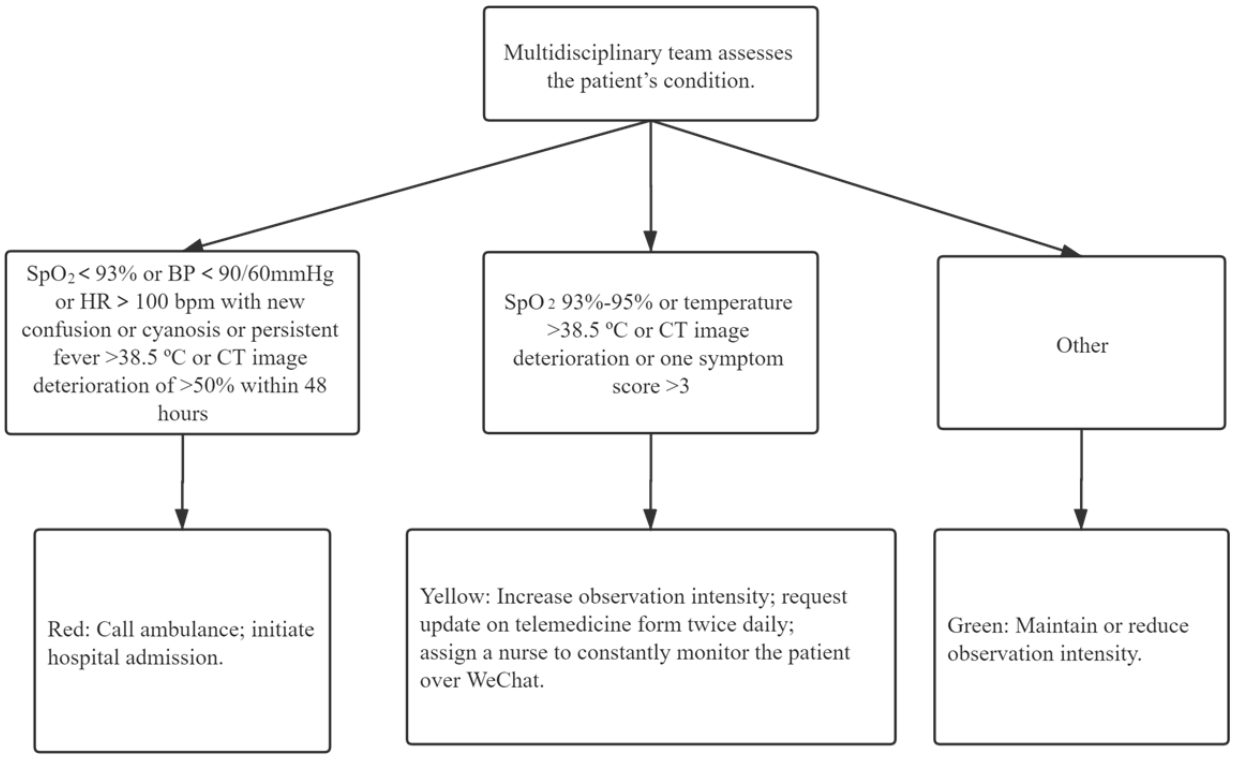
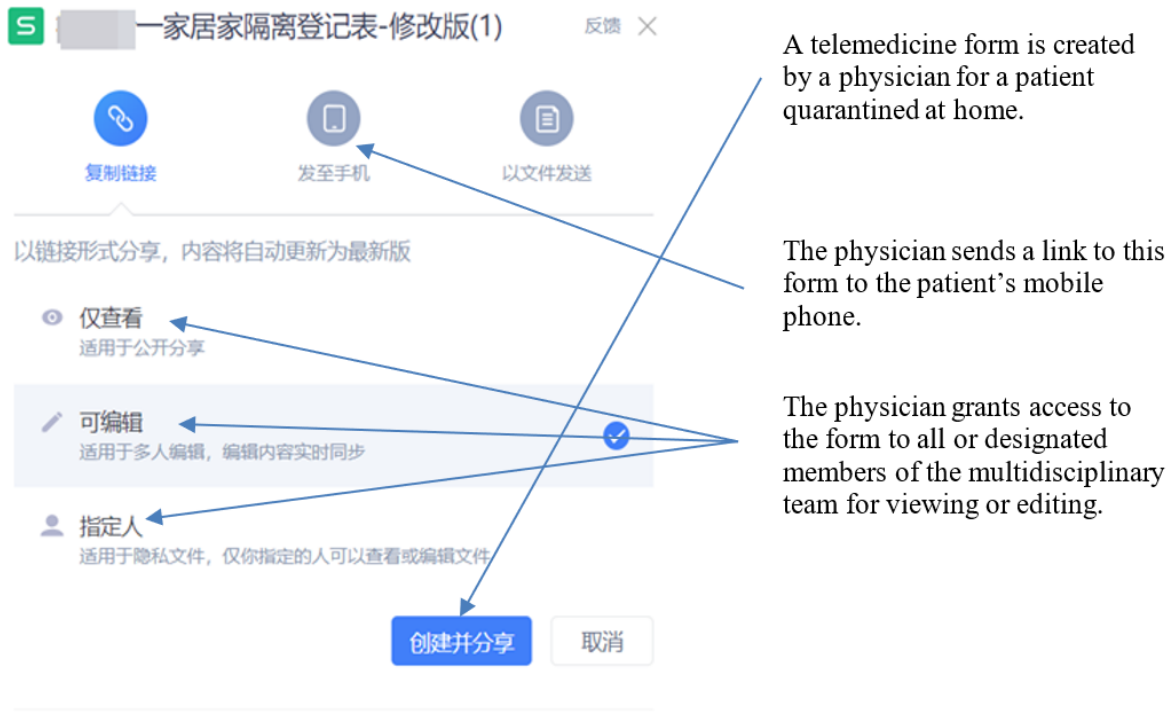
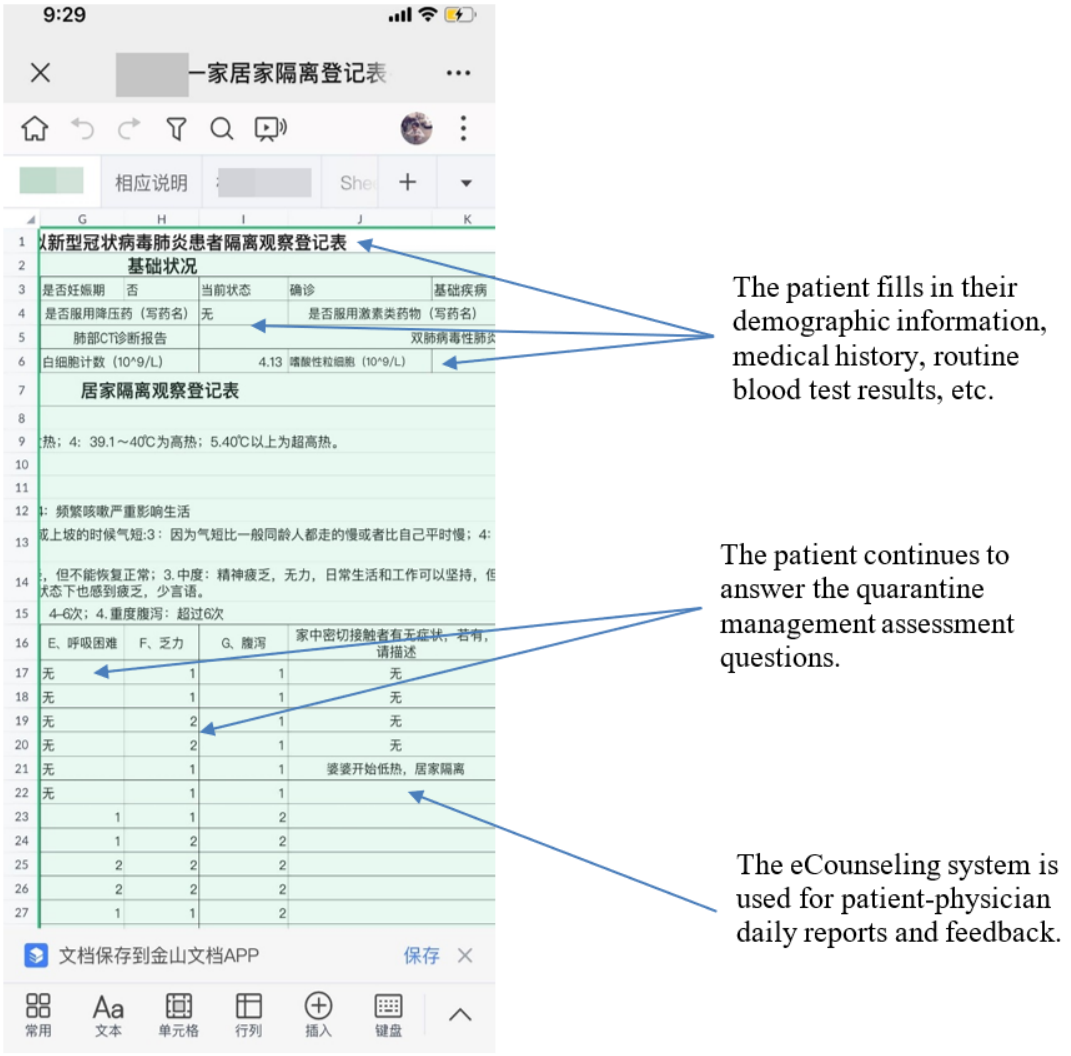


Figure 3. Screenshot of the creation of the telemedicine form by the multidisciplinary team (in Chinese, with English annotations).



XSL•FO
RenderX

Figure 4. Screenshot of the patient interface of the telemedicine form (in Chinese, with English annotations). eCounseling: electronic counseling.



Composition of the Multidisciplinary Team

The multidisciplinary team consisted of multidisciplinary medical workers, including 2 physicians, 3 nurses, 1 rehabilitation physician, and 1 psychologist. The physicians gave treatment advice according to the patients' condition updates through the eCounseling system. The nurses guided the methods of quarantine and disinfection, supervised the patient's regular work, rest, nutrition, and diet, and urged the patients to update their conditions in the telemedicine form on a daily basis. The rehabilitation physician helped the patients develop practical rehabilitation plans, and the psychologist helped them maintain a positive mood.

Participating Patients

The system initially recruited 188 individuals through word of mouth and physician referrals. Of these individuals, 114/188 (60.6%) showed no symptoms of COVID-19 during the 14-day quarantine period. These individuals were dismissed from the telemedicine system. The remaining 74 patients were confirmed COVID-19 cases. Among these patients, 26/74 (35%) left the system prematurely.

The multidisciplinary team followed up with every patient who dropped out; they found that all 26 patients had recovered and felt that their continued participation in the telemedicine system was unnecessary.

The observation endpoints for the patients in the telemedicine system were set as follows: patients were clinically cured (normal CT imaging reports plus a minimum of 2 rounds of negative results for SARS-CoV-2 nucleic acid detection), patients were admitted to hospital because of progression of the disease, or patients were deceased (although this did not occur in our study).

Diagnostic Criteria, Inclusion Criteria, and Exclusion Criteria of Patients With COVID-19

The diagnostic criteria for confirmed cases were defined as the presentation of either 1 of 2 pieces of etiological evidence: testing positive for SARS-CoV-2 nucleic acid in respiratory or blood samples by reverse transcriptase–polymerase chain reaction (RT-PCR) or virus sequences detected in respiratory or blood samples sharing high homology with the known sequence of SARS-CoV-2.

The inclusion criteria of the telemedicine system were set as confirmed or suspected cases of COVID-19 and voluntary

XSL•FO
RenderX

participation in medical observation. Patients who were not diagnosed with COVID-19 with certainty, pregnant or breastfeeding women, patients younger than 18 years or older than 75 years, and patients who were unable to cooperate with the data reporting were excluded from the analysis.

Results

Development of the Quarantine Management Assessment

We designed the quarantine management assessment scales using common symptoms presented by the patients with COVID-19 we treated at the Tongji Hospital clinic. We then gradually added new symptoms to the scales as our understanding of the disease advanced. The final version of the quarantine management assessment was designed based on these scales, which consisted of 5 primary indicators, 22 secondary indicators, and 83 tertiary indicators (Multimedia Appendix 1).

To calibrate the validity of the quarantine management assessment scales, they were submitted to 34 medical experts for evaluation. Two rounds of expert consultations were conducted. In each round, 17 experts from the emergency department, respiratory department, intensive care unit (ICU), and infectious disease department were invited to evaluate the scales. Statistical analysis of the consultation results showed that the response rates of the two rounds of expert consultations were 100% and 88.24%, respectively; the mean authoritative coefficient was 0.855. The W values for the degree of coordination of the expert opinions were 0.204 and 0.293 for the two rounds, respectively. The W values were also statistically significant (P=.003).

Demographics and Baseline Characteristics on Admission

Of the 48 patients in the study, 35 (73%) were female. The median age was 37.5 years (IQR 30.00-45.00). Fever was the most common initial symptom among the 48 patients (28, 58%), followed by cough (16, 33%), lack of strength (6, 13%), muscle soreness (5, 10%), and nasal congestion (3, 6%) (Table 1). One patient had no symptoms upon their initial clinic visit. Six patients were admitted to hospital later because their conditions worsened during home quarantine; all these patients had fever. Of the hospitalized patients, 3 had coughs and 3 also showed infiltration in both lungs at the time of diagnosis. However, 2 hospitalized patients exhibited normal CT imaging at initial diagnosis (Table 2).

XSL•FO
RenderX

Table 1. Demographics, baseline characteristics, and clinical outcomes of the home-quarantined patients with COVID-19 (N=48).

Characteristic	All patients (N=48)	Nonhospitalized patients (n=42)	Hospitalized patients (n=6)
Age (years), median (IQR)	37.50 (30.00-45.00)	35.50 (29.50-44.25)	55.00 (37.25-70.25)
Age (years), n (%)			
≤39	28 (58)	26 (62)	2 (33)
40-49	12 (25)	11 (26)	1 (17)
50-59	4 (8)	4 (10)	0 (0)
60-69	1 (2)	0 (0)	1 (17)
≥70	3 (6)	1 (2)	2 (33)
Gender, n (%)			
Male	13 (27)	10 (24)	3 (50)
Female	35 (73)	32 (76)	3 (50)
Initial symptoms, n (%)			
Headache	3 (6)	3 (7)	0 (0)
Muscle soreness	5 (10)	5 (12)	0 (0)
Cough	16 (33)	13 (31)	3 (50)
Diarrhea	4 (8)	4 (10)	0 (0)
Dyspnea	1 (2)	1 (2)	0 (0)
Sore throat	1 (2)	1 (2)	0 (0)
Dizziness	2 (4)	2 (5)	0 (0)
Chest tightness	4 (8)	3 (7)	1 (17)
Fever	28 (58)	22 (52)	6 (100)
Chest pain	1 (2)	1 (2)	0 (0)
Nasal congestion	3 (6)	3 (7)	0 (0)
Loss of appetite	2 (4)	2 (5)	0 (0)
Lack of strength	6 (13)	5 (12)	1 (17)
Number of initial symptoms, n (%)			
No symptoms	1 (2)	1 (2)	0 (0)
1 symptom	27 (56)	25 (60)	2 (33)
2 symptoms	13 (27)	10 (24)	3 (50)
3 symptoms	5 (10)	4 (10)	1 (17)
4 symptoms	2 (4)	2 (5)	0 (0)
CT[a] manifestation at onset, n (%)			
Normal	7 (15)	5 (12)	2 (33)
Ground-glass opacity in one lung	13 (27)	11 (26)	2 (33)
Ground-glass opacity in both lungs	19 (40)	19 (45)	0 (0)
Infiltration in both lungs	9 (19)	7 (17)	2 (33)
White blood cell count, n (%)			
Normal	25 (52)	22 (52)	3 (50)
Decreased	18 (38)	17 (41)	1 (17)
Elevated	5 (10)	3 (7)	2 (33)
Lymphocyte count, n (%)			
Normal	20 (42)	20 (48)	0 (0)

XSL•FO
RenderX

Characteristic	All patients (N=48)	Nonhospitalized patients (n=42)	Hospitalized patients (n=6)
Decreased	28 (58)	22 (52)	6 (100)

[a]CT: computed tomography.

Of the 6 hospitalized patients, 2 patients (33%) progressed to critical condition. One patient (labeled as Patient No. 3 in Table 2) was directly admitted to the ICU. ECMO support was provided to this patient for 17 days. He was transferred to the general ward when his condition stabilized. Eventually, the patient recovered (Table 2). Another hospitalized patient (labeled as Patient No. 4 in Table 2) was found to have a persistent fever on day 5 of observation. The patient's CT results identified ground-glass opacities in both lungs (Figure 5). Emergency hospital admission was requested for the patient. Noninvasive ventilation was provided, and the patient's condition gradually improved. This patient was discharged from hospital after 30 days of treatment (Table 2).

Based on the information collected through the eCounseling system, we found that there were differences in disease progression between patients with mild conditions and the patients who were hospitalized. The hospitalized patients had appreciably elevated body temperature at onset, which scored between 3 and 4 points (approximately 38-40 ºC) and persisted longer than that of nonhospitalized patients. The mean body temperature of the nonhospitalized patients became normal by day 4 or 5 (Figure 6A). In comparison, the body temperature of the hospitalized patients remained elevated on day 5 to 6 along with exacerbated cough (Figure 6B).

The results of the quarantine management assessment of symptom progression also provided direct insight into timely intervention for patients whose conditions deteriorated over time. The separation between the dyspnea curves was very distinct between nonhospitalized and hospitalized patients. For nonhospitalized patients with mild symptoms, dyspnea peaked on day 6 with a score of 2 to 3, which manifests as shortness of breath when hurrying on a level surface. However, these patients' symptoms gradually improved over time. In contrast, hospitalized patients had sustained dyspnea that continued to worsen over time (Figure 6C).

Although the degree of lack of strength appeared to be unrelated to disease severity at onset, this symptom was never alleviated among hospitalized patients and continued to worsen over time. In comparison, among nonhospitalized patients, strength was nearly normal by day 4 (Figure 6D). In addition, we observed that the mental state of hospitalized patients gradually deteriorated over the course of the disease, showing a similar pattern to lack of strength (Figure 6E). Furthermore, while muscle soreness was alleviated on day 4 for all patients, the degree of muscle soreness in hospitalized patients continued to be higher than normal (Figure 6F). The progression of diarrhea was inconsistent for both groups (Figure 6G). This may be due to the side effects of certain antiviral drugs. We also found that 4 days after the onset of illness, the chest tightness of hospitalized patients gradually worsened (Figure 6H).

XSL•FO
RenderX

Table 2. Clinical characteristics of the participating patients admitted to hospital (n=6) based on the medical observation scales described in Multimedia Appendix 1.

Characteristic	Patient No. 1	Patient No. 2	Patient No. 3	Patient No. 4	Patient No. 5	Patient No. 6
Age (years)	38	67	43	35	71	70
Gender	Female	Male	Male	Male	Female	Female
Presenting symptoms and signs at onset						
Fever[a]	Moderate	None	Moderate	Moderate	Low-grade	Moderate
Cough	Frequent and slightly interferes with daily activities	Occasional	Frequent and slightly interferes with daily activities	Occasional	Frequent and slightly interferes with daily activities	Frequent and slightly interferes with daily activities
Lack of strength	Moderate	Mild	Severe	Moderate	Mild	Severe
Diarrhea	Moderate	Mild	Mild	Mild	Moderate	None
Chest tightness	None	Moderate	Moderate	Moderate	Moderate	None
Dyspnea level[b]	1	2	3	2	3	4
Mental state	Poor	Average	Poor	Average	Average	Poor
Muscle soreness	Occasional	None	Occasional	None	Occasional	Light
Laboratory results at onset						
Procalcitonin	Normal	Normal	Normal	Normal	Normal	Normal
White blood cell count	Normal	Normal	Elevated	Decreased	Elevated	Normal
Lymphocyte percentage	Decreased	Decreased	Decreased	Decreased	Decreased	Decreased
CT[c] image characteristics at onset	Few ground-glass opacities in left lung	Few ground-glass opacities in right lung	Large infiltration shadow in both lungs	Normal	Multiple infections in both lungs	Normal
Number of days since onset before admission	5	7	1	5	1	7
Signs and symptoms at admission						
Fever	None	Moderate	Moderate	Low grade	Low grade	None
Cough	None	Frequent and slightly interferes with daily activities	Frequent and slightly interferes with daily activities	Frequent and slightly interferes with daily activities	Frequent and slightly interferes with daily activities	Occasional
Lack of strength	None	Severe	Severe	Mild	Mild	Mild
Diarrhea	None	Mild	Mild	None	Moderate	None
Chest tightness	None	None	Moderate	Mild	Moderate	Moderate
Dyspnea level	1	4	3	2	3	2
Mental state	Good	Poor	Poor	Poor	Average	Average
Muscle soreness	Light	Light	Occasional	Light	Occasional	None
Laboratory results at admission						
White blood cell count	Normal	Normal	Elevated	Normal	Elevated	Normal
Lymphocyte percentage	Decreased	Decreased	Decreased	Decreased	Decreased	Decreased
CT image characteristics at admission	Bilateral diffused patchy shadows	Bilateral diffused patchy shadows	Bilateral infiltration shadow	Bilateral multiple patches of infiltration shadows	Bilateral multiple infections	Bilateral multiple infections
Outcome	Discharged	Discharged	Discharged	Discharged	Discharged	Discharged

[a]Fever: None (37.3 degrees Celsius and below); Low grade (37.3-38 °C); Moderate (38.1-39 ºC).

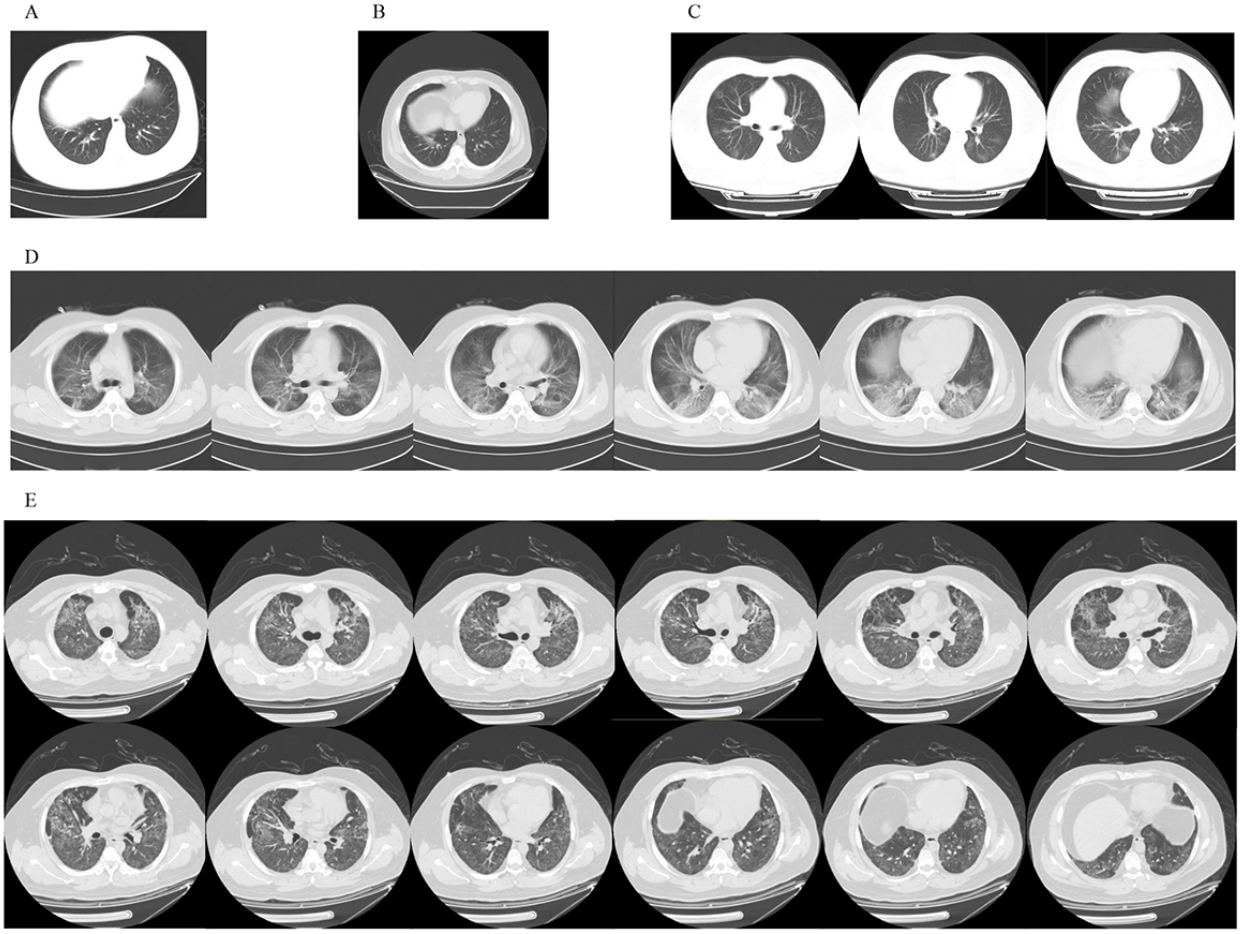
[b]Dyspnea level: 1=Not troubled by breathlessness except with strenuous exercise; 2=Troubled by shortness of breath when hurrying on a level surface or walking up a slight hill; 3=Experience breathlessness or must stop for breath when walking on a level surface at own pace; 4=Stop for breath after

XSL•FO
RenderX

walking 100 meters or after a few minutes on a level surface.

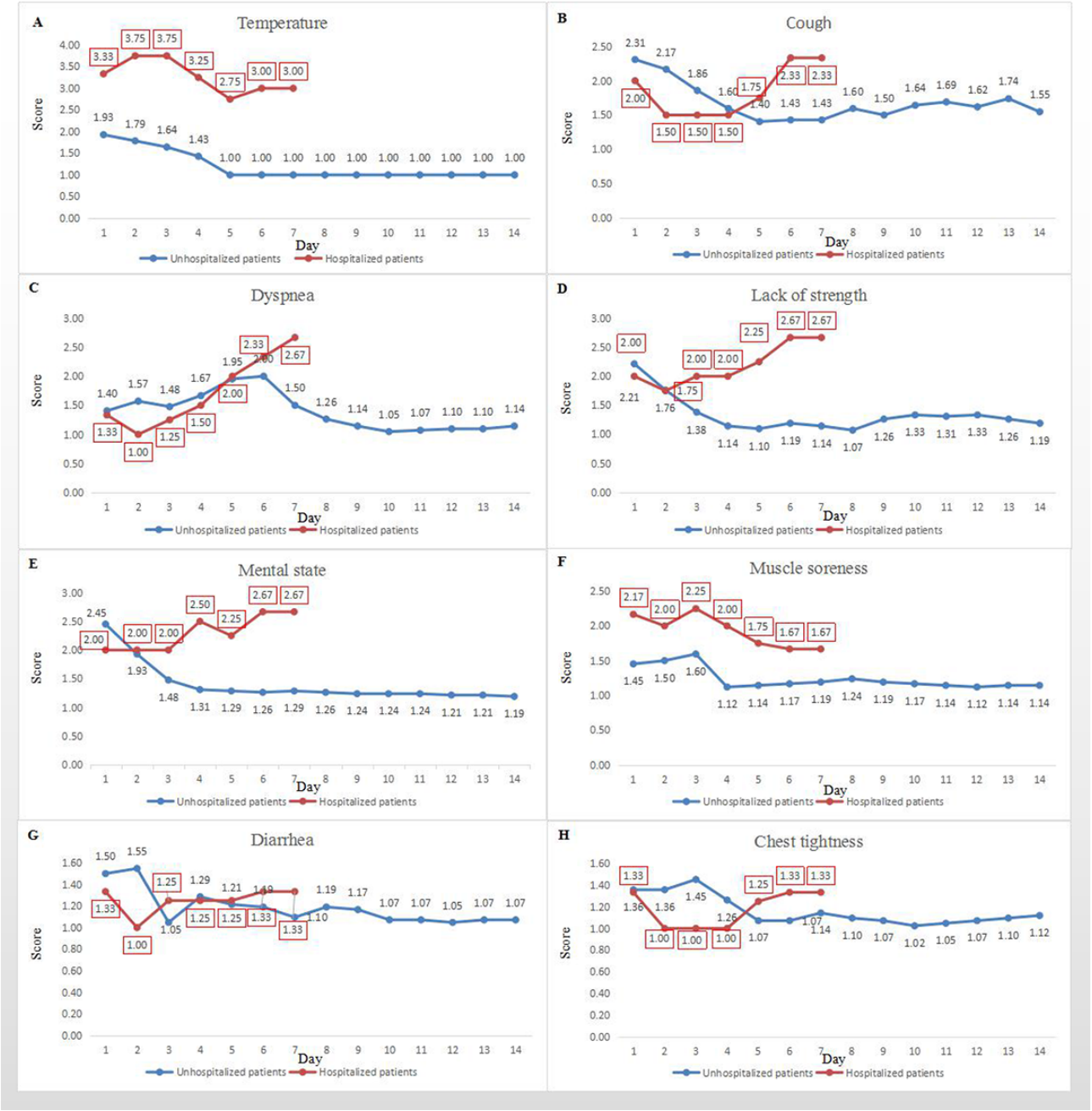
[c]CT: computed tomography.

Figure 5. Chest computed tomography (CT) images of Patient No. 4 showing small patchy shadows on the lower right lung on day 1 (A) and day 3 (B) after symptom onset. Chest CT images showing bilateral ground-glass opacities on day 6 after symptom onset (C) and bilateral large infiltrative shadows with partial consolidation on day 10 after symptom onset (D). Chest CT images showing bilateral diffuse reticular changes and fibrous stripes on day 45 after symptom onset (E).



XSL•FO
RenderX

Figure 6. Comparison of symptom trends between non-hospitalized and hospitalized patients.



Discussion

Principal Findings

World Health Organization (WHO) emergency guidelines recommend considering alternative quarantine methods, including homecare and isolation, for patients with COVID-19 who have mild symptoms in cases of insufficient hospitalization conditions or medical resources [3]. However, these guidelines do not provide details on how this home quarantine should be conducted, nor do they offer instructions on what to do when patients' conditions become severe or critical. A notable example is the British Prime Minister Boris Johnson, who was admitted to the ICU after 11 days of home quarantine [10]. More concerningly, based on initial clinic symptoms and laboratory tests, it is difficult to distinguish patients whose conditions will later become severe or critical. For example, CT results were normal for 7/48 (15%) of the patients in our study during initial diagnosis, including 2 patients who were hospitalized. Our results were consistent with the findings of Guan et al [11], who reported that CT images were normal for 17.9% of patients with mild conditions and 2.9% patients with severe conditions, respectively. Thus, it may be a common challenge to identify patients during an initial clinic screening whose conditions are prone to become severe or critical. Constant observation of home-quarantined patients by medical staff may therefore be of lifesaving importance.

The telemedicine model presented in this paper not only fills the gap in the WHO guidelines on home quarantine but also mitigates the subsequent threats of the disease caused by a lack of initial symptoms. Specifically, through the quarantine

XSL•FO
RenderX

management assessment, the telemedicine system can complement the initial clinic screening and, hence, greatly increase the accuracy of diagnosis. Through the eCounseling system, the telemedicine system can detect any newly emerged symptoms; then, the multidisciplinary team can be promptly informed to make appropriate treatment decisions.

To demonstrate the merits of the telemedicine system, fever can be considered as an example. Current studies list fever as an indispensable or highly prevalent symptom during the initial phase of COVID-19 infection [12-16]. However, in the study by Guan et al [11], fever was only present in about half (48.7%) of the patients during initial diagnosis. The telemedicine system proposed in this paper helped patients report their elevated body temperatures later to medical staff, who then could monitor the changes in body temperature over time to detect any alarming patterns. We found that hospitalized patients had elevated body temperature that was sustained considerably longer than that of nonhospitalized patients. In contrast, the body temperature of nonhospitalized patients returned to normal by day 4 to 5. This difference can be detected only through continuous observation implemented by the telemedicine system or another similar system. Thus, the telemedicine system greatly aids medical staff in making correct treatment decisions without being misguided by the initial clinic diagnosis.

The telemedicine system also contributes to clinical practice by identifying the key roles of dynamic changes in four diagnostic symptoms: fever, dyspnea, lack of strength, and muscle soreness. Dyspnea peaked on day 6 for nonhospitalized patients but persisted in hospitalized patients and was exacerbated over time. Similarly, both lack of strength and muscle soreness returned to normal by day 4 for nonhospitalized patients but not for hospitalized patients. When a patient reported via the eCounseling system that she was still experiencing the symptoms listed above after day 4, the multidisciplinary team went on alert and paid closer attention to that patient.

Further, these symptoms have not been fully studied in the literature on COVID-19. Therefore, tracing dynamic changes in the abovementioned symptoms paves the way for future studies to investigate whether these time markers can serve as turning points of the disease. In comparison, we found that excessive laboratory data and physical examination were not strictly necessary to evaluate patients with mild symptoms. Instead, the evaluation could be performed through patients' subjective initiative and active participation through self-monitoring of the disease.

Another contribution of this telemedicine system is the quarantine management assessment scales. This set of scales provides hands-on and easy-to-use self-diagnosis tools for home-quarantined patients. It also helps medical staff obtain more details about the clinical symptoms of the patients without the need for close physical contact in a clinic. We found that the quarantine management assessment worked well in identifying patients with disease progression.

From the resource management perspective, the telemedicine system enabled management of 188 individuals initially and 74 patients later by a team consisting of only 7 medical workers. None of the patients died, and none of the multidisciplinary team members were infected with COVID-19. Efficiency is important for regions whose medical workforce has been impacted by medical worker infection. Many infected medical workers were required to remain at home for at least two weeks after recovery. One merit of the telemedicine system is that medical staff who have recovered from COVID-19 and are self-quarantined can be enlisted to help. Thus, the shortage in the medical workforce can be alleviated, and other medical workers can be freed to fight COVID-19 on the front line. In conclusion, the implementation of such a system can optimize the usage of local medical resources and reduce cross-infection among medical workers and patients.

Limitations and Future Work

One limitation of this study is that its scope was restricted by the suddenness and complexity of the COVID-19 outbreak as well as the diversity and latency of the clinical manifestations of the disease. As a result, we were unable to obtain a larger sample size to achieve a higher level of validity of the findings. It is worth mentioning that we are currently testing the telemedicine system globally. As of April 9, 2020, 1421 patients worldwide are using this system, including 804 in Italy, 250 in the United Kingdom, 181 in France, and 168 in the United States. More than 60% of the participating patients are quarantined at home. We expect to report further findings in the future.

Conclusion

Continuous monitoring of patients with COVID-19 quarantined at home with a telemedicine system helped greatly reduce the risks of delayed hospitalization due to disease progression. Through this system, medical workers can trace changes in several key symptoms and intervene in the home quarantine in time for hospitalization. The telemedicine system proposed in this study was proven to be effectual and efficient. Implementation of this system will optimize the usage of local medical resources and reduce cross-infection among medical workers and patients.

Acknowledgments

This study was supported by the Special Fund for Novel Coronavirus Pneumonia from the Science and Technology Department, Hubei Province (Grant No. 2020FCA035) and the COVID-19 Prompt Response Research Special Project from Huazhong University of Science and Technology (Grant No. 2020kfyXGYJ023). The funding organizations had no role in the study design, data collection, analysis, interpretation, or writing of the paper.

XSL•FO
RenderX

Authors' Contributions

HX and SH designed the study, had full access to all data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. HX and SH contributed equally and share first authorship. CQ, SL, JD, HX, and LY contributed to the data analysis, data interpretation, literature search, and writing of the manuscript. BJ, LA, YX, MB, and LY had roles in patient recruitment, data collection, data acquisition, and clinical management. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quarantine management assessment for home-quarantined patients with COVID-19.
[DOC File , 41 KB - jmir_v22i7e19514_app1.doc]

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-04-20]
2. Truog RD, Mitchell C, Daley GQ. The Toughest Triage - Allocating Ventilators in a Pandemic. N Engl J Med 2020 May 21;382(21):1973-1975. [doi: 10.1056/NEJMp2005689] [Medline: 32202721]
3. Home care for patients with COVID-19 presenting with mild symptoms and management of their contacts. World Health Organization. 2020 Mar 17. URL: https://www.who.int/publications/i/item/home-care-for-patients-with-suspected-novel-coronavirus-(ncov)-infection-presenting-with-mild-symptoms-and-management-of-contacts [accessed 2020-03-17]
4. Huang S, Xiao Y, Yan L, Deng J, He M, Lu J, et al. Implications for Online Management: Two Cases with COVID-19. Telemed J E Health 2020 Apr;26(4):487-494. [doi: 10.1089/tmj.2020.0066] [Medline: 32233973]
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
6. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. N Engl J Med 2020 Feb 27;382(9):872-874 [FREE Full text] [doi: 10.1056/NEJMc2001272] [Medline: 31991079]
7. Cai JH, Zong WH. Analysis on the Application of Medical O2O Model in China. Chinese Journal of Health Informatics and Management 2015 Jan:346. [doi: 10.3969/j.issn.1672-5166.2015.04.02]
8. Coronavirus disease (COVID-19) technical guidance: Patient management. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/patient-management [accessed 2020-02-07]
9. Diagnosis and treatment of pneumonia infected by the new novel coronavirus (the trial Fifth Edition). Webpage in Chinese. National Health Commission of the People's Republic of China. 2020 Feb 04. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml [accessed 2020-06-25]
10. Coronavirus: Boris Johnson admitted to hospital over virus symptoms. BBC News. 2020 Apr 06. URL: https://www.bbc.com/news/uk-52177125 [accessed 2020-06-25]
11. Guan W, Ni Z, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
12. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
13. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Feb 07:1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
14. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
15. Chen L, Liu HG, Liu W, Liu J, Liu K, Shang J, et al. Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia. Article in Chinese. Zhonghua Jie He He Hu Xi Za Zhi 2020 Mar 12;43(3):203-208. [doi: 10.3760/cma.j.issn.1001-0939.2020.03.013] [Medline: 32164089]
16. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]

XSL•FO
RenderX

Abbreviations

COVID-19: coronavirus disease
CT: computed tomography
ECMO: extracorporeal membrane oxygenation
eCounseling: electronic counseling
ICU: intensive care unit
RT-PCR: reverse transcriptase–polymerase chain reaction
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 21.04.20; peer-reviewed by H Li, J Zhang, K Kreiner; comments to author 27.05.20; accepted 21.06.20; published 02.07.20.

Please cite as:
Xu H, Huang S, Qiu C, Liu S, Deng J, Jiao B, Tan X, Ai L, Xiao Y, Belliato M, Yan L
Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study
J Med Internet Res 2020;22(7):e19514
URL: https://www.jmir.org/2020/7/e19514
doi:10.2196/19514
PMID:32568727

©Hui Xu, Sufang Huang, Chun Qiu, Shangkun Liu, Juan Deng, Bo Jiao, Xi Tan, Ling Ai, Yaru Xiao, Mirko Belliato, Li Yan. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 02.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Racial and Ethnic Digital Divides in Posting COVID-19 Content on Social Media Among US Adults: Secondary Survey Analysis

Celeste Campos-Castillo[1], PhD; Linnea I Laestadius[2], PhD, MPP

[1]Department of Sociology, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

[2]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Corresponding Author:
Celeste Campos-Castillo, PhD
Department of Sociology
University of Wisconsin-Milwaukee
3210 N Maryland Ave
Milwaukee, WI, 53201
United States
Phone: 1 414 229 1113
Email: camposca@uwm.edu

Abstract

Background: Public health surveillance experts are leveraging user-generated content on social media to track the spread and effects of COVID-19. However, racial and ethnic digital divides, which are disparities among people who have internet access and post on social media, can bias inferences. This bias is particularly problematic in the context of the COVID-19 pandemic because due to structural inequalities, members of racial and ethnic minority groups are disproportionately vulnerable to contracting the virus and to the deleterious economic and social effects from mitigation efforts. Further, important demographic intersections with race and ethnicity, such as gender and age, are rarely investigated in work characterizing social media users; however, they reflect additional axes of inequality shaping differential exposure to COVID-19 and its effects.

Objective: The aim of this study was to characterize how the race and ethnicity of US adults are associated with their odds of posting COVID-19 content on social media and how gender and age modify these odds.

Methods: We performed a secondary analysis of a survey conducted by the Pew Research Center from March 19 to 24, 2020, using a national probability sample (N=10,510). Respondents were recruited from an online panel, where panelists without an internet-enabled device were given one to keep at no cost. The binary dependent variable was responses to an item asking whether respondents "used social media to share or post information about the coronavirus." We used survey-weighted logistic regressions to estimate the odds of responding in the affirmative based on the race and ethnicity of respondents (white, black, Latino, other race/ethnicity), adjusted for covariates measuring sociodemographic background and COVID-19 experiences. We examined how gender (female, male) and age (18 to 30 years, 31 to 50 years, 51 to 64 years, and 65 years and older) intersected with race and ethnicity by estimating interactions.

Results: Respondents who identified as black (odds ratio [OR] 1.29, 95% CI 1.02-1.64; *P*=.03), Latino (OR 1.66, 95% CI 1.36-2.04; *P*<.001), or other races/ethnicities (OR 1.33, 95% CI 1.02-1.72; *P*=.03) had higher odds than respondents who identified as white of reporting that they posted COVID-19 content on social media. Women had higher odds of posting than men regardless of race and ethnicity (OR 1.58, 95% CI 1.39-1.80; *P*<.001). Among men, respondents who identified as black, Latino, or members of other races/ethnicities were significantly more likely to post than respondents who identified as white. Older adults (65 years or older) had significantly lower odds (OR 0.73, 95% CI 0.57-0.94; *P*=.01) of posting compared to younger adults (18-29 years), particularly among those identifying as other races/ethnicities. Latino respondents were the most likely to report posting across all age groups.

Conclusions: In the United States, members of racial and ethnic minority groups are most likely to contribute to COVID-19 content on social media, particularly among groups traditionally less likely to use social media (older adults and men). The next step is to ensure that data collection procedures capture this diversity by encompassing a breadth of search criteria and social media platforms.

(J Med Internet Res 2020;22(7):e20472) doi:10.2196/20472

XSL•FO
RenderX

KEYWORDS

COVID-19; digital divides; user characteristics; race; ethnicity; algorithm bias; social media; bias; surveillance; public health

Introduction

Background

Amid the novel coronavirus disease (COVID-19) pandemic of 2020, US adults are turning to social media to consume and share information [1]. Public health surveillance experts are already leveraging the data created by social media users to track the spread and effects of COVID-19 across populations [2-6]. Sets of social media data are also appearing in open access repositories to facilitate this work [7]. In light of COVID-19, several scientific journals are fast-tracking the peer-review process, creating an unprecedented pace of publication for new social media surveillance studies [8,9]. This in turn is accelerating the availability of analyses that can be used to inform interventions and policies aimed at combating the direct and indirect effects of COVID-19 in society.

Despite the potential benefits of social media data in providing real-time insight into COVID-19, important limitations remain. As noted in prior studies, representativeness is an ongoing challenge for social media surveillance systems [10-12]. Among the factors shaping the representativeness of social media data are racial and ethnic digital divides [13-15], in which members of racial and ethnic minority groups are less likely to access the internet and use social media [16]. In turn, the absence of social media data from racial and ethnic minority groups may bias findings from social media surveillance systems. In the context of the COVID-19 pandemic, biases due to the underrepresentation of social media data from members of racial and ethnic minority groups are particularly problematic because structural inequalities have generated an environment in which these groups are among the most vulnerable [17-23].

In this study, we analyzed a nationally representative survey of US adults conducted in late March 2020 to determine how race and ethnicity, and their intersections with gender and age, are associated with posting content related to COVID-19 on social media. Demographic intersections of age and gender with race and ethnicity are rarely inspected in studies characterizing social media users to understand potential biases in data sets [13,15]. However, it is particularly critical to examine these intersections during the COVID-19 pandemic because age and gender are additional fault lines shaping both social media use and the risk of experiencing direct and indirect effects of COVID-19. This analysis of digital divides can potentially offer guidance to researchers conducting public health surveillance studies who are concerned about biases in the social media data they are analyzing.

Inequalities in Social Media Use and Effects of COVID-19

At present, little is known about how social media posting about COVID-19 varies across racial and ethnic groups. This is particularly problematic given the context of the COVID-19 pandemic. Within the United States, members of racial and ethnic minority groups experience heightened susceptibility to serious complications from COVID-19 because of underlying health and health care inequalities predating the pandemic [17-19]. Racial and ethnic minority groups are also overrepresented among essential workers, which elevates their risk of exposure to COVID-19 [21]. Further, they are more susceptible to indirect effects of the pandemic, such as job loss or pay cuts [20].

Digital divides raise questions about whether members of racial and ethnic minority groups are posting COVID-19 content on social media. Racial and ethnic minority groups in the United States have only recently made gains in internet access, narrowing the advantage held by white people [14,24]. Among people with internet access, members of racial and ethnic minority groups use social media at approximately the same rate as white people or more, depending on the platform [13,15,25]. Further, it has been suggested that the COVID-19 pandemic will exacerbate existing digital divides, due in part to the economic impacts of the pandemic and reduced access to public, workplace, and school internet connections [26]. The findings of public health surveillance tools leveraging social media data to track the COVID-19 pandemic may thus be biased toward underrepresenting the experiences of racial and ethnic minority groups.

Alternatively, members of racial and ethnic minority groups may be most likely to post COVID-19 content on social media. Given their higher risk of contracting COVID-19 and experiencing its deleterious effects, members of these groups are also more likely to struggle with mental health during the pandemic [23,27]. In response, they may turn to social media as a coping mechanism or to seek social support [28,29]. Social media data would then offer an opportunity for public health surveillance systems to understand racial and ethnic inequities during the pandemic.

Age is another axis of inequality for both social media use and complications arising directly and indirectly from COVID-19. Older adults (65 years and older) are less likely to access the internet and use it to share content on social media [30,31]. However, this demographic group was identified early in the pandemic as being at high risk for complications from COVID-19 [32] as well as for experiencing social isolation from mitigation efforts such as stay-at-home restrictions [33,34]. The intersection between race/ethnicity and age is also important because the older adult population is currently majority white but will become increasingly diverse in the next few years [35]. Public health surveillance tools relying on social media data must consider the impacts of these shifting demographics on digital divides.

Lastly, gender inequalities are also critical for identifying potential biases present in social media data. Women in the United States have outgained men in internet access [14,36]. Among internet users, women are generally more likely than men to use social media [13,15]; thus, it is probable that they are overrepresented in social media data. This digital divide in which women make greater use of social media may be

XSL•FO
RenderX

problematic given that men appear to be at increased risk from COVID-19 complications [37]. Regarding the indirect effects of COVID-19, there are compelling reasons to expect distinct experiences for men and women. For example, men are more likely than women to rely on the workplace for confidants [38,39]. As a result, men who are working remotely during the pandemic, lose their job, or experience a reduction in work hours may have mental wellness impacts caused by diminished access to their usual sources of social support. By contrast, women are more likely to work as frontline nursing staff and to suffer indirect effects due to changing work patterns, increased exposure to intimate partner violence, and increased childcare obligations [40,41]. Race and ethnicity likely further modify these gender differences in experiences during the pandemic [42]; thus, an intersectional approach is required when analyzing which people are likely to post content on social media.

To date, no studies have considered how social media posting about COVID-19 differs by race and ethnicity, how these differences intersect with gender and age, or the implications of these patterns for public health surveillance efforts. Accordingly, the current study reports findings from a secondary analysis of a nationally representative survey fielded late March 2020.

Methods

Data Source

The data we analyzed were obtained from the Pew Research Center, a nonpartisan think tank that conducts regular surveys to track demographic trends. We analyzed the Wave 64 survey of their American Trends Panel, which is a probability-based online panel of US adults (people aged 18 years or older) [43]. Of the 15,433 invited panelists, 11,537 responded (74.8% response rate) in either English or Spanish from March 19 to 24, 2020. Among these panelists, 164/11,537 (1.4%) did not have an internet-enabled device but received an internet-enabled tablet to keep at no cost. The results shown are from analyzing 10,510 respondents, with complete responses on all measures used. Missing values for any measure were <1%, except for income, for which 3% of values were missing. An analysis using a multiple imputed data set (10 iterations) to fill in missing income values afforded results that are comparable to those shown here.

Posting COVID-19–Related Content

The dependent measures were responses to an item asking whether they "used social media to share or post information about the coronavirus" (1=yes, 0=no).

Race/Ethnicity, Gender, and Age

We measured these three demographics using the original groupings provided in the data set. Race and ethnicity were based on the respondent's identification with one of four mutually exclusive categories (white, black, Latino, other). Gender was measured as a sex binary (female, male). Age was measured using four categories (18 to 29 years, 30 to 49 years, 50 to 64 years, 65 years or older).

Covariates

Other measures associated with social media use and differential exposure to the direct and indirect effects of COVID-19 served as covariates in the multivariable analysis. The set of covariates directly measures experiences with economic hardship and struggles with mental health during the pandemic along with sociodemographic factors shaping differential exposure to these and other stressors, such as additional household labor and misinformation [13,15,29,30,44,45]. Covariates were respondents' marital status (never married; currently married/cohabitating; divorced, widowed, or separated), annual family income (less than US $30,000, US $30,000-$74,999, greater than US $75,000), educational attainment (high school or less, some college, college graduate), political leaning (very liberal, liberal, moderate, conservative, very conservative), and mental health, which is an average of five items modified from the Center for Epidemiologic Studies Depression Scale [46] and General Anxiety Disorder Scale [47] asking how frequently (less than 1 day, 1 to 2 days, 3 to 4 days, 5 to 7 days) they experienced the following during the seven days preceding the survey (α=.73): nervous, anxious, or on edge; depressed; lonely; hopeful about the future; trouble sleeping. Higher values indicate poorer mental health. Household characteristics were census division, metropolitan area (yes, no), presence of a child less than 12 years of age (yes, no), presence of a household member who was laid off (yes, no), and presence of a household member who received a pay cut (yes, no). All groupings are original to the data set.

Statistical Analysis

After describing the analytical sample, we present results from three multivariable logistic regressions estimating the odds that respondents would report that they used social media to post COVID-19 content. The first estimated the association of respondents' race and ethnicity on the odds of posting content, adjusted for gender, age, and the covariates. The second examines how the intersection of race/ethnicity with gender shapes the odds of posting by estimating interactions between the racial and ethnic categories with gender. The estimated odds were adjusted for age and the covariates. The estimates were then used to compare the presence and severity of the gender divide by racial and ethnic group. The third regression followed a parallel procedure, enabling an assessment of the presence and severity of the age divide by racial and ethnic group. Specifically, we examined the intersection of race/ethnicity with age by estimating interactions between the racial and ethnic categories and the age groups. In these last two regressions, we corrected for making multiple comparisons.

All results shown (except for frequencies when describing the analytic sample) are based on weighting the data using the set of svy commands available in the statistical program Stata 16. The survey weights used in this process were provided by the Pew Research Center to reflect known US population counts on the following characteristics: age, race and ethnicity, gender, education, census region, metropolitan status, internet access, political party identification, volunteerism, registered voter, and years living in the United States.

XSL•FO
RenderX

Results

Descriptive data for the analytical sample appear in Table 1. Means and percentages are survey-weighted, but frequencies reflect the unweighted sample. Of the respondents in the weighted sample, 4383/10,510 (39.4%) reported posting COVID-19 content on social media. The majority of the 10,510 respondents (7012, 65%) were white, and most (6700, 63.7%) were between 30 and 64 years of age. A slight majority of respondents were women (5724/10,510, 51%).

Table 2 shows the associations between respondents' race and ethnicity and the odds of posting COVID-19 content on social media based on a multivariable logistic regression that adjusts for all covariates. Compared to the odds of posting for white respondents, black respondents had 29% higher odds (P=.03), Latino respondents had 66% higher odds (P<.001), and respondents identifying with other races or ethnicities had 33% higher odds (P=.03).

Table 2 also shows the independent associations of respondents' gender and age with their odds of posting. Estimates depict a gender divide favoring women and an age divide favoring respondents aged 18 to 29 years. Women had 58% higher odds than men to report that they posted COVID-19 content (P<.001). Compared to respondents aged 18 to 29 years, respondents who are 65 and older were significantly less likely to report posting (P=.01).

The next two steps in the analysis examined the intersections between race/ethnicity and gender and age by estimating statistical interactions. Full estimation results are available in Multimedia Appendix 1. Here, we summarize key findings using figures plotting predicted probabilities and discrete changes in predicted probabilities, showing 95% confidence intervals around estimates.

The first two figures summarize how the intersections of race and gender are associated with reports of posting COVID-19 content on social media. Figure 1 plots the predicted probabilities of posting COVID-19 content on social media for each cross-section. Among men, who were less likely to post (Table 2), respondents who identified as black (P=.009), Latino (P<.001), or other races/ethnicities (P=.003) were more likely to post than respondents who identified as white. Among women, only Latino respondents were significantly more likely to post than white respondents (P=.001). Figure 2 summarizes how the severity of the gender divide in posting COVID-19 content among respondents who identified as black, Latino, or other races/ethnicities compares to that of respondents who identified as white. Values crossing zero indicate comparable severity, while negative values indicate an attenuation of the divide. The figure shows a tendency for the gender divide to be slightly narrower among respondents identifying as either black (P=.06) or other races (P=.03); these findings did not hold when adjusting for multiple comparisons.

XSL•FO
RenderX

Table 1. Characteristics of sample respondents (N=10,510).

Characteristic	Value
Posted COVID-19[a] content on social media, n (%)	4383 (39.4)
Race/ethnicity, n (%)	
White	7012 (65.4)
Black	771 (10.3)
Latino	2145 (15.6)
Other	582 (8.7)
Female gender, n (%)	5724 (51.3)
Age (years), n (%)	
18-29	1219 (20.9)
30-49	3569 (35.4)
50-64	3131 (24.6)
≥65	2591 (19.1)
Annual family income (US $), n (%)	
<30,000	1897 (27.7)
30,000-74,999	3618 (36.8)
≥75,000	4995 (35.4)
Education, n (%)	
High school or less	1397 (33.8)
Some college	3141 (32.5)
College graduate	5972 (33.7)
Household member was laid off, n (%)	1833 (19.5)
Household member received pay cut, n (%)	2776 (27.4)
Mental health[b], mean (SD)	2.07 (0.72)
US citizen, n (%)	9999 (93.4)
Marital status, n (%)	
Never married	1823 (17.4)
Currently married or cohabitating	6787 (58.2)
Divorced, widowed, or separated	1900 (17.3)
Young child (<12 years) in household	2370 (24.9)
Political leaning, n (%)	
Very liberal	1030 (8.1)
Liberal	2275 (17.7)
Moderate	4074 (41.4)
Conservative	2347 (24.4)
Very conservative	784 (8.5)
In metropolitan area	9435 (87.2)
Census division, n (%)	
Pacific	1491 (14.5)
Middle Atlantic	1158 (12.6)
East North Central	1467 (14.7)
West North Central	689 (6.5)

XSL•FO
RenderX

Characteristic	Value
South Atlantic	2966 (21.4)
East South Central	458 (5.0)
West South Central	996 (10.9)
Mountain	817 (9.2)
New England	468 (5.1)

[a]COVID-19: coronavirus disease.

[b]Average of five items modified from the Center for Epidemiologic Studies Depression Scale and General Anxiety Disorder Scale. Higher values indicate poorer mental health.

XSL•FO
RenderX

Table 2. Associations between respondent characteristics and reported posting on social media about the COVID-19 pandemic.

Characteristic	Odds ratio (95% CI)	*P* value
Race/ethnicity		
White	Reference	N/A[a]
Black	1.29[b] (1.02-1.64)	.04
Latino	1.66[c] (1.36-2.04)	<.001
Other	1.33[b] (1.02-1.72)	.03
Gender		
Male	Reference	N/A
Female	1.58[c] (1.39-1.80)	<.001
Age (years)		
18-29	Reference	N/A
30-49	0.99 (0.80-1.23)	.94
50-64	0.95 (0.75-1.20)	.67
≥65	0.73[b] (0.57-0.94)	.01
Annual family income (US $)		
<30,000	Reference	N/A
30,000-74,999	0.98 (0.82-1.18)	.86
≥75,000	0.82[b] (0.67-0.99)	.04
Education		
High school or less	Reference	N/A
Some college	1.12 (0.94-1.33)	.20
College graduate	1.10 (0.92-1.30)	.29
Household member was laid off	1.17 (0.97-1.41)	.10
Household member received pay cut	1.05 (0.89-1.23)	.57
Mental health[d]	1.27[c] (1.15-1.40)	<.001
US citizen	0.69[b] (0.49-0.97)	.03
Marital status		
Never married	Reference	N/A
Currently married or cohabitating	1.16 (0.96-1.40)	.13
Divorced, widowed, or separated	1.18 (0.94-1.47)	.15
Young child (<12 years) in household	1.18 (1.00-1.41)	.06
Political leaning		
Very liberal	Reference	N/A
Liberal	0.80 (0.62-1.01)	.07
Moderate	0.66[c] (0.52-0.83)	<.001
Conservative	0.77[b] (0.6-0.99)	.045
Very conservative	0.70[b] (0.51-0.96)	.03
In metropolitan area	1.01 (0.82-1.25)	.90
Census division		
Pacific	Reference	N/A
Middle Atlantic	1.29[b] (1.01-1.64)	.04

XSL•FO
RenderX

Characteristic	Odds ratio (95% CI)	*P* value
East North Central	1.21 (0.95-1.53)	.12
West North Central	1.03 (0.77-1.38)	.85
South Atlantic	1.21 (0.98-1.50)	.07
East South Central	1.45[b] (1.04-2.03)	.03
West South Central	1.11 (0.85-1.44)	.44
Mountain	0.98 (0.73-1.30)	.87
New England	1.27 (0.91-1.77)	.16

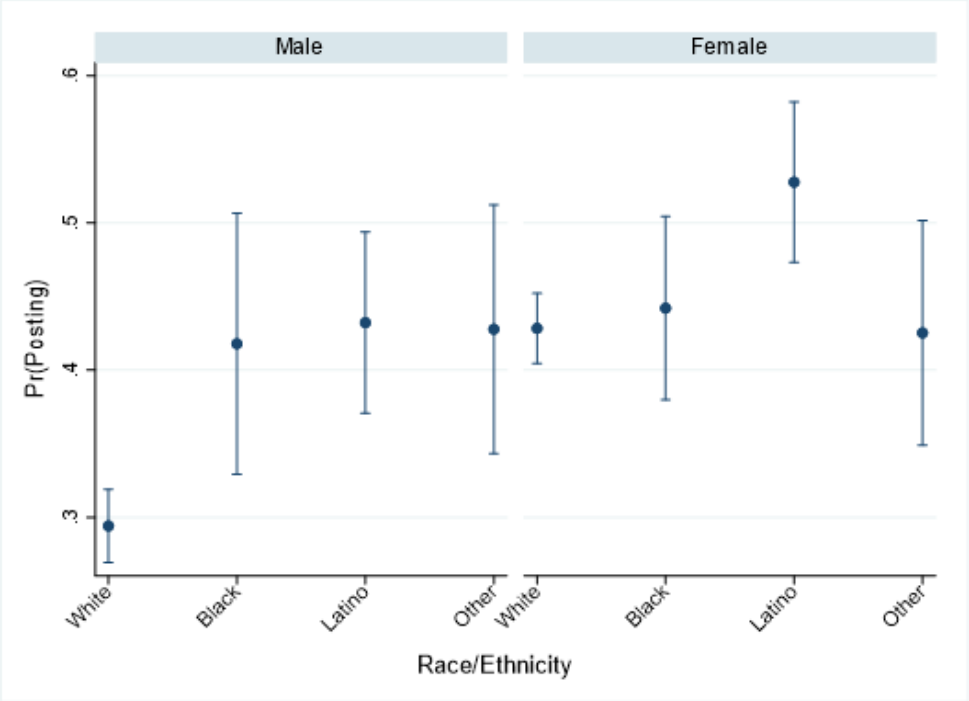
[a]N/A: not applicable.

[b]*P*<.05.

[c]*P*<.001.

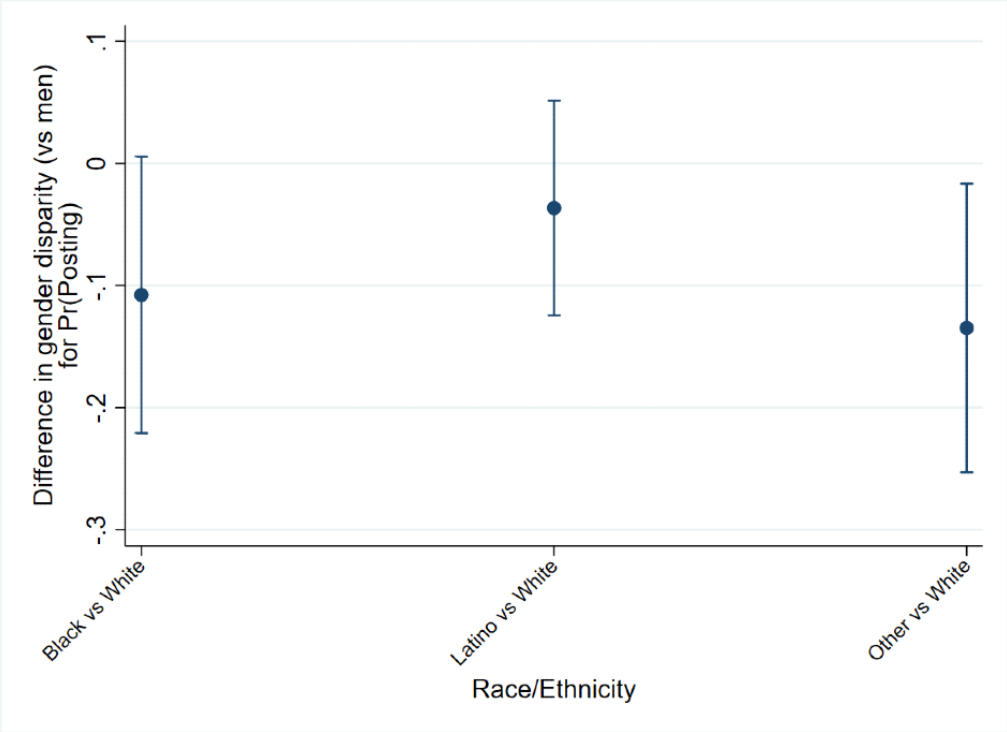
[d]Average of five items modified from the Center for Epidemiologic Studies Depression Scale and General Anxiety Disorder Scale. Higher values indicate poorer mental health.

Figure 1. Predicted probabilities of posting (95% CI) for each gender by race and ethnicity. Pr(Posting): probability of posting.



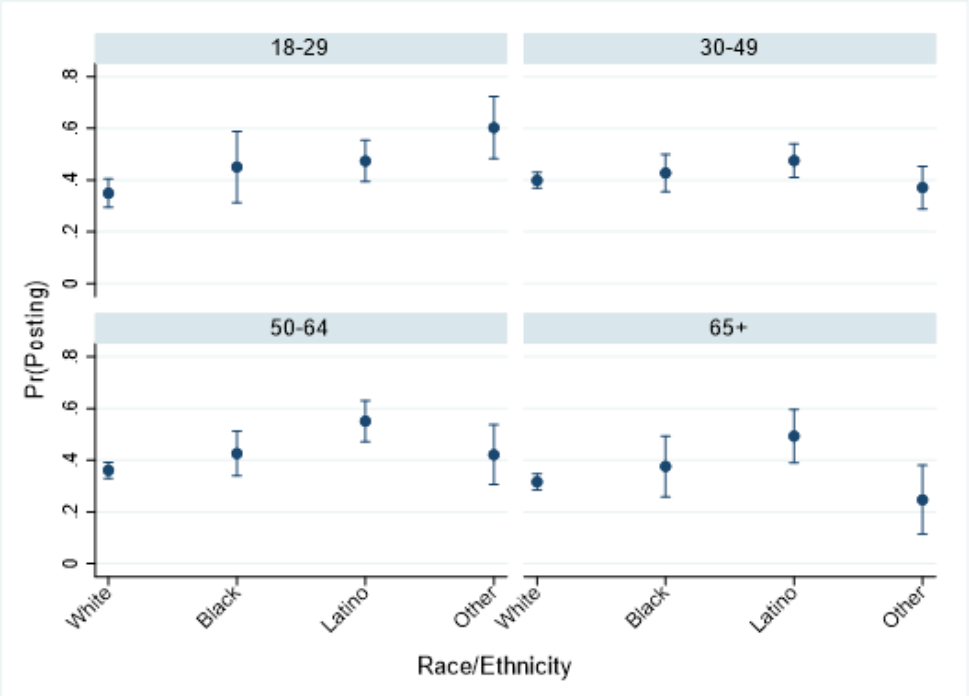
XSL•FO
RenderX

Figure 2. Differences in the gender divide in posting (95% CI) among respondents who identified as black, Latino, or other races/ethnicities relative to that of respondents who identified as white. Pr(Posting): probability of posting.



The remaining two figures offer a parallel summary of the intersections between the race of the respondents and their ethnicity and age. Figure 3 shows the predicted probabilities of posting for all four age groups by race and ethnicity of the respondents. Latino respondents were significantly more likely than white respondents across all age groups to report posting COVID-19 content on social media (18 to 29 years: P=.01, 30 to 49 years: P=.04, 50 to 64 years: P<.001, ≥65 years: P=.001). While respondents aged 65 years or older were less likely to post COVID-19 content than respondents aged 18 to 29 years (Table 2), those who did post content were likely to identify as Latino. The only other significant finding from this analysis was that among respondents aged 18 to 29 years, respondents identifying as a race/ethnicity other than black, white, or Latino had a higher probability of posting than white respondents (P<.001).

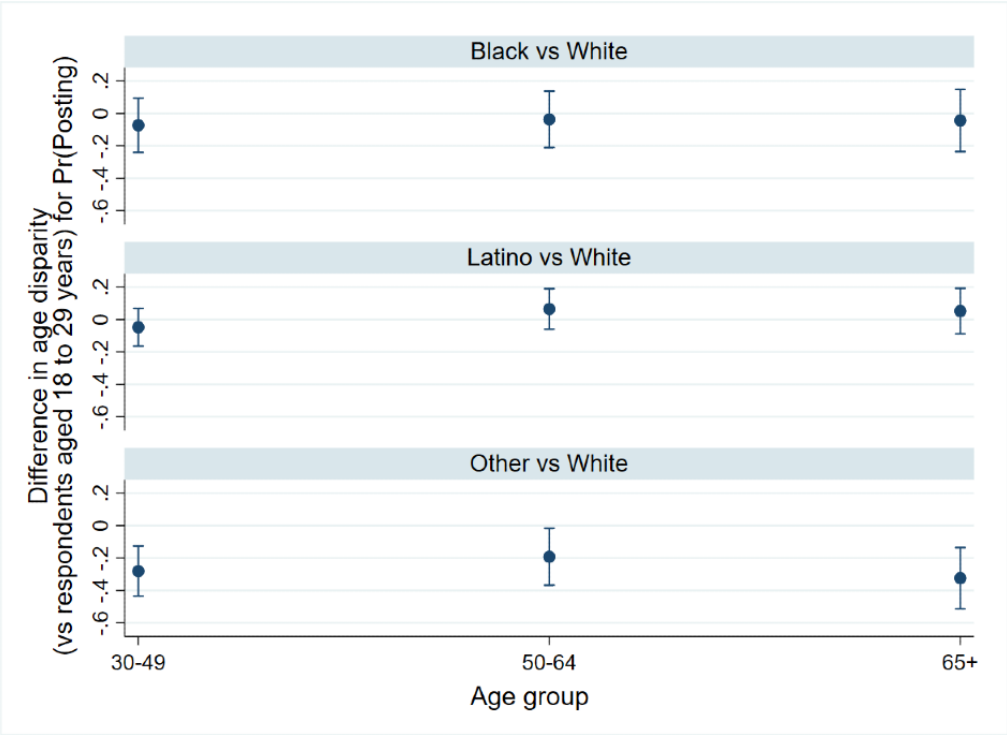
Figure 3. Predicted probabilities of posting (95% CI) for each age group by race and ethnicity. Pr(Posting): probability of posting.



XSL•FO
RenderX

Figure 4 continues the summary of the intersections between race and age by displaying the additional contributions of race and ethnicity to the age divide among respondents who post COVID-19 content on social media. The figure shows how the differences in the predicted probability between respondents aged 18 to 29 years and each of the older age groups change for respondents who identified as black, Latino, or other races/ethnicities relative to respondents who identified as white. A value crossing zero indicates comparable predicted differences in probability, with positive values indicating a wider age divide among the group compared to white respondents. Age divides were significantly intensified among respondents identifying as other races (31 to 49 years: $P<.001$, 50 to 64 years: $P=.04$, ≥65 years: $P=.001$), perhaps because of the heightened tendency of respondents aged 18 to 29 years from this racial/ethnic group to post (Figure 1). When adjusting for multiple comparisons, these conclusions held except among respondents aged 50 to 64 years. The severity of the age divide was comparable in all other cases.

Figure 4. Differences in the age divide in posting (95% CI) among respondents who identified as black, Latino, or other races/ethnicities relative to respondents who identified as white. Pr(Posting): probability of posting.



Discussion

Principal Findings

US public and public health surveillance analysts are both turning to social media during the COVID-19 pandemic. Specifically, members of the public are posting content related to the pandemic; in turn, analysts are using this content to track the spread of the virus and its effects across the population. However, racial and ethnic digital divides raise questions about the representativeness of the social media data that analysts are using to make inferences and projections. We found that among a nationally representative sample of US adults, members of racial and ethnic minority groups were most likely to report posting COVID-19 content on social media. Further, among men and older adults, who are traditionally less likely to use social media, members of racial and ethnic minority groups were also most likely to report posting content. Studies like this one that examine digital divides are critical to understanding the characteristics of people posting content to better infer the presence of biases in social media data. Other steps are likely necessary to fully address issues of representativeness in data, such as ensuring that data collection processes and search criteria are designed to capture diversity in the data.

Respondents who identified as black, Latino, or other races and ethnicities were more likely to report posting COVID-19 content on social media than respondents who identified as white. This finding may simply be an artifact of a reversal of digital divides, whereby white people are now least likely to use social media. A reversal is possible given trends observed during February 2019 in the United States [25]. While the estimated racial and ethnic differences from February 2019 cannot adequately account for the odds ratios in the current study, it is difficult to rule out this possibility without a demographic profile of social media users during the days preceding the pandemic. Alternatively, members of racial and ethnic minority groups may be most likely to post because the specific pandemic context (in which they are at higher risk of direct and indirect effects of COVID-19) inspires a stronger motivation to post. Specifically, our model controlled for financial ramifications from COVID-19 and mental health; however, there may be other COVID-19–related factors driving posting. Uses and gratifications theory suggests that social media use is often driven by a desire for self-expression, self-documentation, and

XSL•FO
RenderX

social connection [48]. Social media also offers a space for seeking social support and coping with stress [28,29]. Although there may be culturally driven distinctions in posting behaviors related to uses and gratifications, there is little prior research considering how these distinctions differ across race and ethnicity [49]. More work is needed to understand differences in posting behaviors, particularly to better understand COVID-19.

Regardless of the reason that members of racial and ethnic minority groups are most likely to post about COVID-19 on social media, our findings suggest that public health surveillance tools relying on social media data may be well positioned to capture their pandemic experiences and how they are related to structural inequalities within the United States. It will be critical for researchers to continue to both reflect on the overall representativeness of social media data and consider the nuances of social media use and data collection strategies when seeking data on specific populations. For example, a higher proportion of the Latino population than the white population uses Instagram, while these rates are more consistent for Facebook [25].

The intersections with gender and age were also enlightening for understanding digital divides during the pandemic. Women are generally more likely than men to use social media [13,15], and we saw this divide reflected across all racial and ethnic groups with respect to posting about COVID-19. Among men, those in racial and ethnic minority groups were most likely to report posting. These racial and ethnic differences are important to bear in mind depending on which gender disparities in COVID-19 experiences researchers seek to understand. With respect to age, studies consistently show that older adults are less likely to use social media [30], which was reflected in the findings of this study. This is important because of the elevated risks older adults face during the pandemic. Among respondents 65 years or older, Latino respondents are significantly more likely to post, suggesting an opportunity to leverage their social media data to understand their experiences. Such intersections with race and ethnicity are rarely investigated in studies on digital divides.

Limitations

This study is not without limitations. The dependent measure (posting content on social media related to COVID-19) is self-reported; thus, it is susceptible to recall bias. However, the survey was administered shortly after the outbreak occurred in the United States, which mitigates some of the potential bias. The measure also does not allow us to determine the content of posts or which social media platforms were used. Previous work showing racial and ethnic demographic variation by platform can offer some guidance [15]. Moreover, we cannot distinguish between users who post original content versus users who repost others' content (eg, a retweet on Twitter). Other datasets combining a user's social media data with their demographic profile are better suited for identifying users who are likely to repost [44]. Limitations such as these are common within studies relying on a secondary analysis of nationally representative data.

Conclusions

Despite limitations, the results of this examination of digital divides offer key insights into who is contributing to the social media content related to COVID-19 that analysts are studying. Because COVID-19 is exacerbating previously existing inequalities in the United States, members of racial and ethnic minority groups experience elevated risk of experiencing the deleterious health and economic effects of the virus. These populations are also most likely to post COVID-19–related content. There is also evidence of a heightened tendency to post among members of racial and ethnic minority groups who are members of other vulnerable groups that are usually less likely to use social media (people aged 65 years and older, men). These patterns suggest that social media data offer opportunities for public health researchers seeking to understand the impacts and spread of COVID-19 among racial and ethnic minority groups in the United States. We further suggest that it is not sufficient for researchers to rely only on broad social media usage demographics when reflecting on the representativeness of their data, as COVID-19 posting does not perfectly mirror what would be anticipated based on usage statistics alone [25].

To ensure that public health surveillance efforts take digital divides into account, studies like this one are a critical first step in that they provide demographic context for user-generated data. Additional steps are needed to realize this opportunity fully and mitigate bias. For example, collecting data in several languages, including Spanish, may ensure that the data reflect the patterns observed here among Latino respondents. A publicly available dataset of tweets associated with the pandemic is currently being collected [7]. The data collection process must incorporate a breadth of social media platforms and ensure that search criteria are reflective of the linguistic diversity in discourse about COVID-19. Given that the selection of search criteria shapes social media research findings [50], it is also critical for future research to map language use and how it differs across demographics. These steps are critical to meeting the challenges presented by the COVID-19 pandemic and understanding its lasting effects on the US population.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Estimates for multivariable logistic regressions.
[DOCX File , 28 KB - jmir_v22i7e20472_app1.docx]

XSL•FO
RenderX

References

1. Anderson M, Vogels E. Americans turn to technology during the COVID-19 outbreak, say an outage would be a problem. Pew Research Center. 2020 Mar 31. URL: https://www.pewresearch.org/fact-tank/2020/03/31/americans-turn-to-technology-during-covid-19-outbreak-say-an-outage-would-be-a-problem/ [accessed 2020-06-24]
2. Jimenez-Sotomayor MR, Gomez-Moreno C, Soto-Perez-de-Celis E. Coronavirus, Ageism, and Twitter: An Evaluation of Tweets about Older Adults and COVID-19. J Am Geriatr Soc 2020 Apr 27:online ahead of print [FREE Full text] [doi: 10.1111/jgs.16508] [Medline: 32338787]
3. Budhwani H, Sun R. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
4. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
5. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
6. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
7. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
8. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
9. Callaway E, Cyranoski D, Mallapaty S, Stoye E, Tollefson J. The coronavirus pandemic in five powerful charts. Nature 2020 Mar;579(7800):482-483. [doi: 10.1038/d41586-020-00758-2] [Medline: 32203366]
10. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
11. Aiello AE, Renson A, Zivich PN. Social Media- and Internet-Based Disease Surveillance for Public Health. Annu Rev Public Health 2020 Apr 02;41:101-118. [doi: 10.1146/annurev-publhealth-040119-094402] [Medline: 31905322]
12. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, et al. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? PLoS Curr 2016 Dec 07;8:ecurrents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6 [FREE Full text] [doi: 10.1371/currents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6] [Medline: 28123858]
13. Blank G. The Digital Divide Among Twitter Users and Its Implications for Social Research. Social Science Computer Review 2016 Oct 12;35(6):679-697. [doi: 10.1177/0894439316671698]
14. Campos-Castillo C. Revisiting the First-Level Digital Divide in the United States: Gender and Race/Ethnicity Patterns, 2007–2012. Social Science Computer Review 2014 Aug 21;33(4):423-439. [doi: 10.1177/0894439314547617]
15. Hargittai E. Potential Biases in Big Data: Omitted Voices on Social Media. Social Science Computer Review 2018 Jul 30;38(1):10-24. [doi: 10.1177/0894439318788322]
16. Hargittai E. Second-Level Digital Divide: Differences in People's Online Skills. First Monday 2002 Apr 01;7(4):online. [doi: 10.5210/fm.v7i4.942]
17. Millett G, Jones A, Benkeser D, Baral S, Mercer L, Beyrer C, et al. Assessing Differential Impacts of COVID-19 on Black Communities. Ann Epidemiol 2020 May 14:online ahead of print [FREE Full text] [doi: 10.1016/j.annepidem.2020.05.003] [Medline: 32419766]
18. Yancy CW. COVID-19 and African Americans. JAMA 2020 Apr 15:1891-1892. [doi: 10.1001/jama.2020.6548] [Medline: 32293639]
19. Laurencin CT, McClinton A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. J Racial Ethn Health Disparities 2020 Jun;7(3):398-402 [FREE Full text] [doi: 10.1007/s40615-020-00756-0] [Medline: 32306369]
20. Krogstad J, Gonzalez-Barrera A, Noe-Bustamante L. U.S. Latinos among Hardest Hit by Pay Cuts, Job Losses due to Coronavirus. Pew Research Center. 2020 Apr 03. URL: https://www.pewresearch.org/fact-tank/2020/04/03/u-s-latinos-among-hardest-hit-by-pay-cuts-job-losses-due-to-coronavirus/ [accessed 2020-06-24]
21. Bibbins-Domingo K. This Time Must Be Different: Disparities During the COVID-19 Pandemic. Ann Intern Med 2020 Apr 28 [FREE Full text] [doi: 10.7326/M20-2247] [Medline: 32343767]

XSL•FO
RenderX

22. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst Innovations in Care Delivery. 2020 May 04. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123 [accessed 2020-06-24]
23. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. JAMA 2020 Jun 03:online ahead of print [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
24. Internet/Broadband Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2020-06-24]
25. Social Media Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-06-24]
26. Beaunoyer E, Dupéré S, Guitton M. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 May 11:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
27. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Apr 13:online ahead of print. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
28. Davis M, Anthony D, Pauls S. Seeking and receiving social support on Facebook for surgery. Soc Sci Med 2015 Apr;131:40-47 [FREE Full text] [doi: 10.1016/j.socscimed.2015.02.038] [Medline: 25753284]
29. Wang J, Gaskin J, Rost DH, Gentile DA. The Reciprocal Relationship Between Passive Social Networking Site (SNS) Usage and Users' Subjective Well-Being. Soc Sci Comput Rev 2017 Jul 27;36(5):511-522. [doi: 10.1177/0894439317721981]
30. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media Soc 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
31. Anderson M, Perrin AJ. Tech Adoption Climbs Among Older Adults. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2020-06-24]
32. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3]
33. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256. [doi: 10.1016/s2468-2667(20)30061-x]
34. Steinman MA, Perry L, Perissinotto CM. Meeting the Care Needs of Older Adults Isolated at Home During the COVID-19 Pandemic. JAMA Intern Med 2020 Apr 16:online ahead of print. [doi: 10.1001/jamainternmed.2020.1661] [Medline: 32297903]
35. The Administration for Community Living. Profile of Older Americans. US Department of Health and Human Services. 2018. URL: https://acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2018OlderAmericansProfile.pdf [accessed 2020-06-24]
36. Ono H, Zavodny M. Gender and the Internet. Social Science Q 2003 Mar;84(1):111-121. [doi: 10.1111/1540-6237.00143]
37. Sharma G, Volgman AS, Michos ED. Sex Differences in Mortality from COVID-19 Pandemic: Are Men Vulnerable and Women Protected? JACC Case Rep 2020 May 04:online ahead of print [FREE Full text] [doi: 10.1016/j.jaccas.2020.04.027] [Medline: 32373791]
38. Campos-Castillo C, shuster S, Groh SM, Anthony DL. Warning: Hegemonic Masculinity May Not Matter as Much as You Think for Confidant Patterns among Older Men. Sex Roles 2020 Mar 7:online. [doi: 10.1007/s11199-020-01131-3]
39. McPherson M, Smith-Lovin L, Brashears ME. Social Isolation in America: Changes in Core Discussion Networks over Two Decades. Am Sociol Rev 2016 Jun 23;71(3):353-375. [doi: 10.1177/000312240607100301]
40. Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. Lancet 2020 Mar;395(10227):846-848. [doi: 10.1016/s0140-6736(20)30526-2]
41. Bahn K, Cohen J, Meulen Rodgers Y. A feminist perspective on COVID‐19 and the value of care work globally. Gender Work Organ 2020 Jun 11:online ahead of print. [doi: 10.1111/gwao.12459]
42. Khalatbari-Soltani S, Cumming RG, Delpierre C, Kelly-Irving M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. J Epidemiol Community Health 2020 May 08:online ahead of print [FREE Full text] [doi: 10.1136/jech-2020-214297] [Medline: 32385126]
43. Election News Pathways Survey. Pew Research Center. 2020. URL: https://www.pewresearch.org/pathways-2020/covidchg/total_us_adults/us_adults/ [accessed 2020-06-24]
44. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
45. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
46. Radloff LS. The CES-D Scale. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
47. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
48. Alhabash S, Ma M. A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? Social Media + Society 2017 Feb;3(1):205630511769154. [doi: 10.1177/2056305117691544]

XSL•FO
RenderX

49. Wang W, Segev S, Liu Y. Use of social networking sites for product communication: A comparative study of Hispanics and non-Hispanic whites. Computers Hum Behav 2015 Aug;49:86-93 [FREE Full text] [doi: 10.1016/j.chb.2015.02.051]
50. Rogers R. Doing Digital Methods. Los Angeles, CA: SAGE Publishing; Apr 2019.

Abbreviations

COVID-19: coronavirus disease

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by JP Allem, S Doan; comments to author 10.06.20; revised version received 15.06.20; accepted 21.06.20; published 03.07.20.

Please cite as:
Campos-Castillo C, Laestadius LI
Racial and Ethnic Digital Divides in Posting COVID-19 Content on Social Media Among US Adults: Secondary Survey Analysis
J Med Internet Res 2020;22(7):e20472
URL: https://www.jmir.org/2020/7/e20472
doi:10.2196/20472
PMID:32568726

©Celeste Campos-Castillo, Linnea I. Laestadius. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 03.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis

Peter E Lonergan[1*], MD, FRCS (Urol); Samuel L Washington III[1*], MD, MAS; Linda Branagan[2], PhD; Nathaniel Gleason[3,4], MD; Raj S Pruthi[1], MD, MHA; Peter R Carroll[1], MD, MPH; Anobel Y Odisho[1,2,3], MD, MPH

[1]Department of Urology, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, United States

[2]Telehealth Resource Center, University of California, San Francisco, San Francisco, CA, United States

[3]Center for Digital Health Innovation, University of California, San Francisco, San Francisco, CA, United States

[4]Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Anobel Y Odisho, MD, MPH
Department of Urology
Helen Diller Family Comprehensive Cancer Center
University of California, San Francisco
550 16th Street
San Francisco, CA, 94143
United States
Phone: 1 (415) 353 2200
Email: anobel.odisho@ucsf.edu

Abstract

Background: The emergence of the coronavirus disease (COVID-19) pandemic in March 2020 created unprecedented challenges in the provision of scheduled ambulatory cancer care. As a result, there has been a renewed focus on video-based telehealth consultations as a means to continue ambulatory care.

Objective: The aim of this study is to analyze the change in video visit volume at the University of California, San Francisco (UCSF) Comprehensive Cancer Center in response to COVID-19 and compare patient demographics and appointment data from January 1, 2020, and in the 11 weeks after the transition to video visits.

Methods: Patient demographics and appointment data (dates, visit types, and departments) were extracted from the electronic health record reporting database. Video visits were performed using a HIPAA (Health Insurance Portability and Accountability Act)-compliant video conferencing platform with a pre-existing workflow.

Results: In 17 departments and divisions at the UCSF Cancer Center, 2284 video visits were performed in the 11 weeks before COVID-19 changes were implemented (mean 208, SD 75 per week) and 12,946 video visits were performed in the 11-week post–COVID-19 period (mean 1177, SD 120 per week). The proportion of video visits increased from 7%-18% to 54%-72%, between the pre– and post–COVID-19 periods without any disparity based on race/ethnicity, primary language, or payor.

Conclusions: In a remarkably brief period of time, we rapidly scaled the utilization of telehealth in response to COVID-19 and maintained access to complex oncologic care at a time of social distancing.

(J Med Internet Res 2020;22(7):e19322) doi:10.2196/19322

KEYWORDS

health informatics; telehealth; video visits; COVID-19; video consultation; pandemic; electronic health record; EHR

Introduction

The emergence of the coronavirus disease (COVID-19) pandemic in the United States in March 2020 created unprecedented challenges in the provision of scheduled health care and, in particular, ambulatory cancer care. The rapid spread of COVID-19 has renewed focus on telehealth [1], including video consultations [2], as a means of continuing ambulatory care without increasing the risk of potential exposure for patients, clinicians, and staff.

XSL•FO
RenderX

Telehealth is the provision of health care remotely by means of a variety of telecommunication platforms such as messaging, audio, and video [3]. While the use of telehealth to deliver cancer care is not new [4] and has already been well described [5], delivering it at the current scale as a result of the COVID-19 pandemic is unprecedented.

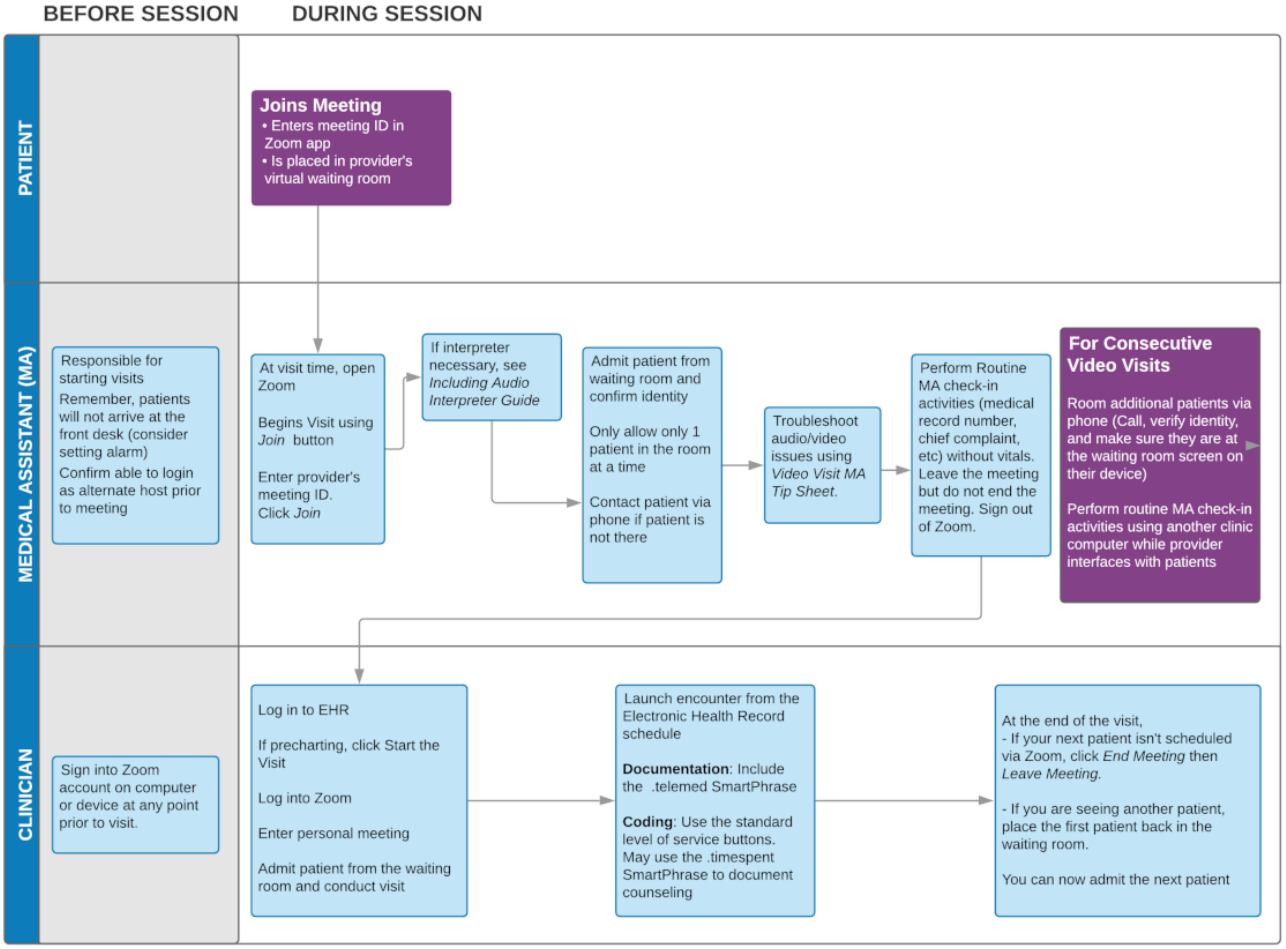
University of California, San Francisco (UCSF) Health established a telehealth program in 2015, which offers video visits in all practices. In response to the evolving pandemic, leadership challenged the organization to transition all in-person clinic visits, beginning March 15, 2020, to video visits with exceptions only for specific, urgent cases.

In this study, we analyze the change in video visit volume at the UCSF Comprehensive Cancer Center in response to COVID-19 and compare demographics and appointment data from January 1, 2020, and in the 11 weeks after the transition to video visits.

Methods

Demographics and appointment data (dates, visit types, and departments) were extracted from the electronic health record reporting database. The pre–COVID-19 period was defined as the 11 weeks from January 1 to March 13, 2020, prior to the transition, and the post–COVID-19 period as the 11 weeks following the transition to video visits on March 16, 2020, up to May 31, 2020. All video visits were performed using a HIPAA (Health Insurance Portability and Accountability Act)-compliant video conferencing platform (Zoom Video Communications Inc) with a pre-existing workflow (Figure 1). The statistical program R (version 3.5.3, R Foundation for Statistical Computing) was used for analysis [6] and a *P* value less than .05 was considered significant.

Figure 1. Video visit workflow used at the University of California, San Francisco.



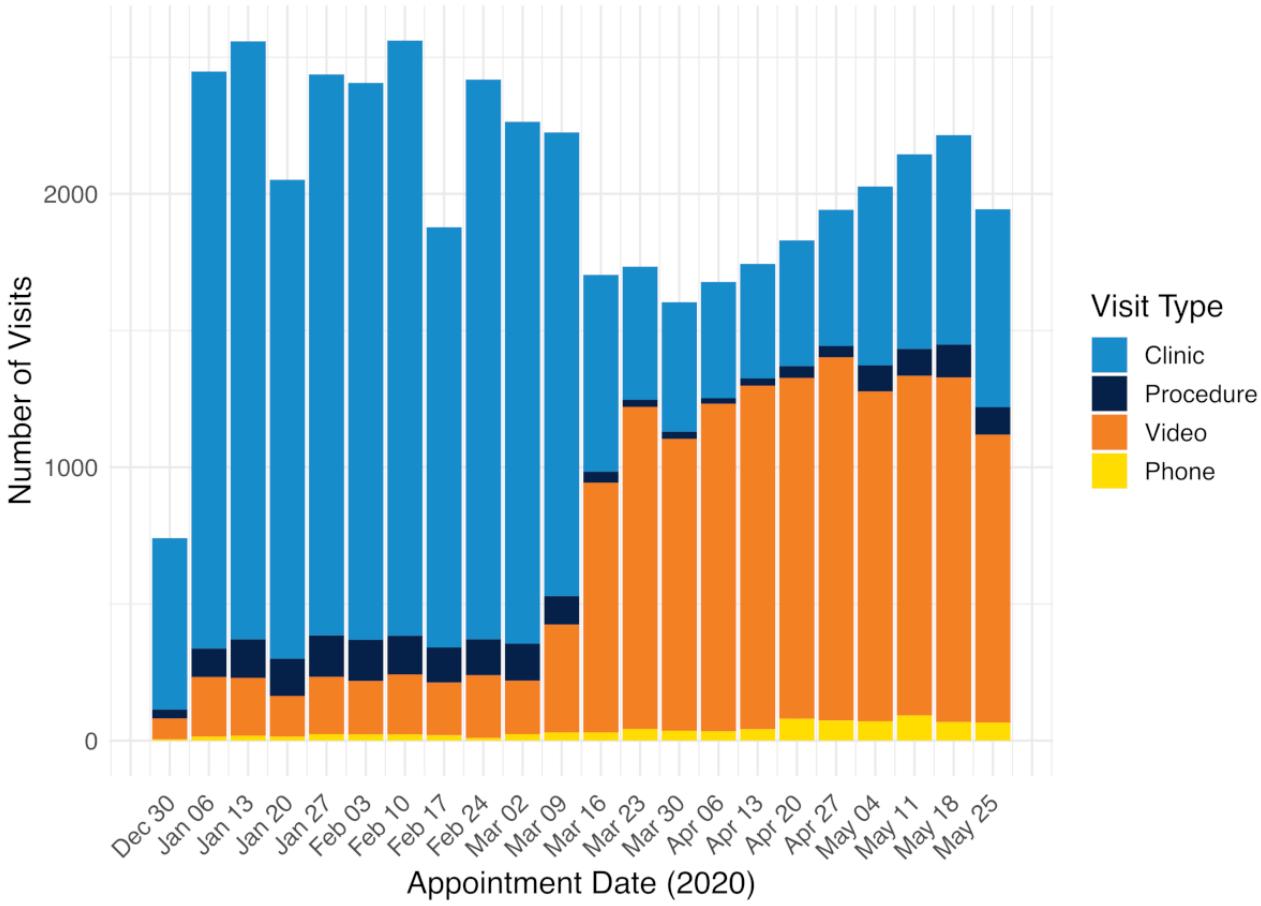
Results

In the pre–COVID-19 period from January 1 to March 13, 2020, there were a total of 23,988 ambulatory care episodes, with a mean of 2181 (SD 522) episodes per week across 17 departments and divisions at the UCSF Cancer Center (Figure 2). During this period, 2284 video visits were performed, with a mean of 208 (SD 75) video visits being performed per week. The proportion of video visits ranged from 7%-18% whereas the proportion of in-person visits ranged from 76%-86% (Figure 3). In the post–COVID-19 period from March 16 to May 31, 2020, there was a total of 20,567 ambulatory care episodes (mean 1870, SD 200 per week). A total of 12,946 video visits were performed in the post–COVID-19 period (mean 1177, SD 120 per week). The proportion of video visits increased to 54%-72%. The proportion of episodes during which a procedure was performed ranged from 4%-7% in the pre–COVID-19 period and 1%-5% in the post–COVID-19 period.

XSL•FO
RenderX

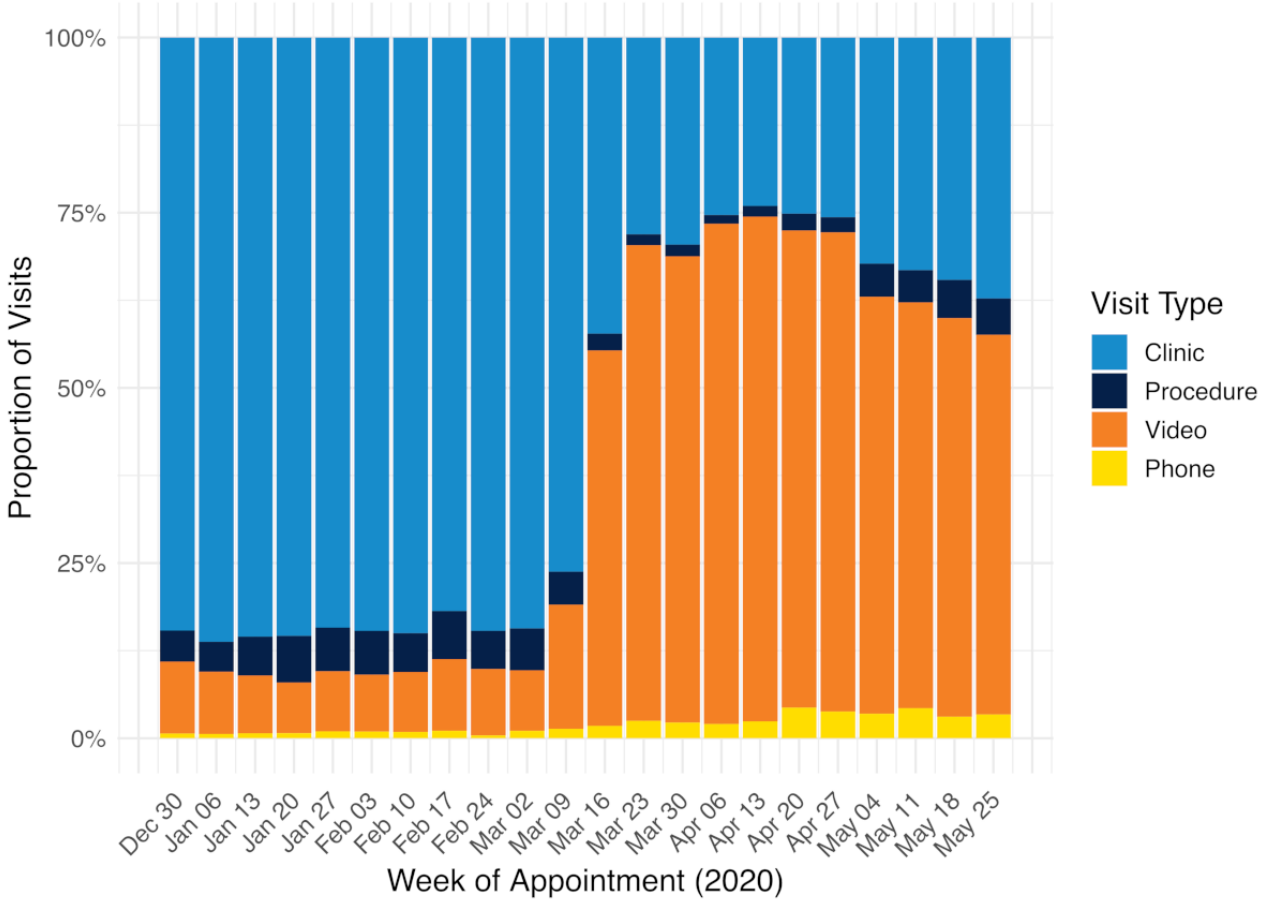
Table 1 displays the demographic data of patients who had a video visit in the pre–COVID-19 period (n=2284) and patients who had a video visit in the post–COVID-19 period (n=12,946). In the post–COVID-19 period, more black/African American (531 [4.1%] vs 78 [3.4%]; *P*<.001), Hispanic/Latino (1450 [11.2%] vs 215 [9.4%]; *P*<.001), and Asian (1903 [14.7%] vs 197 [8.6%]; *P*<.001) patients received care via video visits compared to the pre–COVID-19 period. There was increased post–COVID-19 utilization of video visits for patients in urban areas (12,014 [92.8%] vs 2026 [88.7%]; *P*<.001). We did not find any difference in the insurance status of patients using video visits during either period. In the post–COVID-19 period, first clinic encounter (2822 [21.8%] vs 417 [18.3%]; *P*<.001) and physician-provided visits (10,590 [81.8%] vs 1538 [67.3%]; *P*<.001) increased.

Figure 2. Number of clinic encounters stratified by visit type (in-person visits, procedural visits, video visits, and phone visits) from January 1 to May 31, 2020, with March 16, 2020, denoting the institution-wide transition to video visits in response to coronavirus disease (COVID-19).



XSL•FO
RenderX

Figure 3. Relative weekly trends in clinic encounters, stratified by visit type (in-person visits, procedural visits, video visits, and phone visits) from January 1 to May 31, 2020, with March 16, 2020, denoting the institution-wide transition to video visits in response to coronavirus disease (COVID-19).



XSL•FO
RenderX

Table 1. Demographics of patients who had video visits pre– and post–COVID-19.

Characteristic	Pre–COVID-19 period (n=2284)	Post–COVID-19 period (n=12,946)	*P* value
Age (years), median (IQR)	64.3 (54.9-71.5)	63.6 (52.8-71.8)	.11
Female, n (%)	844 (37)	6123 (47.3)	<.001
Ethnicity/race, n (%)			<.001
White	1606 (70.3)	7988 (61.7)	
Black/African American	78 (3.4)	531 (4.1)	
Hispanic/Latino	215 (9.4)	1450 (11.2)	
Asian	197 (8.6)	1903 (14.7)	
Other/unknown	188 (8.2)	1075 (8.3)	
Primary language: English	2212 (96.9)	11,962 (92.4)	<.001
Interpreter requested	84 (3.7)	945 (7.3)	<.001
Marital status, n (%)			.34
Married/partnered	1580 (69.2)	8415 (65.0)	
Single/separated	373 (16.3)	2369 (18.3)	
Unknown/declined	108 (4.7)	583 (4.5)	
Patient residence, n (%)			<.001
Urban	2026 (88.7)	12,014 (92.8)	
Rural	255 (11.2)	919 (7.1)	
Payor, n (%)			.14
Commercial	929 (40.7)	5398 (41.7)	
Medicare	1077 (47.2)	5709 (44.1)	
Medicaid	224 (9.8)	1541 (11.9)	
Self-pay	32 (1.4)	155 (1.2)	
First clinic encounter	417 (18.3)	2822 (21.8)	<.001
First video visit	1054 (46.2)	5865 (45.3)	.51
Provider type, n (%)			<.001
Physician	1538 (67.3)	10,590 (81.8)	
Advanced practice provider	746 (32.7)	2356 (18.2)	

Discussion

Principal Findings

We demonstrate a rapid expansion (from <20% to 72%) in telehealth use in a comprehensive cancer center over a remarkably brief time period in response to COVID-19 without differences in race or insurance type. Medicare telehealth visits have increased by more than 25% annually for the past decade [7], yet absolute adoption numbers remain low and fragmented with concerns about potentiating disparities in health care access [8]. The vast majority of cancer care cannot be delayed and the COVID-19 pandemic has presented new challenges that telehealth is uniquely situated to solve. Changes that would typically encompass months of planning, pilot testing, and education have been compressed into days. The use of telehealth has grown exponentially with some practices transitioning to near-complete virtual care in as little as a few days [9-11].

Several factors have likely enabled the rapid expansion of video visits at our institution. First, we had an established telehealth structure and workflow familiar to providers and practice staff. Second, UCSF made a strategic decision to provide work Relative Value Unit (wRVU) credit to providers for telehealth visits since 2015, irrespective of payer reimbursement. Third, new regulatory changes at the federal and state levels as a response to COVID-19 have reduced barriers, including the ability to see new patients (including Medicare beneficiaries) without a prior in-person visit to establish care and reimbursement for telehealth encounters by the Centers for Medicare & Medicaid Services (CMS) at parity with in-person visits beginning March 17, 2020 [12]. Finally, CMS now permits providers licensed in any state to provide telehealth services across the country [13].

Limitations

There are a number of limitations to this study that need to be acknowledged. We did not evaluate the outcome of the video

XSL•FO
RenderX

visits (eg, patient satisfaction and qualitative or clinical outcomes). The study is from a single, large, urban academic cancer center in the United States and our findings may not be generalizable to other specialties, practices, or locations. We believe that this is the first report of the utilization of video visits in a comprehensive cancer center in response to COVID-19 with a detailed description of the changing demographic of patients utilizing video visits before and after the COVID-19 pandemic.

Conclusions

Overall, the proportion of video visits increased from 7%-18% to 54%-72% between the pre– and post–COVID-19 periods while maintaining access to complex oncologic care at a time of social distancing. The COVID-19 pandemic has forced us to radically rethink and change our cancer care delivery models. In many health systems, there will undoubtedly be many lessons learned from this “natural experiment,” which has the potential to permanently change care delivery patterns.

Authors' Contributions

AYO had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. AYO, PEL, and SLW were responsible for study concept and design; and acquisition, statistical analysis, and interpretation of data. PEL and SLW drafted the manuscript. Administrative, technical, and material support was provided by LB, and supervision was provided by PRC, RSP, and AYO. All authors critically reviewed the manuscript for important intellectual content.

Conflicts of Interest

AYO was a consultant for VSee from December 2019 to January 2020.

References

1. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
2. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
3. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
4. Hazin R, Qaddoumi I. Teleoncology: current and future applications for improving cancer care globally. The Lancet Oncology 2010 Feb;11(2):204-210. [doi: 10.1016/s1470-2045(09)70288-8]
5. Pennell NA, Dicker AP, Tran C, Jim HSL, Schwartz DL, Stepanski EJ. mHealth: Mobile Technologies to Virtually Bring the Patient Into an Oncology Practice. Am Soc Clin Oncol Educ Book 2017;37:144-154 [FREE Full text] [doi: 10.1200/EDBK_176093] [Medline: 28561720]
6. R Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.R-project.org/
7. Mehrotra A, Jena AB, Busch AB, Souza J, Uscher-Pines L, Landon BE. Utilization of Telemedicine Among Rural Medicare Beneficiaries. JAMA 2016 May 10;315(18):2015-2016. [doi: 10.1001/jama.2016.2186] [Medline: 27163991]
8. Wilcock AD, Rose S, Busch AB, Huskamp HA, Uscher-Pines L, Landon B, et al. Association Between Broadband Internet Availability and Telemedicine Use. JAMA Intern Med 2019 Jul 29;179(11):1580. [doi: 10.1001/jamainternmed.2019.2234] [Medline: 31355849]
9. Mehrotra A, Ray K, Brockmeyer D, Barnett M, Bender J. Rapidly Converting to "Virtual Practices": Outpatient Care in the Era of Covid-19. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0091 [accessed 2020-06-14]
10. Spiegelman J, Krenitsky N, Syeda S, Sutton D, Moroz L. Rapid Development and Implementation of a Covid-19 Telehealth Clinic for Obstetric Patients. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0170 [accessed 2020-06-14]
11. Kasle DA, Torabi SJ, Savoca EL, Judson BL, Manes RP. Outpatient Otolaryngology in the Era of COVID-19: A Data-Driven Analysis of Practice Patterns. Otolaryngol Head Neck Surg 2020 May 12:194599820928987 [FREE Full text] [doi: 10.1177/0194599820928987] [Medline: 32393101]
12. Medicare Telemedicine Health Care Provider Fact Sheet. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2020-06-14]
13. COVID-19 Emergency Declaration Health Care Providers Fact Sheet. URL: https://www.cms.gov/files/document/covid19-emergency-declaration-health-care-providers-fact-sheet.pdf [accessed 2020-06-14]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
COVID-19: coronavirus disease
HIPAA: Health Insurance Portability and Accountability Act
UCSF: University of California, San Francisco

XSL•FO
RenderX

wRVU: work Relative Value Unit

Edited by G Eysenbach; submitted 15.04.20; peer-reviewed by I Mircheva, A Behmanesh, G Levin; comments to author 11.06.20; revised version received 16.06.20; accepted 21.06.20; published 06.07.20.

Please cite as:
Lonergan PE, Washington III SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, Odisho AY
Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis
J Med Internet Res 2020;22(7):e19322
URL: https://www.jmir.org/2020/7/e19322
doi:10.2196/19322
PMID:32568721

©Peter E Lonergan, Samuel L Washington III, Linda Branagan, Nathaniel Gleason, Raj S Pruthi, Peter R Carroll, Anobel Y Odisho. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Similarities and Differences in COVID-19 Awareness, Concern, and Symptoms by Race and Ethnicity in the United States: Cross-Sectional Survey

Jeb Jones[1], PhD, MPH, MS; Patrick S Sullivan[1], DVM, PhD; Travis H Sanchez[1], DVM, MPH; Jodie L Guest[1], PhD, MPH; Eric W Hall[1], PhD, MPH; Nicole Luisi[1], MPH, MS; Maria Zlotorzynska[1], MPH, PhD; Gretchen Wilde[2], MPH; Heather Bradley[3], PhD, MHS; Aaron J Siegler[2], PhD, MHS

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[2]Department of Behavioral, Social and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[3]Department of Epidemiology & Biostatistics, School of Public Health, Georgia State University, Atlanta, GA, United States

Corresponding Author:
Jeb Jones, PhD, MPH, MS
Department of Epidemiology
Rollins School of Public Health
Emory University
1518 Clifton Road
Atlanta, GA, 30322
United States
Phone: 1 4047122275
Fax: 1 4047128392
Email: jeb.jones@emory.edu

Abstract

Background: Existing health disparities based on race and ethnicity in the United States are contributing to disparities in morbidity and mortality during the coronavirus disease (COVID-19) pandemic. We conducted an online survey of American adults to assess similarities and differences by race and ethnicity with respect to COVID-19 symptoms, estimates of the extent of the pandemic, knowledge of control measures, and stigma.

Objective: The aim of this study was to describe similarities and differences in COVID-19 symptoms, knowledge, and beliefs by race and ethnicity among adults in the United States.

Methods: We conducted a cross-sectional survey from March 27, 2020 through April 1, 2020. Participants were recruited on social media platforms and completed the survey on a secure web-based survey platform. We used chi-square tests to compare characteristics related to COVID-19 by race and ethnicity. Statistical tests were corrected using the Holm Bonferroni correction to account for multiple comparisons.

Results: A total of 1435 participants completed the survey; 52 (3.6%) were Asian, 158 (11.0%) were non-Hispanic Black, 548 (38.2%) were Hispanic, 587 (40.9%) were non-Hispanic White, and 90 (6.3%) identified as other or multiple races. Only one symptom (sore throat) was found to be different based on race and ethnicity (P=.003); this symptom was less frequently reported by Asian (3/52, 5.8%), non-Hispanic Black (9/158, 5.7%), and other/multiple race (8/90, 8.9%) participants compared to those who were Hispanic (99/548, 18.1%) or non-Hispanic White (95/587, 16.2%). Non-Hispanic White and Asian participants were more likely to estimate that the number of current cases was at least 100,000 (P=.004) and were more likely to answer all 14 COVID-19 knowledge scale questions correctly (Asian participants, 13/52, 25.0%; non-Hispanic White participants, 180/587, 30.7%) compared to Hispanic (108/548, 19.7%) and non-Hispanic Black (25/158, 15.8%) participants.

Conclusions: We observed differences with respect to knowledge of appropriate methods to prevent infection by the novel coronavirus that causes COVID-19. Deficits in knowledge of proper control methods may further exacerbate existing race/ethnicity disparities. Additional research is needed to identify trusted sources of information in Hispanic and non-Hispanic Black communities and create effective messaging to disseminate correct COVID-19 prevention and treatment information.

(J Med Internet Res 2020;22(7):e20001) doi:10.2196/20001

XSL•FO
RenderX

KEYWORDS

COVID-19; SARS-CoV-2; race; ethnicity; awareness; concern; symptom; cross-sectional; knowledge; health disparity; inequality

Introduction

A novel coronavirus, severe acute respiratory syndrome 2 (SARS-CoV-2), was identified in Wuhan, China, in December 2019; the virus quickly spread worldwide and was labeled a pandemic by the World Health Organization on March 11, 2020 [1]. SARS-CoV-2 can result in severe respiratory infection, causing coronavirus disease (COVID-19). The rapid spread and high estimated infectivity of SARS-CoV-2 coupled with the severity of COVID-19 led to widespread shuttering of businesses and implementation of mandatory stay-at-home orders across the United States. Although reported cases are not stratified by race and ethnicity in all US jurisdictions or may be incompletely reported due to missing information, available data indicate the emergence of racial and ethnic disparities in the occurrence of COVID-19 [2], mirroring disparities that have been well characterized in other disease processes.

Racial and ethnic disparities in COVID-19 occurrence may be due to multiple factors. Asian, Black, and Hispanic Americans are all more likely to be uninsured than non-Hispanic White Americans [3], resulting in reduced access to health care. Increased prevalence of preexisting conditions that are known to be associated with poor outcomes in COVID-19 patients, including diabetes [4,5] among Black and Hispanic people and hypertension [5-7] among Black people, may increase susceptibility and disease severity in these populations. The long-term effects of structural racism and income inequality that cause larger proportions of people of color to work in conditions that are not conducive to social distancing [8] or to live in crowded conditions [8,9] also facilitate efficient transmission of the virus. In addition to combatting these existing disparities, in a rapidly evolving pandemic, accurate information is necessary for individuals to exercise appropriate prevention and care-seeking behavior. We examined whether there were racial or ethnic differences in knowledge, stigma, and experience of symptoms during the early stages of the COVID-19 pandemic.

Methods

Advertisements on Facebook, Snapchat, and Twitter were used to conduct a nationwide web-based survey with US adults aged 18 years or older from March 27 through April 1, 2020 [10]. Participants were eligible to participate if they were at least 18 years of age. We sought to obtain a racially and ethnically diverse sample by overrecruiting people of color and Hispanic participants. Therefore, because a sufficient number of non-Hispanic White participants had been enrolled, enrollment of non-Hispanic White participants was stopped on the final day of recruiting to increase the representation of communities of color in the survey. Participants were not compensated for completing the survey. All study procedures were reviewed and approved by the Emory University Institutional Review Board.

The survey included questions on knowledge of effective prevention methods, a COVID-19 stigma scale, and current COVID-19–related symptoms (see Multimedia Appendix 1). We used chi-square tests with Holm Bonferroni correction to account for multiple comparisons to assess differences across these domains based on race and ethnicity (Asian, Hispanic, non-Hispanic Black, non-Hispanic White, and other/multiple race). Corrected P values <.05 were considered statistically significant. The full survey is included in Multimedia Appendix 1. All analyses were conducted using SAS 9.4 (SAS Institute).

Results

A total of 1435 participants consented to participate and completed the survey. Of the 1435 participants, 52 (3.6%) were Asian, 548 (38.2%) were Hispanic, 158 (11.0%) were non-Hispanic Black, 587 (40.9%) were non-Hispanic White, and 90 (6.3%) were of other or multiple races. There were race and ethnicity differences in gender, annual income, educational status, and geographic region of residence across categories of race and ethnicity (Table 1).

When asked whether they were currently experiencing symptoms possibly related to COVID-19, a plurality of participants (635/1435, 44.3%) reported no symptoms (Table 2). Of those experiencing COVID-19 symptoms, the most frequently reported symptoms were cough (328/1435, 22.9%), sneezing (322/1435, 22.4%), and headache (311/1435, 21.7%). The only symptom that differed by race/ethnicity was sore throat. Racial/ethnic differences were observed in perceived likelihood of currently having COVID-19 by race/ethnicity. Hispanic (192/544, 35.3%) and non-Hispanic Black (54/157, 34.4%) participants were more likely to report that it was somewhat likely, likely, or very likely that they currently had COVID-19 compared to Asian (11/51, 21.6%) and non-Hispanic White (121/580, 20.9%) participants.

When asked about their knowledge of numbers of cases and COVID-19 deaths, non-Hispanic White (376/587, 64.1%) and Asian (32/52, 61.5%) participants were more likely to correctly estimate that there were currently 100,000 or more cases in the US compared to Hispanic (269/548, 49.1%) and non-Hispanic Black (73/158, 46.2%) participants. No differences were observed in the estimated number of deaths expected from COVID-19 by the end of 2020.

Non-Hispanic White (180/587, 30.7%) and Asian (13/52, 25.0%) participants were more likely to answer all 14 COVID-19 knowledge scale questions correctly compared to Hispanic (108/548, 19.7%) and non-Hispanic Black (25/158, 15.8%) participants. Overall, 46.6% of participants endorsed at least one stigmatizing statement related to COVID-19, with no difference by race.

Overall, 954/1247 participants (76.5%) indicated willingness to participate in future research studies of diagnostic and serologic testing for SARS-CoV-2, with no differences by race and ethnicity.

XSL•FO
RenderX

Table 1. Demographic characteristics of participants in a cross-sectional web-based survey of COVID-19 symptoms and knowledge from March 27 through April 1, 2020 (N=1435).

Characteristic	Total (N=1435)	Asian (n=52, 3.6%)	Hispanic (n=548, 38.2%)	Non-Hispanic Black (n=158, 11.0%)	Non-Hispanic White (n=587, 40.9%)	Other/multiple race (n=90, 6.3%)	*P* value[a]
Age (years), median (IQR)	33 (24-57)	25 (21-31)	30 (22-44)	25 (20-38)	54 (31-64)	31 (21-58)	.002
Gender, n (%)[b]							.002
Male	536 (40.2)	20 (41.7)	236 (48.4)	49 (36.8)	202 (34.4)	29 (37.7)	
Female	761 (57.1)	26 (54.2)	249 (51.0)	83 (62.4)	360 (61.3)	43 (55.8)	
Other	36 (2.7)	2 (4.2)	3 (0.6)	1 (0.8)	25 (4.3)	5 (6.5)	
Annual Income (US $), n (%)							.002
<30,000	376 (36.1)	9 (25.7)	150 (41.3)	54 (49.1)	139 (29.3)	24 (40.7)	
30,000-74,999	397 (38.1)	12 (34.3)	127 (35.0)	43 (39.1)	192 (40.5)	23 (39.0)	
≥75,000	268 (25.7)	14 (40.0)	86 (23.7)	13 (11.8)	143 (30.2)	12 (20.3)	
Education, n (%)							.004
High school or less	202 (16.7)	5 (10.4)	99 (22.9)	15 (11.8)	69 (12.9)	14 (19.4)	
At least some college	1011 (83.4)	43 (89.6)	333 (77.1)	112 (88.2)	465 (87.1)	58 (80.6)	
Region, n (%)							.002
Midwest	280 (19.6)	7 (13.5)	57 (10.4)	27 (17.1)	177 (30.3)	12 (13.3)	
Northeast	251 (17.5)	9 (17.3)	82 (15.0)	22 (13.9)	117 (20.0)	9 (17.3)	
South	540 (37.7)	8 (15.4)	215 (39.4)	86 (54.4)	200 (34.2)	8 (15.4)	
West	360 (25.2)	28 (53.9)	192 (35.2)	23 (14.6)	91 (15.6)	28 (53.9)	

[a]Holm Bonferroni *P* value to correct for multiple comparisons; results were considered statistically significant if corrected $P<.05$.

[b]Category totals do not sum to the total of 1435 due to missing data.

XSL•FO
RenderX

Table 2. Associations between race/ethnicity and symptoms, likelihood of current COVID-19 infection, estimates of the extent of the COVID-19 pandemic, knowledge, stigma, and interest in participating in research studies among participants in a web-based, cross-sectional survey conducted from March 27 through April 1, 2020 (N=1435).

Variable	Total (N=1435), n (%)	Asian (n=52, 3.6%), n (%)	Hispanic (n=548, 38.2%), n (%)	Non-Hispanic Black (n=158, 11.0%), n (%)	Non-Hispanic White (n=587, 40.9%), n (%)	Other/multiple race (n=90, 6.3%), n (%)	*P* value[a]
Symptoms in last 24 hours							
Fever	35 (2.4)	1 (1.9)	10 (1.8)	6 (3.8)	13 (2.2)	5 (5.6)	>.99
Cough	328 (22.9)	9 (17.3)	119 (21.7)	30 (19.0)	159 (27.1)	11 (12.2)	.07
Sneezing	322 (22.4)	11 (21.2)	124 (22.6)	37 (23.4)	129 (22.0)	21 (23.3)	>.99
Sore throat	214 (14.9)	3 (5.8)	99 (18.1)	9 (5.7)	95 (16.2)	8 (8.9)	.003
Headache	311 (21.7)	10 (19.2)	110 (20.1)	43 (27.2)	127 (21.6)	21 (23.3)	>.99
Shortness of breath	94 (6.6)	3 (5.8)	31 (5.7)	5 (3.2)	47 (8.0)	8 (8.9)	>.99
Diarrhea	108 (7.5)	3 (5.8)	42 (7.7)	9 (5.7)	47 (8.0)	7 (7.8)	>.99
Myalgia	103 (7.2)	2 (3.9)	36 (6.6)	7 (4.4)	54 (9.2)	4 (4.4)	>.99
Feeling of being unwell	175 (12.2)	8 (15.4)	69 (12.6)	17 (10.8)	70 (11.9)	11 (12.2)	>.99
No symptoms	635 (44.3)	23 (44.2)	246 (44.9)	75 (47.5)	251 (42.8)	40 (44.4)	>.99
COVID-19[b] likelihood[c]							.002
Very unlikely	356 (25.0)	8 (15.7)	129 (23.7)	37 (23.6)	157 (27.1)	25 (27.8)	
Unlikely	661 (46.5)	32 (62.8)	223 (41.0)	66 (42.0)	302 (52.1)	38 (42.2)	
Somewhat likely	324 (22.8)	9 (17.7)	149 (27.4)	41 (26.1)	103 (17.8)	22 (24.4)	
Likely	47 (3.3)	2 (3.9)	24 (4.4)	10 (6.4)	9 (1.6)	2 (2.2)	
Very likely	34 (2.4)	0 (0.0)	19 (3.5)	3 (1.9)	9 (1.6)	3 (3.3)	
Estimated number of current cases[d]							.004
<1000	28 (2.0)	1 (1.9)	14 (2.6)	4 (2.5)	9 (1.5)	0 (0.0)	
1000-9999	496 (34.6)	17 (32.7)	215 (39.2)	72 (45.6)	157 (26.8)	35 (38.9)	
10,000-99,999	113 (7.9)	2 (3.9)	50 (9.1)	9 (5.7)	45 (7.7)	7 (7.8)	
100,000-499,999	458 (31.9)	18 (34.6)	161 (29.4)	43 (27.2)	209 (35.6)	27 (30.0)	
500,000-999,999	139 (9.7)	3 (5.8)	52 (9.5)	8 (5.1)	69 (11.8)	7 (7.8)	
≥1,000,000	201 (14.0)	11 (21.2)	56 (10.2)	22 (13.9)	98 (16.7)	14 (15.6)	
Estimated number of deaths[c]							>.99
Fewer than 1000	55 (4.0)	2 (4.0)	20 (3.9)	8 (5.1)	20 (3.6)	5 (6.0)	
1000-10,000	266 (19.5)	8 (16.0)	82 (15.8)	38 (24.2)	125 (22.6)	13 (15.5)	
10,001-100,000	478 (35.0)	16 (32.0)	184 (35.4)	54 (34.4)	194 (35.0)	30 (35.7)	
100,000-1,000,000	430 (31.5)	17 (34.0)	186 (35.8)	48 (30.6)	152 (27.4)	27 (32.1)	
≥1,000,000	136 (10.0)	7 (14.0)	48 (9.2)	9 (5.7)	63 (11.4)	9 (10.7)	
Knowledge index							.002
<12	418 (29.1)	13 (25.0)	196 (35.8)	61 (38.6)	116 (19.8)	32 (35.6)	
12-13	675 (47.0)	26 (50.0)	244 (44.5)	72 (45.6)	291 (49.6)	42 (46.7)	
14	342 (23.8)	13 (25.0)	108 (19.7)	25 (15.8)	180 (30.7)	16 (17.8)	
Stigma index[c]							.09
0	722 (53.4)	31 (62.0)	242 (47.1)	71 (48.6)	329 (59.2)	49 (56.3)	
1-2	525 (38.8)	16 (32.0)	221 (43.0)	65 (44.5)	191 (34.4)	32 (36.8)	
≥3	106 (7.8)	3 (6.0)	51 (9.9)	10 (6.9)	36 (6.5)	6 (6.9)	

XSL•FO
RenderX

Variable	Total (N=1435), n (%)	Asian (n=52, 3.6%), n (%)	Hispanic (n=548, 38.2%), n (%)	Non-Hispanic Black (n=158, 11.0%), n (%)	Non-Hispanic White (n=587, 40.9%), n (%)	Other/multiple race (n=90, 6.3%), n (%)	*P* value[a]
Interest in learning about another study[c]							>.99
Yes	954 (76.5)	35 (72.9)	352 (78.4)	94 (71.2)	416 (76.5)	57 (77.0)	
No	293 (23.5)	13 (27.1)	97 (21.6)	38 (28.8)	128 (23.5)	17 (23.0)	

[a]Holm Bonferroni *P* value to correct for multiple comparisons; results were considered statistically significant if corrected *P*<.05.

[b]COVID-19: coronavirus disease.

[c]Category total does not sum to the column total due to missing data.

[d]Values greater than the population of the United States (328 million) were excluded. At the beginning of the study period (March 27, 2020), there were 107,000 cumulative confirmed cases in the United States. At the end of the study period (April 1, 2020), there were 213,400 cumulative confirmed cases in the United States.

Discussion

Principal Findings

We observed few differences in prevalent symptoms consistent with COVID-19 among a web-based sample of adults in the United States at the end of March through the beginning of April 2020. With respect to estimating the extent of the epidemic in the United States during the study period, there were 101,700 cumulative confirmed cases and 1600 cumulative deaths as of March 27, 2020 and 213,400 cumulative confirmed cases and 5000 cumulative deaths by April 1, 2020 [11].

Despite the similarity in experiences of COVID-19 symptoms, we did observe differences in participants' self-assessed likelihood of having COVID-19 at the time of survey completion. Hispanic and non-Hispanic Black participants were more likely to suspect that they were currently infected, which may reflect an awareness of the differential impact of the pandemic on communities of color in the United States or a better understanding of the potential for asymptomatic infection. These differential self-assessments of the likelihood of current infection may translate to differential testing and care-seeking behavior. Trends in self-assessment of symptoms and knowledge of appropriate prevention strategies should be continuously monitored to identify any persistent differences based on race and ethnicity.

Limitations

This study has limitations. Participants were recruited on the internet via advertisements on social media sites and are not representative of all adults in the United States. Participation in an uncompensated survey may reflect a prior interest in COVID-19; therefore, the respondents may have more COVID-19 knowledge compared to the general population. However, we do not think that this selection bias would be differential by race or ethnicity.

Conclusions

Racial and ethnic disparities in morbidity and mortality due to COVID-19 are already apparent. Black and Hispanic populations bear a disproportionate burden of medical conditions across their lifespans, including obesity, diabetes, and heart disease [4,5,12]. Worse health outcomes are exacerbated by structural racism, which in turn impacts housing, economic opportunities, education, transportation, food availability, and health care access [13]. Previously existing disparities in chronic disease incidence and prevalence due to unequal access and treatment [14] that are at least partially contributing to the observed COVID-19 disparities could increase if people are relying on incorrect prevention methods. Knowledge gaps in effective prevention methods and an inability to adhere to known best prevention methods due to crowding and work conditions have the potential to exacerbate these underlying disparities.

Addressing the structural racism that results in differential access to care will necessarily involve structural changes to health care delivery [15]. Universal access to COVID-19 testing, treatment, and if available, a SARS-CoV-2 vaccine will be critical but not sufficient. Medical mistrust must also be addressed [16]. Lack of trust in the medical establishment is an artifact of historical mistreatment of populations of color and of misinformation. It will be necessary to identify methods to overcome medical mistrust to ensure acceptance of treatment and prevention measures. Additionally, research is needed to identify trusted sources of information, as well as modes and mechanisms of communication, in Black and Hispanic communities to ensure that correct prevention and treatment information can be disseminated most effectively. Reducing disparities in knowledge, health care access, prevention, and treatment will be necessary conditions for successfully combatting the COVID-19 pandemic.

Acknowledgments

We appreciate and acknowledge the contributions of our study participants. This work was supported by the National Institute of Allergy and Infectious Diseases (3R01AI143875-02S1). The study was facilitated by the Center for AIDS Research at Emory University (P30AI050409). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

XSL•FO
RenderX

Authors' Contributions

All authors had full access to study data, and JJ and AJS had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full survey.
[PDF File (Adobe PDF File), 289 KB - jmir_v22i7e20001_app1.pdf]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-04-29]
2. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244 [FREE Full text] [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]
3. Berchick ER, Barnett JC, Upton RD. Health Insurance Coverage in the United States: 2018 Current Population Reports. United States Census Bureau. 2019 Nov. URL: https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf [accessed 2020-07-06]
4. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: Knowledge in progress. Diabetes Res Clin Pract 2020 Apr 09;162:108142 [FREE Full text] [doi: 10.1016/j.diabres.2020.108142] [Medline: 32278764]
5. Balfour PC, Ruiz JM, Talavera GA, Allison MA, Rodriguez CJ. Cardiovascular Disease in Hispanics/Latinos in the United States. J Lat Psychol 2016 May;4(2):98-113 [FREE Full text] [doi: 10.1037/lat0000056] [Medline: 27429866]
6. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020 May;109(5):531-538 [FREE Full text] [doi: 10.1007/s00392-020-01626-9] [Medline: 32161990]
7. Louisiana Coronavirus COVID-19. Louisiana Department of Health. URL: http://ldh.la.gov/Coronavirus/ [accessed 2020-05-08]
8. COVID-19 in Racial and Ethnic Minority Groups. US Centers for Disease Control and Prevention. 2020 Jun 25. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html [accessed 2020-06-23]
9. Burr JA, Mutchler JE, Gerst K. Patterns of residential crowding among Hispanics in later life: immigration, assimilation, and housing market factors. J Gerontol B Psychol Sci Soc Sci 2010 Nov;65(6):772-782 [FREE Full text] [doi: 10.1093/geronb/gbq069] [Medline: 20937707]
10. Siegler A, Hall E, Luisi N, Zlotorzynska M, Wilde G, Sanchez T. Willingness to seek laboratory testing for SARS-CoV-2 with home, drive-through, and clinic-based specimen collection locations. Open Forum Infect Dis 2020 Jun 30:ofaa269. [doi: 10.1093/ofid/ofaa269]
11. COVID-19 United States Cases by County. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/us-map [accessed 2020-05-08]
12. Wang Y, Beydoun MA. The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. Epidemiol Rev 2007;29:6-28. [doi: 10.1093/epirev/mxm007] [Medline: 17510091]
13. Laurencin CT, McClinton A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. J Racial Ethn Health Disparities 2020 Jun;7(3):398-402 [FREE Full text] [doi: 10.1007/s40615-020-00756-0] [Medline: 32306369]
14. Manuel JI. Racial/Ethnic and Gender Disparities in Health Care Use and Access. Health Serv Res 2018 Jun;53(3):1407-1429 [FREE Full text] [doi: 10.1111/1475-6773.12705] [Medline: 28480588]
15. Williams DR, Cooper LA. Reducing Racial Inequities in Health: Using What We Already Know to Take Action. Int J Environ Res Public Health 2019 Feb 19;16(4):606 [FREE Full text] [doi: 10.3390/ijerph16040606] [Medline: 30791452]
16. Jaiswal J, Halkitis PN. Towards a More Inclusive and Dynamic Understanding of Medical Mistrust Informed by Science. Behav Med 2019;45(2):79-85. [doi: 10.1080/08964289.2019.1619511] [Medline: 31343962]

Abbreviations

COVID-19: coronavirus disease
SARS-CoV-2: severe acute respiratory syndrome 2

XSL•FO
RenderX

Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by D Guillaume, S Badawy; comments to author 18.06.20; revised version received 26.06.20; accepted 01.07.20; published 10.07.20.

Please cite as:
Jones J, Sullivan PS, Sanchez TH, Guest JL, Hall EW, Luisi N, Zlotorzynska M, Wilde G, Bradley H, Siegler AJ
Similarities and Differences in COVID-19 Awareness, Concern, and Symptoms by Race and Ethnicity in the United States: Cross-Sectional Survey
J Med Internet Res 2020;22(7):e20001
URL: http://www.jmir.org/2020/7/e20001/
doi:10.2196/20001
PMID:32614778

©Jeb Jones, Patrick S Sullivan, Travis H Sanchez, Jodie L Guest, Eric W Hall, Nicole Luisi, Maria Zlotorzynska, Gretchen Wilde, Heather Bradley, Aaron J Siegler. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study

Yulan Lin[1], MD, MSc, PhD; Zhijian Hu[1], MD, MSc, PhD; Haridah Alias[2], BSc, MSc; Li Ping Wong[1,2], BSc, MSc, PhD

[1]Department of Epidemiology and Health Statistics, Fujian Provincial Key Laboratory of Environment Factors and Cancer, School of Public Health, Fujian Medical University, Fuzhou, China

[2]Centre for Epidemiology and Evidence-Based Practice, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Zhijian Hu, MD, MSc, PhD
Department of Epidemiology and Health Statistics, Fujian Provincial Key Laboratory of Environment Factors and Cancer
School of Public Health
Fujian Medical University
1 Xue Yuan Road
University Town
Fuzhou, 350122
China
Phone: 86 (0)591 2286 ext 2573
Email: huzhijian@fjmu.edu.cn

Abstract

Background: An extensive amount of information related to the novel coronavirus (COVID-19) pandemic was disseminated by mass and social media in China. To date, there is limited evidence on how this infodemic may influence psychobehavioral responses to the crisis.

Objective: The aim of this study is to assess the psychobehavioral responses to the COVID-19 outbreak and examine their associations with mass and social media exposure.

Methods: A cross-sectional study among medical and health sciences students from the Fujian Medical University in Fuzhou, China, was conducted between April 6-22, 2020.

Results: A total of 2086 completed responses were received. Multivariable analyses demonstrated that four constructs of the Health Belief Model (HBM)—higher perception of susceptibility (odds ratio [OR] 1.44; 95% CI 1.07-1.94), severity (OR 1.32; 95% CI 1.10-1.59), self-efficacy (OR 1.61; 95% CI 1.21-2.15), and perceived control or intention to carry out prevention measures (OR 1.32; 95% CI 1.09-1.59)—were significantly associated with a higher mass media exposure score, whereas only three constructs—higher perception of severity (OR 1.43; 95% CI 1.19-1.72), self-efficacy (OR 1.85; 95% CI 1.38-2.48), and perceived control or intention to carry out prevention measures (OR 1.32; 95% CI 1.08-1.58)—were significantly associated with a higher social media exposure score. Lower emotional consequences and barriers to carry out prevention measures were also significantly associated with greater mass and social media exposure. Our findings on anxiety levels revealed that 38.1% (n=795; 95% CI 36.0-40.2) of respondents reported moderate-to-severe anxiety. A lower anxiety level was significantly associated with higher mass and social media exposure in the univariable analyses; however, the associations were not significant in the multivariable analyses.

Conclusions: In essence, both mass and social media are useful means of disseminating health messages and contribute to the betterment of psychobehavioral responses to COVID-19. Our findings stress the importance of the credibility of information shared through mass and social media outlets and viable strategies to counter misinformation during a pandemic.

(J Med Internet Res 2020;22(7):e19982) doi:10.2196/19982

KEYWORDS

psychobehavioral; COVID-19; mass media; social media; medical students; China

XSL•FO
RenderX

Introduction

With rapid increases in the number of internet users, both mass and social media have a prominent role to play in modern society. In China, there were approximately 688 million internet users, of whom 75.1% were aged 10-39 years, in 2015 [1]. As the general public becomes more health conscious, the popularity of social media as a means of acquiring health-related information has been growing in recent years [2,3]. Of note, social media tools are readily accessible on the internet and have become even easier to access via apps on smartphones. As a result, the role of social media as a pathway to news is very popular [4]. However, social media users may be exposed to untrustworthy news or information of questionable accuracy. Inaccurate information acquisition could have detrimental effects, since passive acquisition through social media, particularly through WeChat Moments, is an important medium for health information acquisition among college students in China [2]. (WeChat is the most popular social media platform in China and includes instant messaging and service platforms to carry out payment, marketing, and promotion activities. WeChat Moments is an interactive platform that allows users to share information/news articles, photos, and video.) Moreover, almost 60% of social media users admitted that internet-based health information impacted their health management strategy [5]. Mass media, in contrast, provides more credible information and has been used as a means of communication of scientifically accurate information about health more often than social media. Mass media can influence health behaviors and promote health behavior change in the public [6].

In late December 2019, an unknown form of pneumonia—caused by a novel coronavirus—surfaced in Wuhan, China, and rapidly spread across the globe. By the end of April, the overall number of the coronavirus disease (COVID-19) cases worldwide increased to 2,878,196 and the death count reached 198,668 [7]. China, after over 3 months of battling COVID-19, has managed to control the outbreak. Nonetheless, the community at large in China remains vulnerable, and prevention from rebound is essential since lockdown regulations have been relaxed. During the early phase and the peak of the COVID-19 epidemic in China, various issues surrounding mental distress among the general public caught the attention of researchers. Studies showed that a great proportion of the general public was found to have severe depressive symptoms, even during the early phase of the outbreak [8,9]. It is important to address mental health issues during a disease outbreak, as it may weaken social and other areas of functioning, including an impairment in prevention measures [10,11]. Psychobehavioral responses have been understudied after the cessation of the COVID-19 outbreak in China and this warrants attention. The lay public's psychobehavioral responses during a disease outbreak play an important role in bringing the outbreak under control [10]. Hence, to avoid a resurgence of infections, investigation into preventive behavioral responses in addition to the psychological well-being of the public post–COVID-19 warrant attention. Attitude is a key factor that determines behavioral intention. The Health Belief Model (HBM) has been used as the theoretical framework to explain the health behaviors of individuals. It includes the following concepts: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy [12,13]. Adopting the HBM to explain psychobehavioral changes during the COVID-19 outbreak is essential.

One study in China, conducted during the early phase of the outbreak, found a high prevalence of mental health problems among the public, which was positively associated with frequent exposure to social media [14]. To the best of our knowledge, little research has been conducted on social or mass media exposure now that China has entered the downward trend of COVID-19 transmission. Thus, an investigation of exposure to both mass and social media and linkage to the psychobehavioral health outcomes of the public is needed. Accurate information-seeking behaviors during the COVID-19 outbreak has important implications for health-related behavior change and may strengthen infection prevention and control. The traditional mass media is message-driven; in contrast, social media is conversation-driven, and during the COVID-19 outbreak, it is unclear which form of media influences the public and shapes their psychobehavioral responses. Therefore, this study aimed to (1) assess the level of mass and social media exposure related to COVID-19 and (2) identify the association between both forms of media exposure with HBM constructs, psychological and behavioral responses, and anxiety levels.

Methods

Participants and Study Design

An anonymous internet-based, cross-sectional, open survey was distributed to medical and health sciences students at Fujian Medical University, Fuzhou, China, between April 6-22, 2020. Convenience sampling was used to recruit subjects for this study. The link to the survey questions was sent to administrators or lecturers of all departments to be disseminated to registered students at the university. In an attempt to reach comprehensive recipient coverage, the link to the survey was also sent to students' social media groups and forums. All respondents were informed that their participation was voluntary, and consent was implied through their completion of the questionnaire. No incentives were provided to the study participants.

The questionnaire was developed in English, then translated into Chinese. Local experts performed face validation on the content of the questionnaire. The online questionnaire was subsequently pilot tested for readability and clarity of items on 30 participants from the general public. A minor revision was made based on the results of the pilot study. The revised questionnaire was further pretested before field administration. The survey consisted of questions that assessed demographic background, mass media and social media exposure, constructs from the HBM, psychological and behavioral responses, and anxiety levels associated with the COVID-19 outbreak.

XSL•FO
RenderX

Instruments

Mass Media and Social Media Exposure

Questions on mass media (8 items) and social media (10 items) exposure queried participants about types of information acquisition. The response options were scored on a 4-point Likert scale (0=never, 1=rarely, 2=sometimes, and 3=often). The scores were summed, with higher scores representing higher usage. The possible score range for mass media exposure and social media exposure was 0-24 and 0-30, respectively. The participants were informed that the term "mass media" refers to both traditional and online mass media (written or broadcast), including television, radio, advertising, newspapers, magazines, and newsfeeds. In contrast, "social media" refers to websites and apps such as WeChat, Weibo, and Youku, which are among the most commonly used social media platforms in China. Weibo shares features similar to Twitter (eg, allows users to share content up to a 140-Chinese-character limit). On the other hand, Youku, often called the YouTube of China, is an online video and streaming service platform.

HBM Constructs

Questions related to HBM constructs include perceived severity, perceived susceptibility, perceived efficacy, and perceived control or intention [12,13,15]. Perceived severity was measured using a 1-item question (*How serious do you think COVID-19 is?*) on a 4-point scale (*not at all serious* to *very serious*). Perceived susceptibility was a 1-item question (*What do you think are your chances of getting COVID-19?*) on a 4-point scale (*not at all* to *very large chance*). Perceived efficacy was measured using a 1-item question (*Do you think that you will manage to carry out prevention measures currently recommended by the authorities?*) on a 4-point scale (*certainly cannot* to *most certainly yes*). Perceived control or intention was measured using a 1-item question (*Would you carry out prevention measures currently recommended by the authorities?*) on a 4-point scale (*certainly cannot* to *most certainly yes*).

Psychological and Behavioral Responses

Psychological responses measure the emotional consequences of the COVID-19 outbreak. The emotional consequences consist of questions about feelings of fear, avoidance, keeping a secret, embarrassment, and stigma associated with COVID-19 (5 items). Optional answers were on a 4-point Likert scale, with the items scored as 1 (strongly disagree), 2 (disagree), 3 (agree), or 4 (strongly agree). The possible total emotional consequences score ranged from 5-20, with higher scores representing higher levels of emotional consequences.

Behavioral response measures relating to preventive barriers consist of 5 sections (8 items) that comprise questions about personal protection (3 items), cough etiquette (3 items), and contact precautions (2 items). The question queried participants' level of difficulty in practicing physical prevention measures. A 4-point Likert scale was used to report responses, with scores of 1 (very easy), 2 (easy), 3 (difficult), or 4 (very difficult). The total physical prevention barriers score ranged from 8-32, with higher scores representing higher difficulty levels of physical prevention.

Anxiety

Anxiety was measured using the 6-item state version of the State-Trait Anxiety Inventory (STAI-6) [16,17]. The respondents rated the frequency of experiencing 6 emotional states (ie, being calm, tense, upset, relaxed, content, and worried) as a result of the COVID-19 outbreak. A 4-point scale was used (ie, 1=not at all, 2=somewhat, 3=moderately, and 4=very much). The scores on the 3 positively worded items were reverse-coded. The total summed scores were prorated (multiplied by 20/6) to obtain scores that were comparable with those from the full 20-item STAI (giving a range of 20-80) [17]. A cut-off score of 44 was used to indicate moderate-to-severe symptoms [10,18].

Statistical Analysis

The reliability of the scales used was evaluated by assessing the internal consistency of the items representing the scores. The mass media and social media exposure items had a reliability (Cronbach α) of 0.958 and 0.940, respectively. The emotional consequences and prevention barrier behavior items had a reliability (Cronbach α) of 0.794 and 0.840, respectively. The reliability computed for the STAI-6 items in the assessment of anxiety was 0.793.

Multivariable logistic regression analysis, using a simultaneous forced-entry method, was used to determine the factors influencing mass media and social media exposure. Multivariable logistic regression analyses were performed on all variables found to have a statistically significant association (two-tailed, $P<.05$) in the univariable analyses. Odds ratios (ORs), 95% CIs, and P values were calculated for each independent variable. All statistical analyses were performed using SPSS, version 20.0 (IBM Corporation). The level of significance was set at $P<.05$.

Ethical Considerations

This research was approved by the Research Ethics Committee of the Fujian Medical University. Written informed consent was not acquired from participants. The committee approved that consent was implied through questionnaire completion and submission.

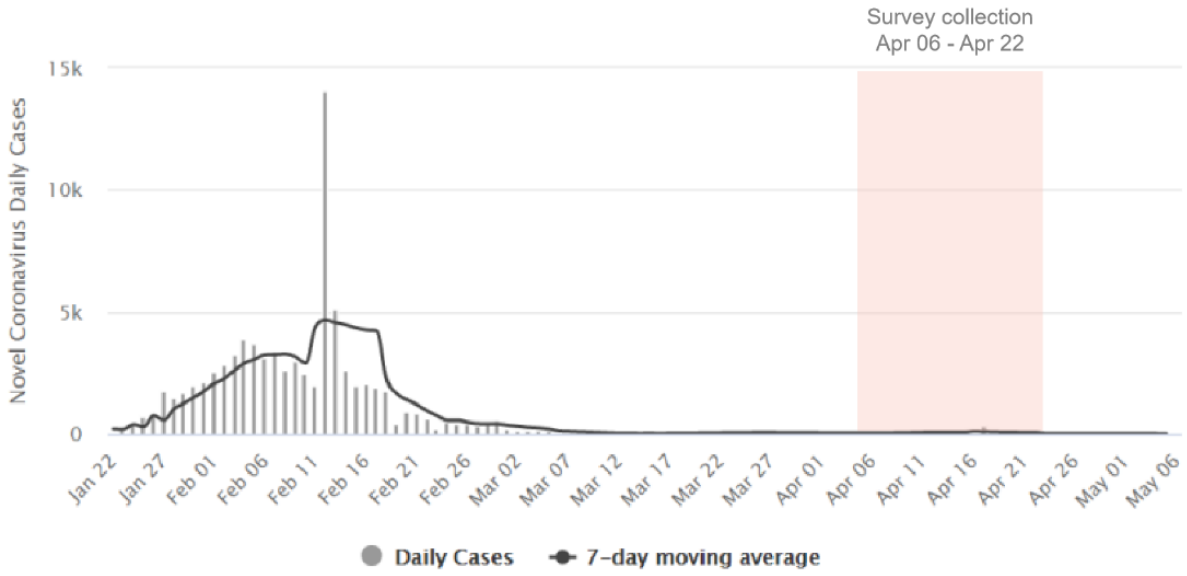
Results

A total of 2086 completed responses were received. Figure 1 shows the number of daily new cases in China since the beginning of the COVID-19 outbreak [19] and the duration of our data collection period. As shown in Figure 1, data collection was carried out past the peak of the COVID-19 outbreak.

As shown in Table 1, more than half of the participants were 18-20 years old (n=1197, 57.4%). Nearly two-thirds of the birthplaces of participants were in rural areas (n=1369, 65.6%). Most participants reported that their annual family income was below CNY 50,000 (n=978, 46.9%) or in the CNY 50,000-120,000 category (n=775, 37.2%). The distribution by university year was approximately equal.

XSL•FO
RenderX

Figure 1. Daily new cases in China since the beginning of the coronavirus disease (COVID-19) outbreak and throughout this study's data collection period.



XSL•FO
RenderX

Table 1. Demographic characteristics of participants (N=2086).

Characteristic	Count, n (%)
Age group (years)	
18-20	1197 (57.4)
21-22	714 (34.2)
23-29	175 (8.4)
Birthplace	
Urban	717 (34.4)
Rural	1369 (65.6)
Annual family income (CNY)	
<50,000	978 (46.9)
50,000-120,000	775 (37.2)
>120,000	333 (16.0)
Year	
1	662 (31.7)
2	490 (23.5)
3	606 (29.1)
4 and postgraduate	328 (15.7)
Health Belief Model	
Perceived susceptibility	
Certainly no/probably no/probably yes	2001 (95.9)
Certainly yes	85 (4.1)
Perceived severity	
Not at all/slightly serious/serious	1101 (52.8)
Very serious	985 (47.2)
Perceived self-efficacy	
Certainly no/probably no/probably yes	580 (27.8)
Certainly yes	1506 (72.2)
Perceived control or intention to carry out preventive measures	
Certainly no/probably no/probably yes	788 (37.8)
Certainly yes	1298 (62.2)
Psychological and behavioral response	
Emotional consequences	
Scores 5-9	1004 (48.1)
Scores 10-20	1082 (51.9)
Barriers to carry out preventive measures	
Scores 8-15	986 (47.3)
Scores 16-32	1100 (52.7)
Anxiety level	
State-Trait Anxiety Inventory	
Scores 20-43	1291 (61.9)
Scores 44-80	795 (38.1)

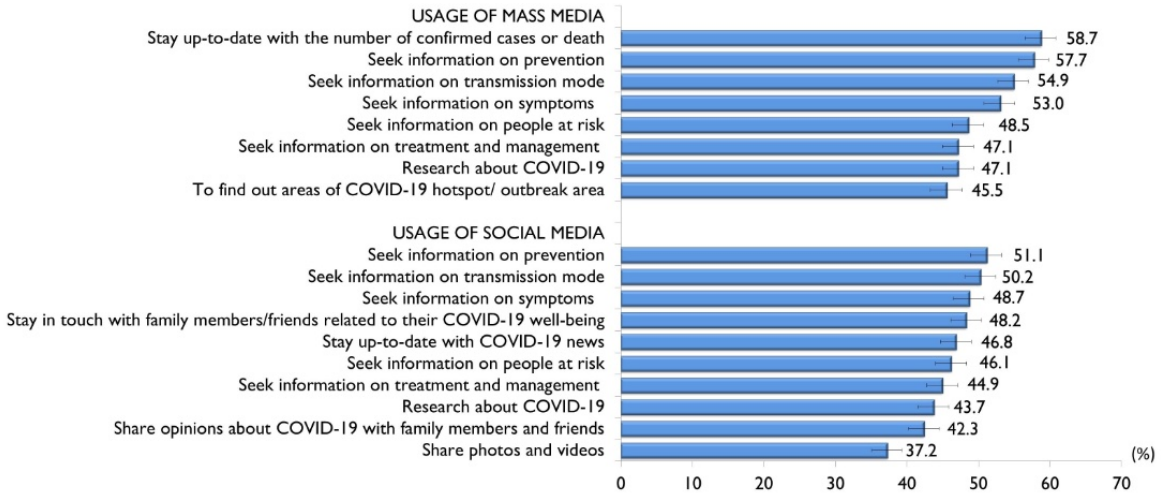
XSL•FO
RenderX

Mass Media and Social Media Exposure

Figure 2 shows the proportion of *often* responses and its corresponding 95% CIs for mass and social media use. The majority of participants relied on mass media for staying up-to-date with information about the number of confirmed COVID-19 cases or deaths (n=1224, 58.7%), followed by information seeking related to prevention (n=1204, 57.7%), transmission (n=1145, 54.9%), symptoms (n=1105, 53%), and risk (n=1012, 48.5%) associated with COVID-19. The most common reasons to use social media were to obtain information about prevention (n=1065, 51.1%), transmission (n=1048, 50.2%), and symptoms (n=1015, 48.7%) of COVID-19.

The mean total mass media exposure was 19.3 (SD 4.9; range 0-24) out of a possible score of 24. The median was 20.0 (IQR 16.0-24.0). The total mass media exposure scores were categorized into two groups (20-24 or 0-23), based on the median split; as such, a total of 1113 (53.5%; 95% CI 51.2-55.5) were categorized as having a score between 20-24 and 973 (46.6%; 95% CI 44.5-48.8) had a score between 0-23. The mean total social media exposure was 23.2 (SD 5.8; range 0-30) out of a possible score of 30. The median was 23.0 (IQR 20.0-29.0). The total social media exposure scores were categorized into two groups (23-30 or 0-22), based on the median split; as such, a total of 1096 (52.5%; 95% CI 50.4-54.7) were categorized as having a score between 23-30 and 990 (47.5%; 95% CI 45.3-49.6) had a score between 0-22.

Figure 2. Proportion of participants who "often" used mass media and social media (N=2086).



HBM Constructs

In total, 1558 participants (74.7%; 95% CI 72.8-76.5) reported *certainly yes*/*probably yes* for perceived susceptibility of getting infected with COVID-19. A relatively lower proportion perceived COVID-19 as very serious (n=985, 47.2%; 95% CI 45.1-49.4). The majority also reported *certainly yes* (n=1506, 72.2%; 95% CI 70.2-74.1) in their ability to carry out recommended prevention measures. A relatively lower proportion reported *certainly yes* (n=1298, 62.2%; 95% CI 60.1-64.3) about their intentions to carry out the recommended prevention measures.

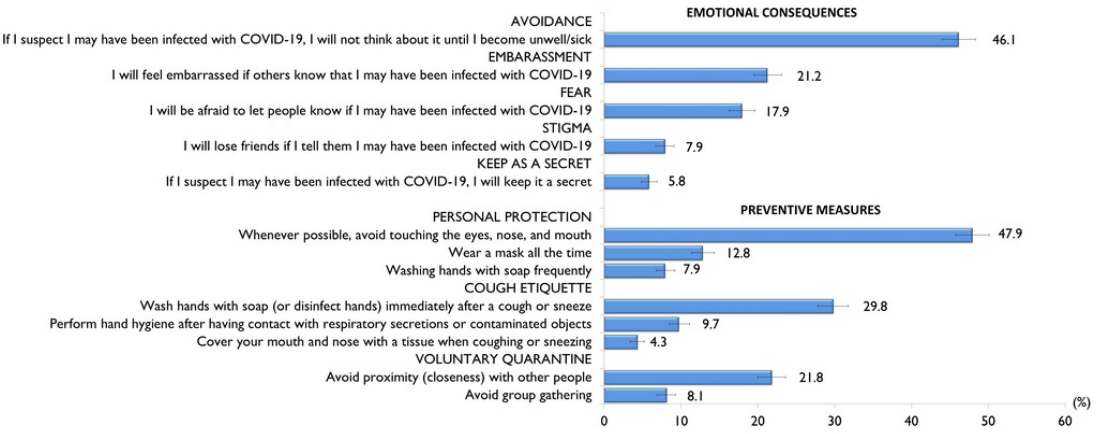
Psychological and Behavioral Responses

Figure 3 shows the proportion and corresponding 95% CIs of responses for items on emotional consequences. Nearly half of the participants answered *strongly agree*/*agree* in regard to avoidance behavior (n=962, 46.1%); 21.2% (n=443) and 17.9% (n=374) strongly agreed or agreed that they felt embarrassment or fear, respectively. The mean total emotional consequences score was 9.4 (SD 2.7; range 5-20). The median was 10 (IQR 7-11). The total emotional consequences scores were categorized into two groups (10-20 or 5-9), based on the median split; as such, a total of 1082 (51.9%; 95% CI 49.7-54.0) were categorized as having a score between 10-20 and 1004 (48.1%; 95% CI 46.0-50.3) were categorized as having a score between 5-9.

The proportions of *difficult/very difficult* responses and the corresponding 95% CIs for difficulties in carrying out preventive measures are also shown in Figure 3. The greatest difficulty reported was avoiding touching one's eyes, nose, and mouth (n=1000, 47.9%). Difficulties in avoiding proximity with other people and wearing a mask all the time were also reported by 21.8% (n=454) and 12.8% (n=267) of participants, respectively. The mean total score for barriers to carry out preventive measure was 15.0 (SD 3.7; range 8-32). The median was 16 (IQR 12-17). The total score for barriers to carry out preventive measures was categorized into two groups (16-32 or 8-15), based on the median split; as such, a total of 1100 (55.7%; 95% CI 50.6-54.9) were categorized as having a score between 16-32, and 986 (47.3%; 95% CI 45.1-49.4) were categorized as having a score between 8-15.

XSL•FO
RenderX

Figure 3. Proportion of participants who answered "agree/strongly agree" for questions related to emotional consequences and "difficult/very difficult" for questions related to carrying out preventive measures (N=2086).



Anxiety

The mean overall anxiety score was 40.4 (SD 10.8; range 20-80). Using a cut-off score of 44 for the STAI score, a total of 38.1% (n=795) (95% CI 36.0-40.2) of participants reported moderate-to-severe anxiety (score=44-80). Participants in the 18-20 years age group (n=477, 39.8%) reported the highest amount of moderate-to-severe anxiety, followed those who were 21-22 years old (n=270, 37.8%) and 23-29 years old (n=48, 27.4%) (χ^2_2=10.027, *P*=.007). There was a gradual decrease in the proportion of moderate-to-severe anxiety by university year, whereby 41.4% (n=274) of year 1 participants reported moderate-to-severe anxiety compared to 40.0% (n=196) among year 2, 39.8% (n=241) among year 3, and only 25.6% (n=84) among year 4 (χ^2_3= 26.198, *P*<.001).

Influence of Mass and Social Media on Psychobehavioral Responses

As shown in Table 2, multivariable regression analysis of factors influencing a higher score of mass media exposure showed significant associations with all the HBM constructs. Higher perception of severity (OR 1.33; 95% CI 1.10-1.60), self-efficacy (OR 2.03; 95% CI 1.64-2.52), and perceived control or intention to carry out prevention measures (OR 1.29; 95% CI 1.07-1.56) were significantly associated with a higher mass media exposure score. Lower emotional consequences (OR 1.51; 95% CI 1.25-1.83) and barriers to carry out preventive measures (OR 1.50; 95% CI 1.26-1.84) were also significantly associated with a higher mass media exposure score.

Multivariable regression analysis of factors influencing a higher score of social media exposure showed significant associations with 3 of the HBM constructs. Higher perception of severity (OR 1.41; 95% CI 1.17-1.69), self-efficacy (OR 2.01; 95% CI 1.67-2.58), and perceived control or intention to carry out prevention measures (OR 1.27; 95% CI 1.05-1.53) were significantly associated with a higher social media exposure score. Likewise, lower emotional consequences (OR 1.50; 95% CI 1.24-1.67) and barriers to carry out preventive measures (OR 1.39; 95% CI 1.15-1.67) were also significantly associated with a higher social media exposure score.

A lower anxiety score was significantly associated with higher mass and social media exposure in the univariable analyses; however, the associations were not significant in the multivariable analyses.

XSL•FO
RenderX

Table 2. Factors associated with mass media and social media exposure (N=2086).

Variable	Univariate analysis (mass media exposure score 20-24 vs 0-19[a])		Multivariable logistic regression (mass media exposure score 20-24 vs 0-19[a])	Univariate analysis (social media exposure score 23-30 vs 0-22[b])		Multivariable logistic regression (social media exposure score 23-30 vs 0-22[b])
	High score (20-24) (n=1113)	*P* value	OR[c] (95%CI)	High score (23-30) (n=1096)	*P* value	OR (95%CI)
Demographic characteristics						
Age group (years)						
18-20	615 (51.4)		—[d]	625 (52.2)		—
21-22	397 (55.6)	.10	—	377 (52.8)	.92	—
23-29	101 (57.7)		—	94 (53.7)		—
Birthplace						
Urban	402 (56.1)	.08	—	395 (55.1)	.09	—
Rural	711 (51.9)		—	701 (51.2)		—
Annual family income (CNY)						
<50,000	503 (51.4)		—	497 (50.8)		Ref[e]
50,000-120,000	421 (54.3)	.19	—	404 (52.1)	.048	1.01 (0.83-1.24)
>120,000	189 (56.8)		—	195 (58.6)		1.28 (0.98-1.66)
Year						
1	338 (51.1)		Ref	338 (51.1)		—
2	234 (47.8)		0.94 (0.73-1.20)	239 (48.8)	.07	—
3	345 (56.9)	.001	1.30 (1.03-1.64)[f]	333 (55.0)		—
4 and postgraduate	196 (59.8)		1.31 (0.99-1.74)	186 (56.7)		—
Health Belief Model						
Perceived susceptibility						
Certainly no/probably no/probably yes	1056 (52.8)		Ref	1033 (51.6)		Ref
Certainly yes	57 (67.1)	.01	1.1 (0.75-1.96)	63 (74.1)	<.001	1.75 (1.05-2.93)
Perceived severity						
Not at all/slightly serious/serious	524 (47.6)	<.001	Ref	509 (46.2)	<.001	Ref
Very serious	589 (59.8)		1.33 (1.10-1.60)[g]	587 (59.6)		1.41 (1.17-1.69)[h]
Perceived self-efficacy						
Certainly no/probably no/probably yes	205 (35.3)	<.001	Ref	199 (34.3)	<.001	Ref
Certainly yes	908 (60.3)		2.03 (1.64-2.52)[h]	897 (59.6)		2.01 (1.67-2.58)[h]
Perceived control or intention to carry out preventive measures						
Certainly no/probably no/probably yes	293 (49.9)	.01	Ref	386 (49.0)	.01	Ref
Certainly yes	720 (55.5)		1.29 (1.07-1.56)[g]	710 (54.7)		1.27 (1.05-1.53)[f]
Psychological and behavioral response						
Emotional consequences						

XSL•FO
RenderX

Variable	Univariate analysis (mass media exposure score 20-24 vs 0-19[a])		Multivariable logistic regression (mass media exposure score 20-24 vs 0-19[a])	Univariate analysis (social media exposure score 23-30 vs 0-22[b])		Multivariable logistic regression (social media exposure score 23-30 vs 0-22[b])
	High score (20-24) (n=1113)	*P* value	OR[c] (95%CI)	High score (23-30) (n=1096)	*P* value	OR (95%CI)
Scores 5-9	625 (62.3)	<.001	1.51 (1.25-1.83)[h]	607 (60.5)	<.001	1.50 (1.24-1.67)[h]
Scores 10-20	488 (45.1)		Ref	489 (45.2)		Ref
Barriers to carry out preventive measure						
Scores 8-15	618 (62.7)	<.001	1.5 (1.26-1.84)[h]	608 (61.7)	<.001	1.39 (1.15-1.67)[g]
Scores 16-32	495 (45.0)		Ref	488 (44.4)		Ref
Anxiety level						
State-Trait Anxiety Inventory						
Score 20-43	731 (56.6)	<.001	1.16 (0.96-1.40)	713 (55.2)	.002	1.11 (0.90-1.34)
Score 44-80	382 (48.1)		Ref	383 (48.2)		Ref

[a]Hosmer–Lemeshow test, chi-square: 4.749, *P*=.78; Nagelkerke R^2=0.116.

[b]Hosmer–Lemeshow test, chi-square: 16.804, *P*=.03; Nagelkerke R^2=0.112.

[c]OR: odds ratio.

[d]Not applicable.

[e]Ref: reference.

[f]*P*<.05.

[g]*P*<.01.

[h]*P*<.001.

Discussion

Principal Findings

This study assessed both mass and social media exposure related to COVID-19 and investigated the association between media exposure and HBM constructs, psychological and behavioral responses, and anxiety levels. This study targeted university students, as university students are among the biggest users of the internet and social media [20]. Since this study was conducted when the country was experiencing a decline in COVID-19 cases, it has the advantage of identifying detrimental psychobehavioral factors to provide insight for interventions to prevent a resurgence of infections. Of note, during the data collection period, the nationwide lockdown and movement control had started to ease; nevertheless, schools and universities in China had not yet reopened.

The high mean total exposure score implies that university students have high exposure to both mass media and social media during the COVID-19 outbreak. This finding replicates evidence from previous research, indicating high use of online media (particularly social media) by the younger generation and specifically university students [21-24]. In this study, we also found that university students were exposed to equal amounts of COVID-19–related information from both mass and social media. Both mass and social media were equally used as information sources for the prevention of infection, symptoms, risk, and mode of transmission.

Despite China's downward COVID-19 trend, the study participants demonstrated a high perceived risk of COVID-19 infection. However, a relatively lower perception of the severity of COVID-19 infection was observed. Many participants also reported high self-efficacy in carrying out recommended prevention measures. During the early phase of the outbreak, the country carried out aggressive public health interventions, such as early detection of cases, contact tracing, and population behavioral changes, which have been reported to have contributed enormously to containing the epidemic [25]. The positive psychobehavioral responses found in this study indicate that population behavioral change interventions have brought about positive behavioral as well as attitudinal changes up to the present time, which is reflected in the success in curbing the spread of the virus to the wider community as observed in the continuous slowdown of COVID-19 cases in China.

The study also found an overall low level of emotional consequences among participants during the off-peak period of the COVID-19 outbreak, as shown by the low mean value of the total emotional consequences score. Despite the low level of emotional consequences, it should be noted that continuous mitigation of the emotional well-being of the public during an infectious disease outbreak is important in controlling

XSL•FO
RenderX

transmission [26]. During the severe acute respiratory syndrome (SARS) epidemic, fear and stigma may have instigated people to delay seeking care and remain in the community undetected. Also, a noteworthy finding is that the most prominent emotional consequence found was avoidance behavior, as it was reported by nearly half of the study participants. It is important to note that cognitive avoidance contributes to a delay in taking precautions to prevent the spread of COVID-19. The implication of this is that prompt action by the public and immediate seeking of medical care upon suspicion of COVID-19 infection are still needed, regardless of the downward trend.

This study's participants found minimal difficulty in carrying out preventive measures. The most prominent difficulty encountered was avoiding touching one's eyes, nose, and mouth; nearly half the participants reported having experienced this difficulty. The importance of refraining from touching one's eyes, nose, and mouth with unwashed hands to prevent the transmission of COVID-19 has been noted previously [27]. Since habitual face-touching behavior has been commonly reported [28], hand hygiene compliance should be encouraged to avoid this route of transmission. Public health interventions to promote and encourage desirable hand-hygiene–compliant behaviors are crucial even though the outbreak is largely under control.

During the early phase of the pandemic, more than half (53.8%) of the general public in China reported the psychological impact associated with COVID-19 as being moderate or severe [8]. In this study, slightly over one-third (38.1%) of university students reported moderate-to-severe anxiety. Although relatively lower anxiety levels were observed after the peak of the outbreak, our results indicate that COVID-19 is still spurring fear in some parts of society. In the case of the Ebola outbreak, anxiety and depression were still prevalent 1 year after the outbreak, especially among those who had been in quarantine and witnessed death associated with the disease [29]. Findings from this study imply that COVID-19–related anxiety among university students warrants special attention. Therefore, it is suggested that continuous assessment and monitoring of COVID-19–associated mental health issues is essential when students resume their studies on campus. Mental health service provision or psychological intervention services to help students who experience loss of family members or friends to COVID-19 should be encouraged in all universities across the country, especially in Wuhan, China's coronavirus epicentre. Furthermore, the study also found that younger university students were more vulnerable to moderate-to-severe anxiety; more attention from university authorities should be allocated to monitor the mental well-being of these students.

The results of the multivariate analyses of this study provide evidence of the important role of both mass and social media in shaping individual health beliefs using the HBM constructs. Substantial mass media exposure was associated with having a higher perception of illness severity and a higher perceived control or intention to carry out prevention measures. Similarly, social media exposure shapes individual health beliefs using the HBM constructs. However, high social media exposure was associated with all the HBM constructs investigated, except for the perception of risk.

Previous reports have noted that emotional consequences such as fear, stigma, and discrimination during the COVID-19 outbreak among people in China were fuelled by misinformation and unfounded rumors [30]. In our study, multivariate analyses revealed that greater mass and social media exposure were also associated with lower emotional consequences, namely, perception of avoidance, embarrassment, fear, and keeping the infection a secret. This perhaps implies that our study participants were exposed to credible and accurate information from both mass and social media, and hence were not negatively impacted. Of note, the Chinese government implemented viable strategies to counter misinformation and fake news during the pandemic such as immediate removal of fake news in the media and strict penalties for offenders.

The behavioral influence of both mass and social media were evident in this study. More mass and social media exposure was also associated with lower barriers to carrying out prevention practices. The findings imply the importance of continuously providing the public with accurate and credible information through mass and social media to enhance emotional well-being and prevention behaviors. It is also vital for media authorities to ensure the credibility of information shared in during an infectious pandemic to elevate negative psychological impact and enhance prevention behaviors. It has been suggested that quick and targeted interventions oriented to delegitimize sources of fake information in the media are important to reduce negative consequences [31]. As such, the findings of this study provide insights into the importance of developing prompt strategies to counter misinformation.

In short, our findings suggest that both mass and social media are useful means of getting health messages across and contribute to improving psychobehavioral responses to COVID-19. Although traditionally the trustworthiness and authenticity of information sourced from social media in relation to mass media has been an issue of concern, this study demonstrated contrary results. Both mass and social media contributed similarly to favorable psychobehavioral responses to COVID-19.

Interestingly, the univariable analyses also observed that both high levels of mass and social media usage were significantly associated with lower anxiety levels. However, the association was not significant in the multivariable analyses. Our finding contradicts recent findings that reported a high prevalence of mental health problems among the public in China that was positively associated with frequent exposure to social media [14]. Of note, our study participants were medical and health sciences students, and this perhaps implies that they were more proficient at identifying and consuming credible information on social media than the general public. In addition, our study has also demonstrated that students with higher exposure to mass and social media tend to have lower negative emotional consequences and fewer barriers to carrying out prevention measures, which might partly contribute to their lower anxiety level. Thus, their increased social media usage does not result in a higher level of mental health problems. This possibly suggests that the proper use of social media for information purposes is beneficial is shaping psychological and behavioral responses during an infectious disease outbreak.

XSL•FO
RenderX

Limitations

This study has several limitations that should be considered. The first pertains to the use of convenience sampling and its cross-sectional nature. It cannot, therefore, be used to infer causality. Despite of the recruitment of a large and diverse sample, the relatively high proportion of young participants in this study may introduce a bias toward greater social media usage. Second, the responses were based on self-report and may be subject to recall bias, self-reporting bias, and a tendency to report socially desirable responses. A third limitation is that the participants were medical and health sciences students; this warrants careful interpretation owing to their comprehensive knowledge and attitude about COVID-19 as well as their higher affinity for health information. Next, the associations found in this study should be interpreted with caution as the psychobehavioral responses were obtained during the off-peak period of the COVID-19 outbreak. Despite these limitations, the study data contribute tremendously to the understanding of the influence of both mass and social media on psychobehavioral responses to the COVID-19 outbreak in China.

Conclusions

Higher exposure to both mass and social media related to the COVID-19 outbreak increased positive attitudes in all the domains of the HBM. Emotional consequences and behavioral prevention barriers also reduced with higher exposure to both mass and social media. In conclusion, based on our results, both mass and social media are useful means of disseminating health-related information to the public and contribute to improvements in psychobehavioral responses to COVID-19. Our findings imply that university students are proficient at identifying and consuming credible information on social media. With much information circulating on the internet, it is challenging for the public to stay informed with reliable, credible, and trustworthy information from the internet. The general public should be informed about proper online health information seeking during disease outbreaks to avoid detrimental psychological and behavioral impacts that may deter outbreak management and control.

Acknowledgments

This study was supported by The Pilot Project of Fujian Provincial Department of Science & Technology (No. 2020Y0005), and Fujian Medical University's Rapid Response Funding Call for COVID-19 Related Research (No. 2020YJ003).

Authors' Contributions

LPW, YL, and ZH conceived the study. YL collected data. LPW and HA analyzed the data. LPW wrote the manuscript. All authors have approved the manuscript.

Conflicts of Interest

None declared.

References

1. Statistical Report on Internet Development in China. China Internet Network Information Center. 2016 Jan. URL: https://cnnic.com.cn/IDR/ReportDownloads/201604/P020160419390562421055.pdf [accessed 2020-06-29]
2. Di Z, Junsheng G, Ruosi S. Clustering analysis of health information acquisition channels: active acquisition and passive receiving. Chinese Journal of Journalism & Communication 2015;05:81-93. [doi: 10.13495/j.cnki.cjjc.2015.05.006]
3. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
4. von Nordheim G, Boczek K, Koppers L. Sourcing the Sources. Digital Journalism 2018 Sep 20;6(7):807-828. [doi: 10.1080/21670811.2018.1490658]
5. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
6. Fishman JM, Casarett D. Mass media and medicine: when the most trusted media mislead. Mayo Clin Proc 2006 Mar;81(3):291-293. [doi: 10.4065/81.3.291] [Medline: 16529129]
7. Coronavirus disease 2019 (COVID-19) Situation Report-98. World Health Organization. 2020 Apr 27. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200427-sitrep-98-covid-19.pdf?sfvrsn=90323472_4 [accessed 2020-04-30]
8. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
9. Zhang J, Lu H, Zeng H, Zhang S, Du Q, Jiang T, et al. The differential psychological distress of populations affected by the COVID-19 pandemic. Brain Behav Immun 2020 Jul;87:49-50 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.031] [Medline: 32304883]

XSL•FO
RenderX

10. Leung GM, Ho LM, Chan SKK, Ho SY, Bacon-Shone J, Choy RYL, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis 2005 Jun 15;40(12):1713-1720. [doi: 10.1086/429923] [Medline: 15909256]
11. Taylor MR, Agho KE, Stevens GJ, Raphael B. Factors influencing psychological distress during a disease epidemic: data from Australia's first outbreak of equine influenza. BMC Public Health 2008 Oct 03;8:347 [FREE Full text] [doi: 10.1186/1471-2458-8-347] [Medline: 18831770]
12. Rosenstock I. The Health Belief Model and Preventive Health Behavior. Health Education Monographs 1974 Dec;2(4):354-386. [doi: 10.1177/109019817400200405]
13. Rosenstock IM. Historical Origins of the Health Belief Model. Health Education Monographs 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
14. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
15. Champion V, Skinner C. The Health Belief Model. In: Glanz K, Rimer B, Viswanath K, editors. Health behavior and health education. San Francisco, CA: Jossey-Bass; 2008:4.
16. Hou WK, Hall BJ, Canetti D, Lau KM, Ng SM, Hobfoll SE. Threat to democracy: Physical and mental health impact of democracy movement in Hong Kong. J Affect Disord 2015 Nov 01;186:74-82 [FREE Full text] [doi: 10.1016/j.jad.2015.07.005] [Medline: 26232750]
17. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
18. Knight RG, Waal-Manning HJ, Spears GF. Some norms and reliability data for the State--Trait Anxiety Inventory and the Zung Self-Rating Depression scale. Br J Clin Psychol 1983 Nov;22 (Pt 4):245-249. [doi: 10.1111/j.2044-8260.1983.tb00610.x] [Medline: 6640176]
19. Daily New Cases in China. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/country/china/ [accessed 2020-05-07]
20. Zachos G, Paraskevopoulou-Kollia E, Anagnostopoulos I. Social Media Use in Higher Education: A Review. Education Sciences 2018 Nov 05;8(4):194. [doi: 10.3390/educsci8040194]
21. Kitazawa M, Yoshimura M, Hitokoto H, Sato-Fujimoto Y, Murata M, Negishi K, et al. Survey of the effects of internet usage on the happiness of Japanese university students. Health Qual Life Outcomes 2019 Oct 11;17(1):151 [FREE Full text] [doi: 10.1186/s12955-019-1227-5] [Medline: 31604455]
22. Apuke OD, Iyendo TO. University students' usage of the internet resources for research and learning: forms of access and perceptions of utility. Heliyon 2018 Dec;4(12):e01052 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e01052] [Medline: 30582057]
23. Almarabeh T, Majdalawi YK, Mohammad H. Internet Usage, Challenges, and Attitudes among University Students: Case Study of the University of Jordan. JSEA 2016;09(12):577-587. [doi: 10.4236/jsea.2016.912039]
24. Christakis D, Moreno M, Jelenchick L, Myaing M, Zhou C. Problematic internet usage in US college students: a pilot study. BMC Med 2011 Jun 22;9:77 [FREE Full text] [doi: 10.1186/1741-7015-9-77] [Medline: 21696582]
25. Lancet. Sustaining containment of COVID-19 in China. Lancet 2020 Apr 18;395(10232):1230 [FREE Full text] [doi: 10.1016/S0140-6736(20)30864-3] [Medline: 32305082]
26. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Inectious Diseases/SARS Community Outreach Team. Fear and stigma: the epidemic within the SARS outbreak. Emerg Infect Dis 2004 Feb;10(2):358-363 [FREE Full text] [doi: 10.3201/eid1002.030750] [Medline: 15030713]
27. Adhikari SP, Meng S, Wu Y, Mao Y, Ye R, Wang Q, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infect Dis Poverty 2020 Mar 17;9(1):29 [FREE Full text] [doi: 10.1186/s40249-020-00646-x] [Medline: 32183901]
28. Kwok YLA, Gralton J, McLaws M. Face touching: a frequent habit that has implications for hand hygiene. Am J Infect Control 2015 Feb;43(2):112-114 [FREE Full text] [doi: 10.1016/j.ajic.2014.10.015] [Medline: 25637115]
29. Jalloh MF, Li W, Bunnell RE, Ethier KA, O'Leary A, Hageman KM, et al. Impact of Ebola experiences and risk perceptions on mental health in Sierra Leone, July 2015. BMJ Glob Health 2018;3(2):e000471 [FREE Full text] [doi: 10.1136/bmjgh-2017-000471] [Medline: 29607096]
30. Ren SY, Gao RD, Chen YL. Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic. World J Clin Cases 2020 Feb 26;8(4):652-657 [FREE Full text] [doi: 10.12998/wjcc.v8.i4.652] [Medline: 32149049]
31. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]

Abbreviations

COVID-19: coronavirus disease

XSL•FO
RenderX

HBM: Health Belief Model
OR: odds ratio
SARS: severe acute respiratory syndrome
STAI: State-Trait Anxiety Inventory

Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by MA Bahrami, W Ahmed, E Chiauzzi; comments to author 18.06.20; revised version received 23.06.20; accepted 25.06.20; published 20.07.20.

Please cite as:
Lin Y, Hu Z, Alias H, Wong LP
Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study
J Med Internet Res 2020;22(7):e19982
URL: https://www.jmir.org/2020/7/e19982
doi:10.2196/19982
PMID:32584779

©Yulan Lin, Zhijian Hu, Haridah Alias, Li Ping Wong. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of National Containment Measures on Decelerating the Increase in Daily New Cases of COVID-19 in 54 Countries and 4 Epicenters of the Pandemic: Comparative Observational Study

Carlos K H Wong[1*], BSc, MPhil, PhD; Janet Y H Wong[2*], PhD; Eric H M Tang[1], BSc; Chi Ho Au[1], BSc; Kristy T K Lau[1], BSc, MSc; Abraham K C Wai[3], MBChB, JD, MBA, FRCPGlasg, FRCEM, FHKAM

[1]Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
[2]School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
[3]Emergency Medicine Unit, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
[*]these authors contributed equally

Corresponding Author:
Abraham K C Wai, MBChB, JD, MBA, FRCPGlasg, FRCEM, FHKAM
Emergency Medicine Unit
Li Ka Shing Faculty of Medicine
The University of Hong Kong
Room 514, William MW Mong Block, Faculty of Medicine Building
21 Sassoon Road, Pokfulam
Hong Kong,
China (Hong Kong)
Phone: 852 3917 9859
Fax: 852 2816 2293
Email: awai@hku.hk

Abstract

Background: Coronavirus disease (COVID-19) is a worldwide epidemic, and various countries have responded with different containment measures to reduce disease transmission, including stay-at-home orders, curfews, and lockdowns. Comparative studies have not yet been conducted to investigate the impact of these containment measures; these studies are needed to facilitate public health policy-making across countries.

Objective: The aim of this study was to describe and evaluate the impact of national containment measures and policies (stay-at-home orders, curfews, and lockdowns) on decelerating the increase in daily new cases of COVID-19 in 54 countries and 4 epicenters of the pandemic in different jurisdictions worldwide.

Methods: We reviewed the effective dates of the national containment measures (stay-at-home order, curfew, or lockdown) of 54 countries and 4 epicenters of the COVID-19 pandemic (Wuhan, New York State, Lombardy, and Madrid), and we searched cumulative numbers of confirmed COVID-19 cases and daily new cases provided by health authorities. Data were drawn from an open, crowdsourced, daily-updated COVID-19 data set provided by Our World in Data. We examined the trends in the percent increase in daily new cases from 7 days before to 30 days after the dates on which containment measures went into effect by continent, World Bank income classification, type of containment measures, effective date of containment measures, and number of confirmed cases on the effective date of the containment measures.

Results: We included 122,366 patients with confirmed COVID-19 infection from 54 countries and 24,071 patients from 4 epicenters on the effective dates on which stay-at-home orders, curfews, or lockdowns were implemented between January 23 and April 11, 2020. Stay-at-home, curfew, and lockdown measures commonly commenced in countries with approximately 30%, 20%, or 10% increases in daily new cases. All three measures were found to lower the percent increase in daily new cases to <5 within one month. Among the countries studied, 20% had an average percent increase in daily new cases of 30-49 over the seven days prior to the commencement of containment measures; the percent increase in daily new cases in these countries was curbed to 10 and 5 a maximum of 15 days and 23 days after the implementation of containment measures, respectively.

Conclusions: Different national containment measures were associated with a decrease in daily new cases of confirmed COVID-19 infection. Stay-at-home orders, curfews, and lockdowns curbed the percent increase in daily new cases to <5 within a month. Resurgence in cases within one month was observed in some South American countries.

XSL•FO
RenderX

(J Med Internet Res 2020;22(7):e19904) doi:10.2196/19904

KEYWORDS

COVID-19; national containment; lockdown; curfew; stay-at-home; epidemic curve

Introduction

The coronavirus disease (COVID-19) epidemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); it is the third coronavirus-associated epidemic to emerge from a species leap from wild animals to humans, after severe acute respiratory syndrome (SARS) in 2003 and Middle East respiratory syndrome (MERS) in 2012 [1,2]. Coronavirus infection causes acute respiratory illness that is usually self-limiting but can be severe in some cases [3]. Coronaviruses primarily infect the upper respiratory and gastrointestinal tracts of birds and mammals. However, once a human is infected, they can transmit the coronavirus to other humans through respiratory droplets and aerosols from coughing and sneezing, like other respiratory pathogens [4]. In Wuhan, China, it was estimated that the basic reproductive number for SARS-CoV-2 was 2.68 (95% CI 2.47-2.86) and its doubling time was 6.4 days (95% CI 5.8-7.1) [5] in the early phase of the pandemic. Since January 2020, following the lockdown of Wuhan, increasing numbers of SARS-CoV-2–infected cases have been reported outside the city [6]. The COVID-19 pandemic began with small chains of transmission in China and nearby cities, which then became large chains of extensive spread in countries worldwide. As of July 2, 2020, over 10 million confirmed cases and more than 510,000 deaths have been recorded globally [7]. National responses of containment measures, such as stay-at-home orders, curfews, and lockdowns, have varied across countries with different characteristics. To our knowledge, there is no comparative study investigating the impact of these containment measures on COVID-19 transmission in countries with respect to geographical location, income status, containment measures imposed, and early and late responses to the COVID-19 pandemic.

From the perspective of public health, principles of controlling contagious disease transmission focus mainly on early detection via testing and contact tracing, in addition to prevention of transmission by containment measures such as stay-at-home orders, curfews, lockdowns, quarantine of exposed individuals [8], and travel or trade restrictions [9]. Additionally, behavioral interventions such as personal protective measures (such as hand hygiene and respiratory etiquette) and social or physical distancing measures (such as isolation of sick individuals, school measures and closures, workplace measures and closures, and avoiding crowding) should be advocated at the community level to facilitate the flattening of the epidemic curve [10].

Quarantine has been used as an effective tool to control communicable disease outbreaks throughout history [11]. The quarantine period provides ample time for the incubation period to complete; therefore, asymptomatic cases will become symptomatic and can therefore be identified. Quarantine is most successful in settings where there is prompt detection of new cases, where contacts can be listed and traced within a short time frame, followed by a prompt issuance of quarantine with voluntary compliance. This may not be applicable to the case of COVID-19 infection because the science and epidemiology of the disease are still largely unknown. At the time of preparation of this paper, in a rapid review assessing the effects of quarantine (alone or in combination with other measures) of individuals who had contact with confirmed cases of COVID-19 infection, who travelled from countries with a declared outbreak, or who lived in regions with high transmission of the disease, it was concluded that insufficient evidence was available [12].

Initial analysis of data from China collected in February 2020 suggested that the epidemic did not expand exponentially. Public response to the epidemic in addition to containment policies was found to be effective despite the initial increase in the number of confirmed cases [13]. The initial transmission rate (R_0) was reduced from 2.6 to 0.62 (95% CI 0.37-0.89) due to the stay-at-home order in the United Kingdom [14]. Similar changes in effective transmission rate (R_t) were observed in the United States and many European countries after population-level containment interventions were implemented [15,16]. However, worldwide evidence to guide policy makers on effective control of the pandemic is still lacking.

Nonpharmaceutical public health measures at the individual level (physical distancing, use of face masks, and wearing of eye protection to reduce person-to-person transmission) have been studied by the COVID-19 Systematic Urgent Review Group Effort (SURGE) [17]. However, the effectiveness of measures implemented at the national level is a knowledge gap that should be filled. In this study, we aimed to describe and evaluate the impact of national containment measures and policies (stay-at-home orders, curfews, and lockdowns) on decelerating the increase in daily new cases of COVID-19 infection in 54 countries and 4 epicenters of the pandemic in different jurisdictions worldwide.

Methods

Data Source

We used an open, crowdsourced, daily-updated COVID-19 data set provided by Our World in Data [18]. This public domain repository provides numbers of cumulative confirmed cases, confirmed cases per 1 million people, confirmed daily new cases, cumulative deaths, and daily deaths for each country associated with the European Centre for Disease Prevention and Control and the total number of tests for COVID-19 performed per 1000 people as reported by the health authority of each country. Because the most common test for COVID-19 is the polymerase chain reaction (PCR) test, our data source tracked the number of tests for COVID-19 in terms of the number of PCR tests performed or the number of individuals who were tested for COVID-19 reported by the health authority of each country. Cumulative confirmed cases and daily new cases in four epicenters (Wuhan, the Lombardy region of Italy, New York State, and Madrid) were obtained from the websites

XSL•FO
RenderX

of national health authorities. Commencement dates of countrywide containment measures were gathered from the national health ministries and health authorities by authors CW and EHMT and were cross-checked by the last author (AW). Characteristics including the continent, income level classified by the World Bank [19], types of containment measures (stay-at-home order, curfew, or lockdown), effective dates of containment measures, and numbers of confirmed cases on the effective dates of containment measures were retrieved for each country and epicenter. All crowdsourced data were available up to June 20, 2020.

Regarding the three containment measures under investigation, a stay-at-home order was defined as limited outdoor movement except essential activity, curfew was defined as a stay-at-home order during specific time periods, and lockdown was defined as restriction of population mobility within a specific region or country. Acknowledging that variations in the adoption of containment measures may exist across regions or countries, the main difference between a stay-at-home order and a lockdown was denoted as the restriction of people moving in or out of a region being imposed with the latter but not the former. Accordingly, the containment measure introduced in the United States would be classified as a stay-at-home order.

Ethics Approval

This study used open-sourced, secondary dataset, and was exempted from ethics review of Institution Review Board.

Outcome Definitions

The primary outcome of this study was the percent increase in daily new cases from 7 days prior (day –7) to 30 days after (day +30) the commencement of measures in countries and epicenters. This specific timeframe between day –7 and day +30 ensured that complete daily new case data were obtained for all included countries and epicenters. The percent increase in daily new cases on day t was defined as the daily new cases on day t divided by the cumulative confirmed cases on day $t-1$, where t was between –7 and 30. When estimating the percent increase in daily new cases by characteristic group, estimates were calculated as the total daily new cases in the locations within the group on day t divided by the total cumulative confirmed cases of the locations within the group on day $t-1$. The denominator and numerator in the fraction were two summations; the summations were taken over the countries with nonmissing numbers of cases on days t and $t-1$, respectively. This percent increase can be represented by the weighted sum of the percent increase of the component locations, where the weight is calculated based on the cumulative cases in the country on day $t-1$:



Average increase in daily new cases from day-7 to day-1 was calculated by method of geometric mean.

Data Analysis

We used line graphs to represent the trends in the percent increase in daily new cases for each type of containment measure and by the measure start date, continent of the country or epicenter, income level, and average percent increase in daily new cases before the start of containment measures. The formula proposed by Waller et al [20] was used to estimate the 95% CI for the weighted proportion. The 95% CI for the percent increase of a location was estimated by the method of the binomial proportion confidence interval, while that for a group was estimated by the method of the weighted binomial proportion confidence interval. The calculation of the percent increase in daily new cases at time t was excluded in a particular group and country if the daily new cases data were missing at time $t-1$ and time t. Scatter plots were used to visualize the percent increase in daily new cases at day +7, day +14, day +21 and day +30 for each country or epicenter by the average percent increase in cases before the commencement of the containment measures and after the start dates of the containment measures. The relationship between the number of diagnosed cases and the number of tests performed for each location was also assessed using line graphs.

All statistical analyses and figure generations were performed using Stata version 16.0 (StataCorp LLC).

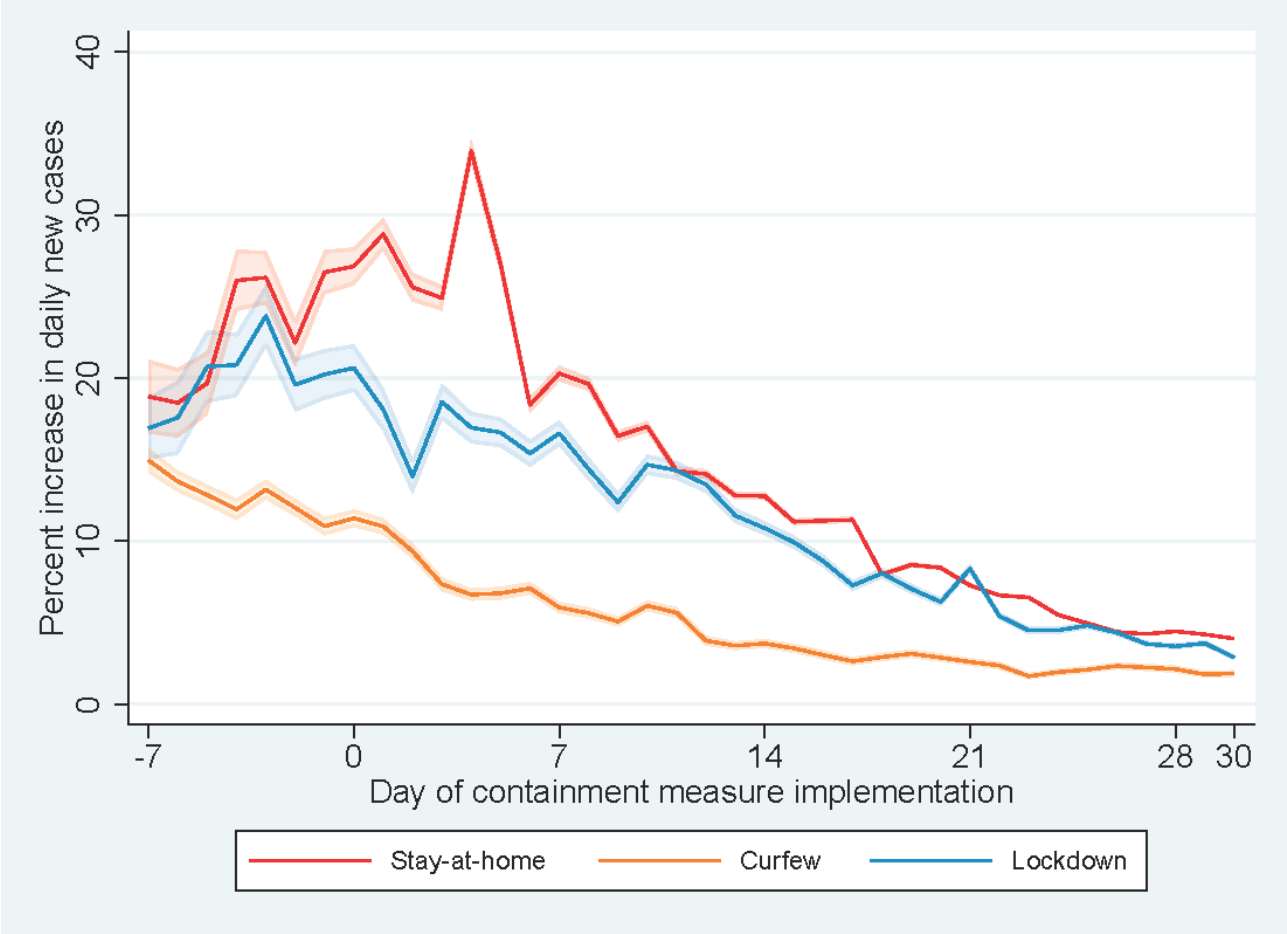
Results

This study included 122,366 patients with confirmed COVID-19 infection from 54 countries and 24,071 patients from 4 epicenters on the days when stay-at-home orders, curfews, or lockdowns were implemented between January 23 and April 11, 2020. The containment measures at the national level are summarized in Multimedia Appendix 1. Of the 54 countries, 31 (57%) adopted lockdowns, 17 (32%) adopted curfews, and 6 (11%) adopted stay-at-home orders. All countries initiated their containment policies after March 9, 2020, except for China (January 23, 2020).

Regarding the type of containment measure, countries adopting stay-at-home orders, curfew, and lockdown demonstrated a decreasing trend in percent increase in daily new cases of COVID-19 infection after the commencement of the measures (Figure 1). In countries with stay-at-home orders, curfews, or lockdowns, the percent increase was observed to be curbed to <5 within one month. The percent increase in daily new cases decreased from 26.9 (95% CI 25.7%-28.0%) at baseline to 20.3 (95% CI 19.8%-20.7%) at day +7, 12.8 (95% CI 12.6%-13.0%) at day +14, 7.29 (95% CI 7.17%-7.41%) at day +21, and 4.03 (95% CI 3.96%-4.10%) at day +30 for countries introducing stay-at-home orders. For countries introducing curfew, the percent increase in daily new cases decreased from 11.4 (95% CI 10.9-11.9) at baseline to 5.93 (95% CI 5.61-6.26) at day +7, 3.73 (95% CI 3.47-3.98) at day +14, 2.60 (95% CI 2.38-2.81) at day +21, and 1.89 (95% CI 1.71-2.07) at day +30. Meanwhile, the percent increase in daily new cases decreased from 20.6 (95% CI 19.2-22.1) at baseline to 16.6 (95% CI 15.9-17.4) at day +7, 10.8 (95% CI 10.4-11.2) at day +14, 8.32 (95% CI 8.06-8.57) at day +21 and 2.88 (95% CI 2.73-3.02) at day +30 for countries introducing lockdown. For the start date of the intervention (Figure 2), a persistent drop in the percent increase in daily new cases was also observed at the initial stage of the containment measures, except in China.

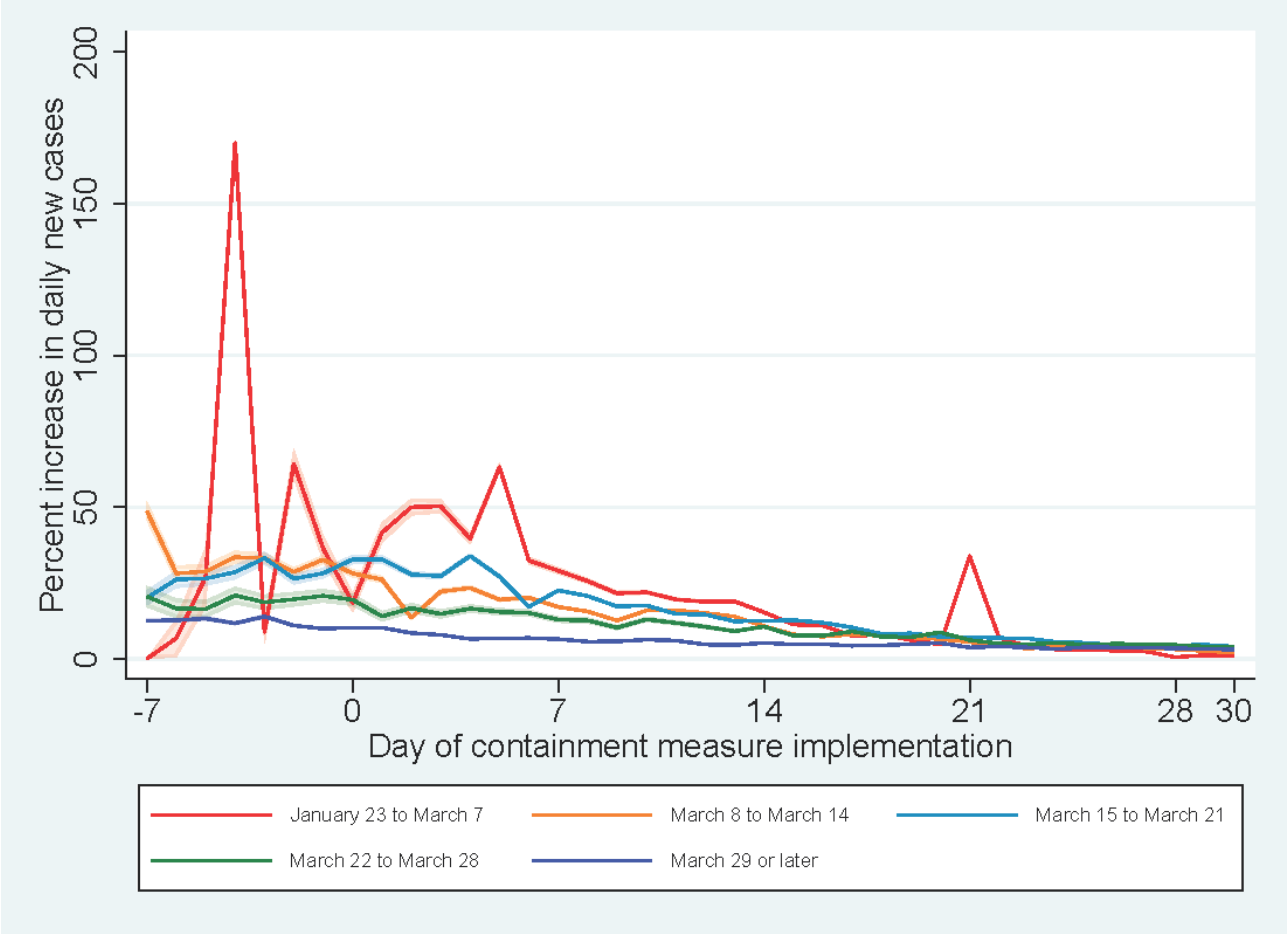
XSL•FO
RenderX

Figure 1. Percent increase in daily new cases vs days since containment measure by type of containment measure.



XSL•FO
RenderX

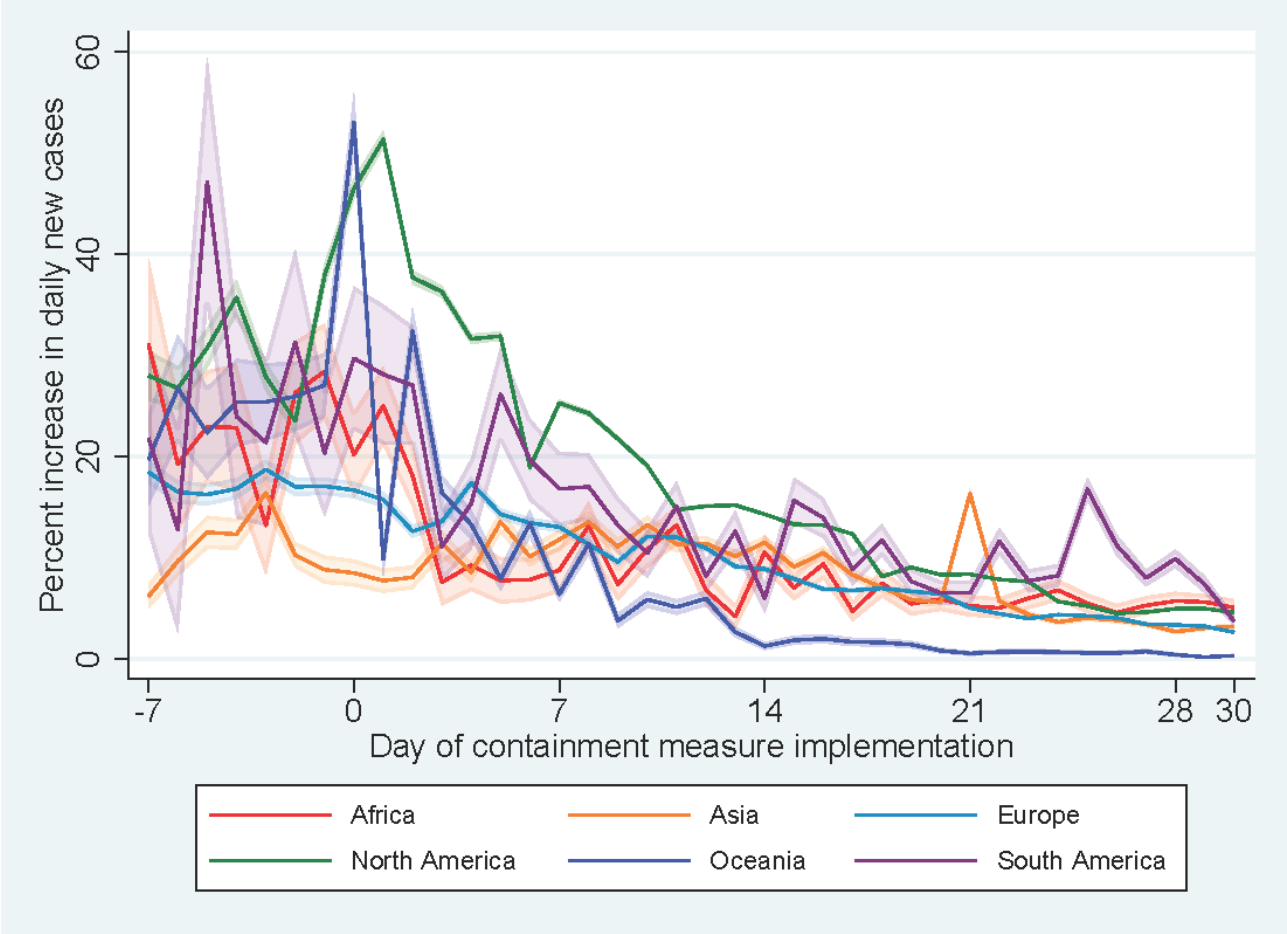
Figure 2. Percent increases in daily new cases vs days since implementation of containment measures by the start date of the containment measure.



Within seven days of commencement of the containment measures, a decreasing trend in the percent increase in daily new cases was observed across continents (Figure 3). The introduction of a new COVID-19 case classification in China at day +21 led to a spike in the percent increase in daily new cases in Asia. Resurgence in COVID-19 cases in South American countries was observed with a spike in percent increase in daily new cases at day +25, more than two weeks after the implementation of containment measures. By income level (Figure 4), a decreasing trend in the percent increase in daily new cases was observed for high-income countries since the initiation of containment measures.

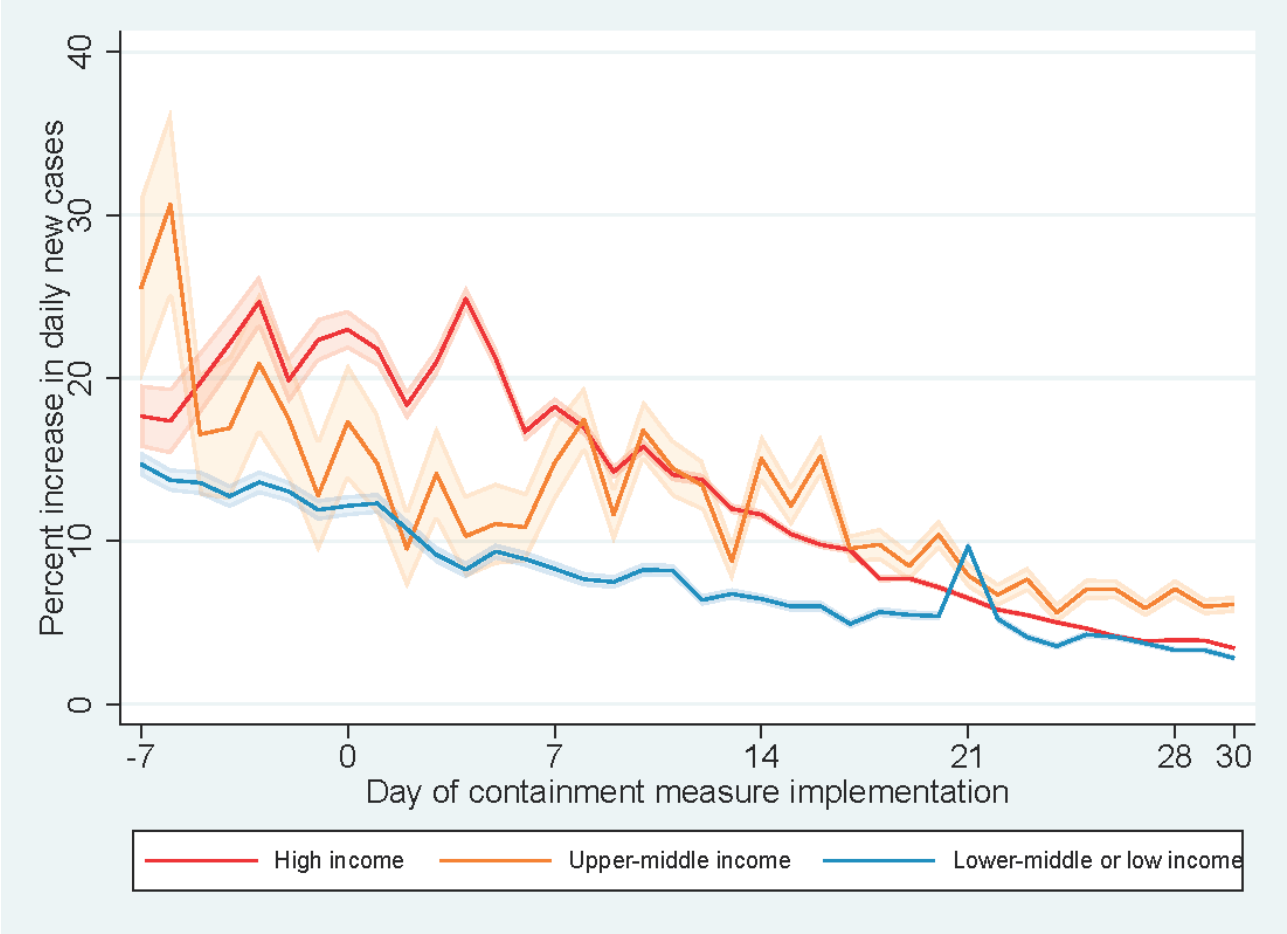
XSL•FO
RenderX

Figure 3. Percent increases in daily new cases against days since implementation of containment measures by continent.



XSL•FO
RenderX

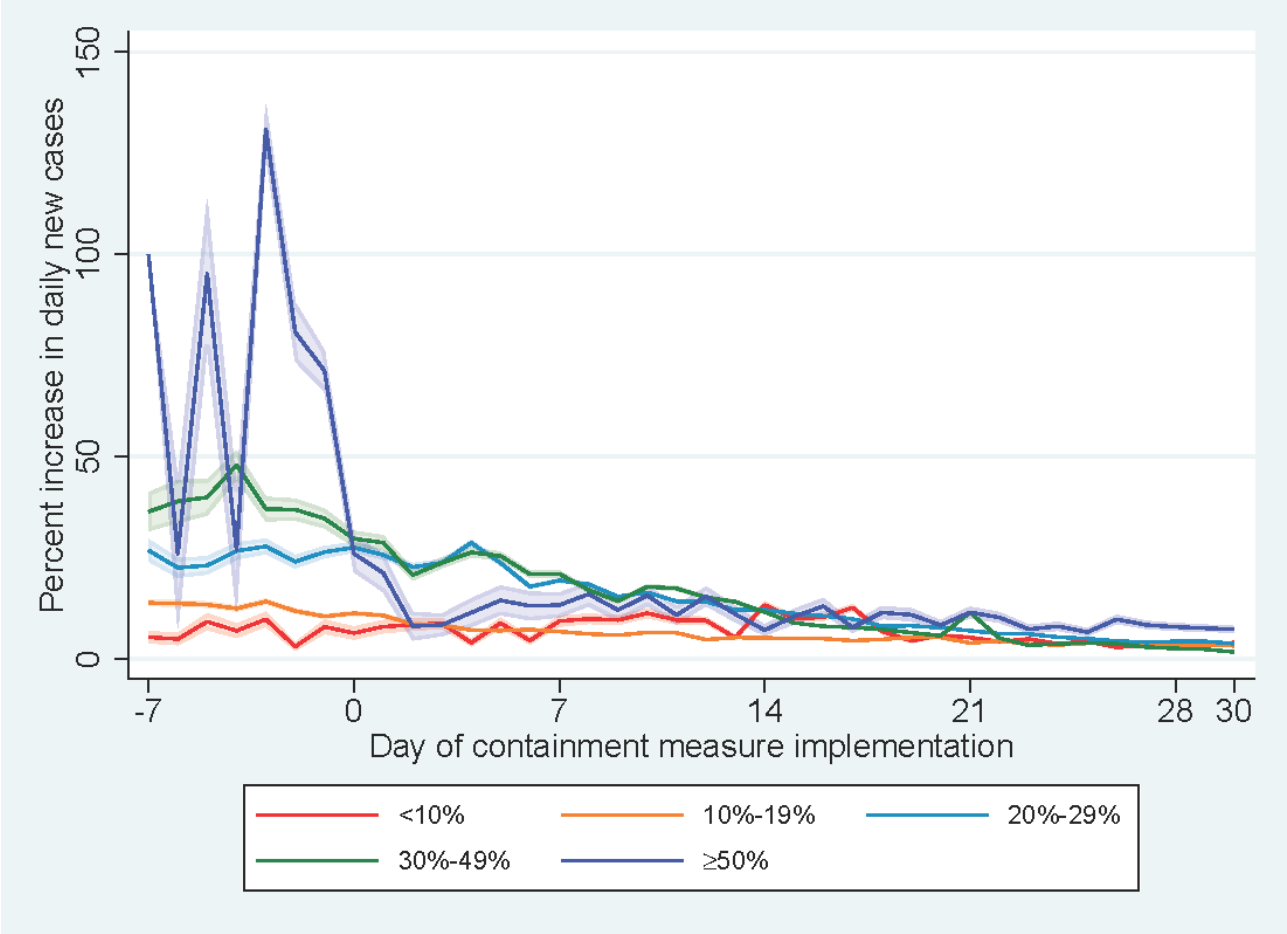
Figure 4. Percent increases in daily new cases vs days since implementation of containment measures by income level.



Among countries with an average percent increase in daily new cases above 10 over seven days prior to the commencement of containment measures, a decreasing trend in percent increase in daily new cases was identified following the implementation of a stay-at-home order, curfew, or lockdown (Figure 5). mong the countries studied, 20% (n=11) had an average increase in daily new cases of 30-49% a week before intervention, while 13% (n=7) had an average increase in daily new cases of ≥50%. For countries with an average percent increase in daily new cases between 30-49%, 15 and 23 days were required to reduce the percent increase in daily new cases to 10 and 5, respectively. The distributions of percent increases of daily new cases at day +7, day +14, day +21 and day +30 by average percent increase before the commencement of containment measures in countries and epicenters are shown in Figure 6. A majority of countries (n=45, 83%) experienced experienced a lower percent increase in daily new cases at day +7 than their respective average percent increases before the commencement of measures, whereas only 2 countries had a higher percent increase in daily new cases at both day +14 and day +21 than that prior to containment intervention. The percent increases in daily new cases at day +7, day +14, day +21, and day +30 for each type of containment measure were similar regardless of the start date of the stay-at-home order, curfew, or lockdown (Figure 7).

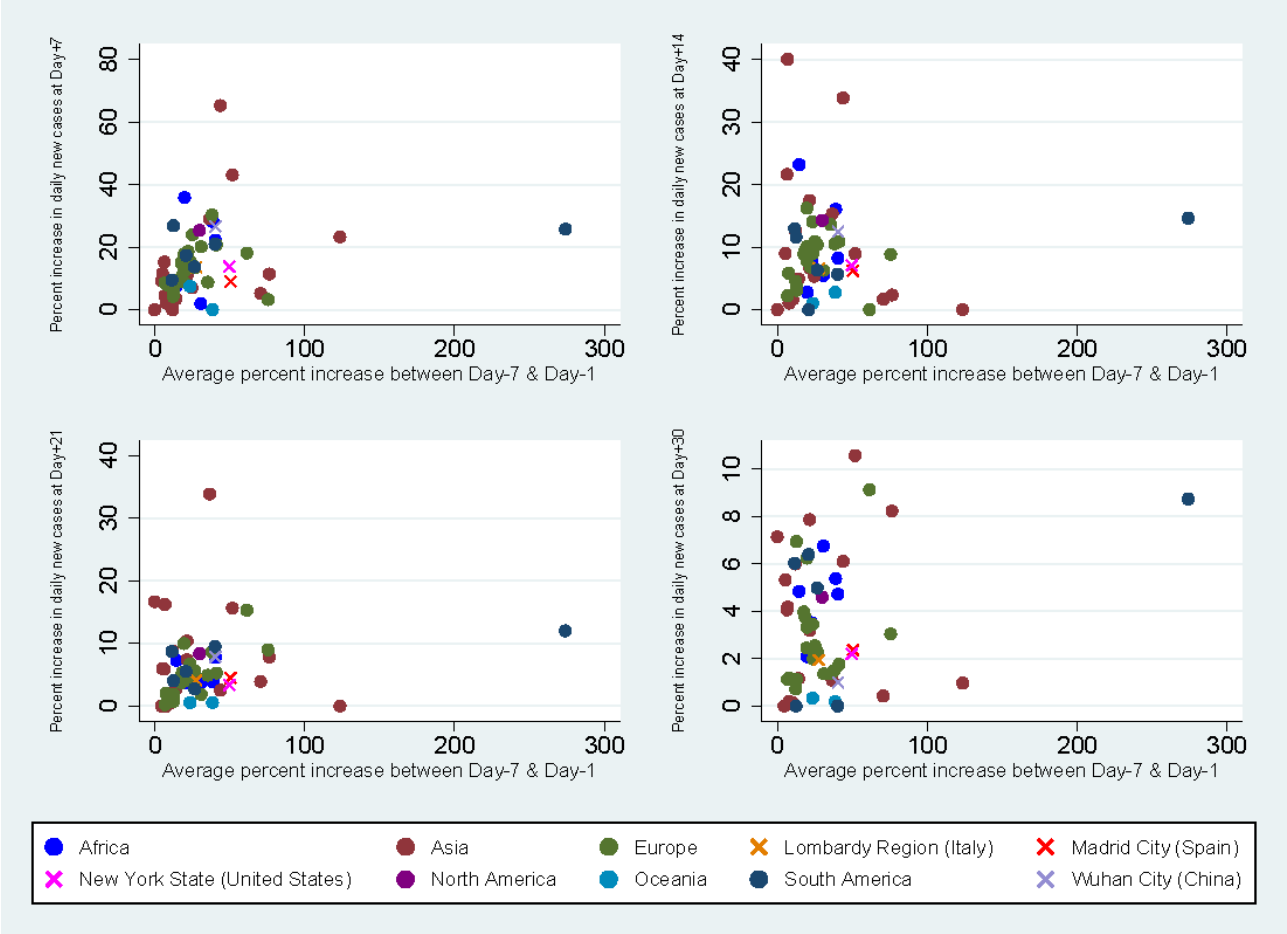
XSL•FO
RenderX

Figure 5. Percent increases in daily new cases vs days since implementation of containment measures by average percent increase in daily new cases before the containment measures.



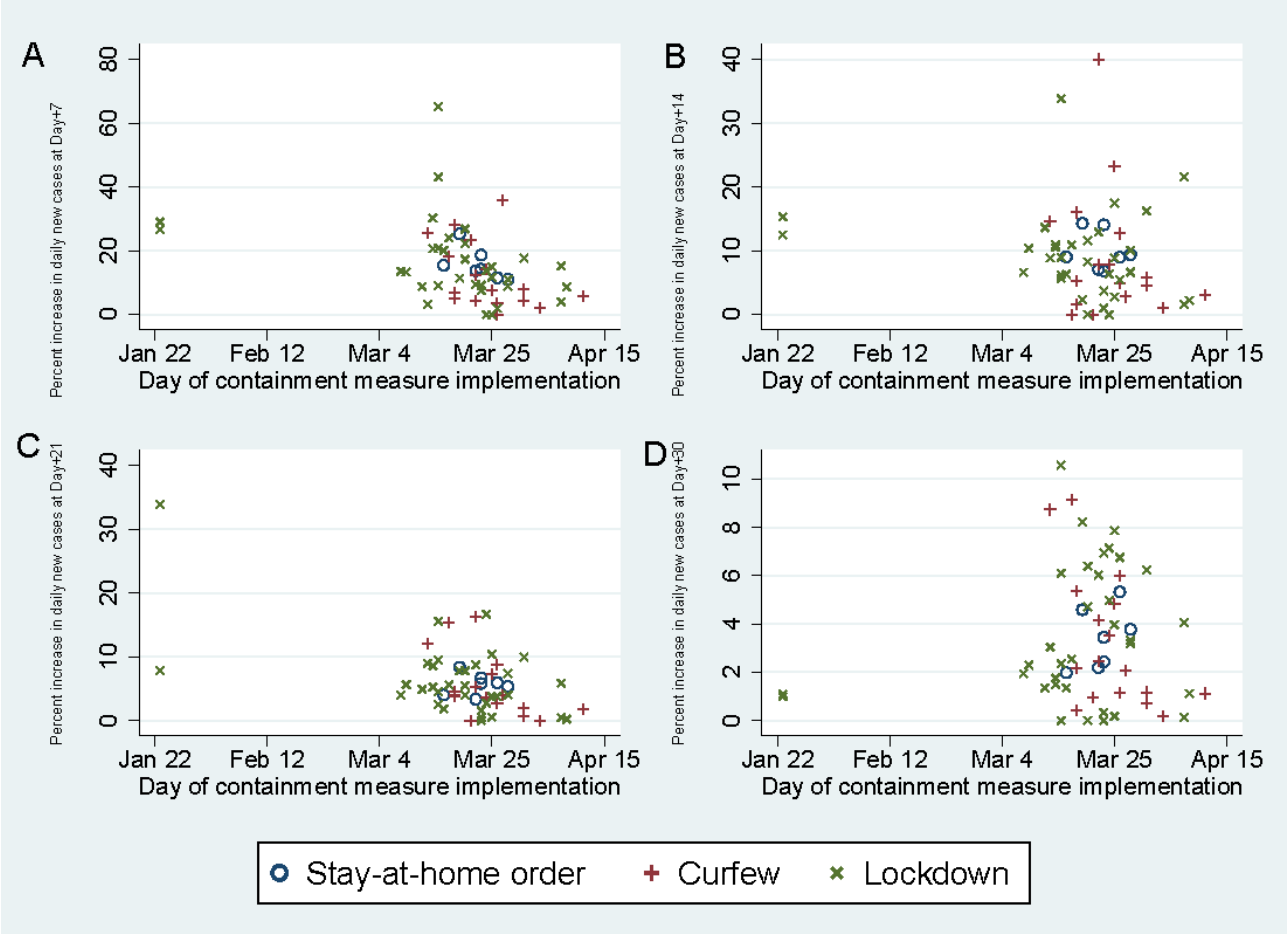
XSL•FO
RenderX

Figure 6. Percent increases in daily new cases at day +7 (A), day +14 (B), day +21 (C) and day +30 (D) against average percent increase in daily new cases before intervention by country and epicenter.



XSL•FO
RenderX

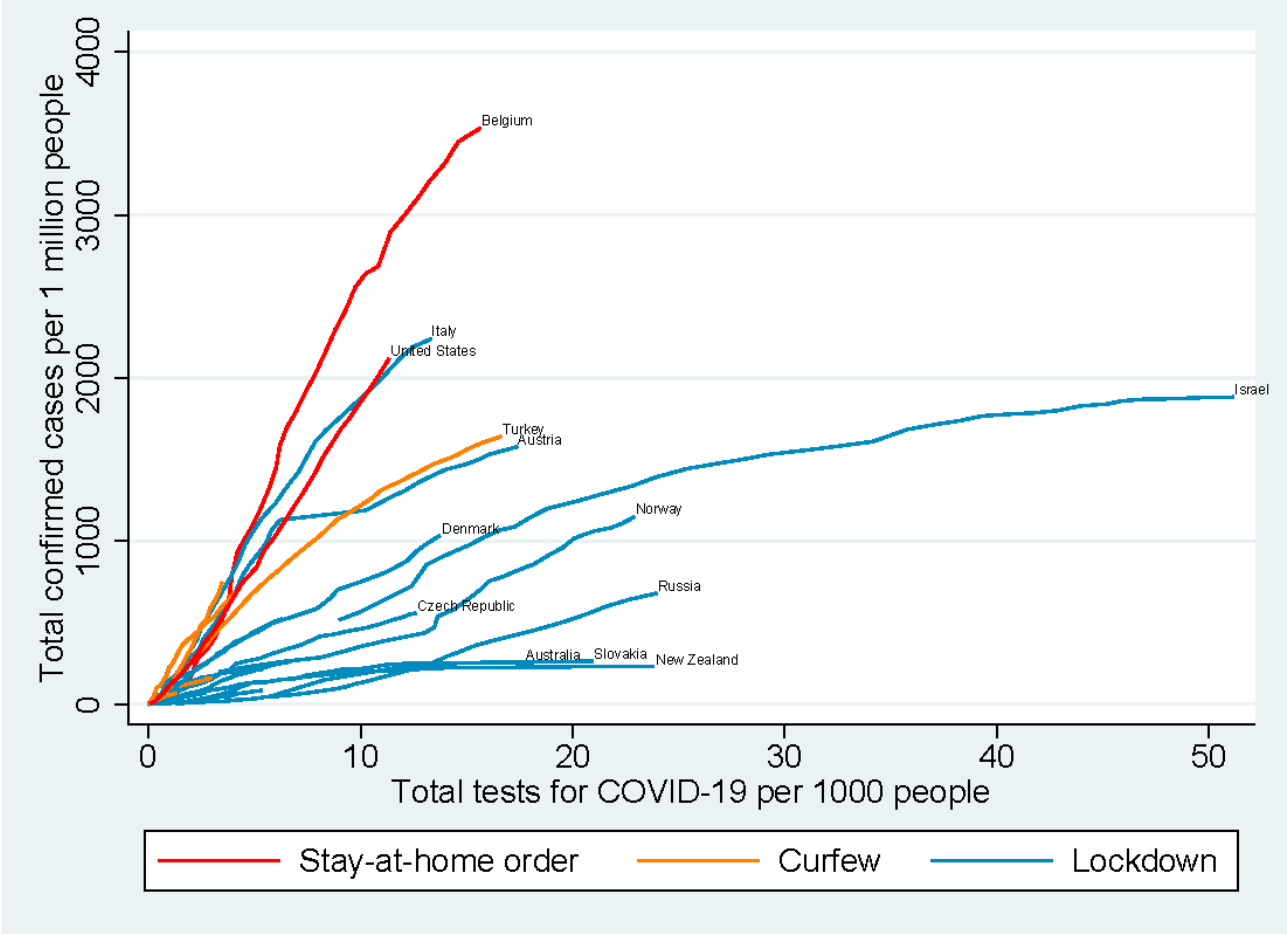
Figure 7. Percent increases in daily new cases at day +7 (A), day +14 (B), day +21 (C), and day +30 (D) vs start date of containment measure by type of containment measure.



With respect to the type of containment measures, Figure 8 illustrates the relationship between the number of COVID-19 tests performed per 1000 people and the number of confirmed cases per 1 million people as reported by the health authority of each country. Countries that initiated lockdown demonstrated flattened curves, implying a reduction of the number of new confirmed cases per 1 million people, while these countries greatly expanded their scope of PCR testing. Countries adopting stay-at-home orders or curfews demonstrated a linear relationship between the number of COVID-19 tests performed and the number of confirmed cases. Increasing the number of COVID-19 tests performed had adjunctive effects on curbing the increase in daily new cases among locked-down countries.

XSL•FO
RenderX

Figure 8. Total confirmed cases per 1 million people vs total tests for COVID-19 per thousand people in 54 countries by type of containment measure. COVID-19: coronavirus disease.



Discussion

Principal Findings

Our findings suggest that different national containment measures, namely stay-at-home orders, curfews, and lockdowns, are associated with a decrease in the percent increase in daily new cases of confirmed COVID-19 infection. This is consistent with our previous understanding that the virus can be transmitted through respiratory droplets; hence, reducing social contact and population movement may help reduce the spread of this infectious disease. Among countries adopting a lockdown approach, increasing the number of COVID-19 tests may have had adjunctive effects on curbing the increase of confirmed daily new cases. The effects of stay-at-home orders, as described in this study, were contributed by the United States, the United Kingdom, Belgium, Germany, Ireland, and Japan. However, the effects of these orders were likely underestimated on a global scale, given that residents in many regions and countries (such as South Korea [21], Vietnam, Taiwan [22], and Hong Kong [23]) were strongly advised to stay home rather than being ordered to do so, which also appears to have been effective in reducing disease transmission [24]. Lockdown serves to prevent movement of infected individuals between regions. Therefore, lockdown was implemented in epicenters as an altruistic measure to contain the epidemic. While lockdown can cause severe disruption of international interactions, especially in this globalized era, further research is warranted to investigate whether entry screening and quarantine would achieve similar outcomes to those of lockdown. It is our view that stay-at-home orders are adopted to minimize community transmission, and lockdowns (or border restrictions) aim to reduce the number of imported cases. Curfew, however, only limits physical contact among individuals during specific time periods; this may lead to a less observable effect, if any.

In the United States, the most prominent country in North America, there was no single effective date of a stay-at-home order at the country level, as containment measures and government responses to the COVID-19 pandemic varied from state to state (ranging from March 19 to April 11). California was the earliest state to enact a stay-at-home order, on March 19, 2020; this was assumed to be the effective date of national containment measures for the United States in this study. Most states in the United States subsequently imposed stay-at-home orders, leading to a time lag between the effective date and actual disease control. It can be speculated that the delay in commencing these containment measures across the country contributed to the spike in percent increase in daily new cases with stay-at-home orders for the first few days immediately following the first effective date of implementation, as the United States accounted for almost half the percent increase in daily new cases during the peak observed.

A persistently low percent increase in daily new cases was observed in Asia throughout the study period. An upsurge of

XSL•FO
RenderX

the number of COVID-19 cases alerted the health authorities in Asian countries and their residents during the early phase of the pandemic. With reference to the lessons learned during the SARS outbreak in 2003, Asian communities adopted various nonpharmaceutical measures to minimize disease transmission even before the announcement of health policies by local governments. This is evident in a report suggesting that nonpharmaceutical measures (such as hand hygiene, use of face masks, respiratory etiquette, and social distancing) have been incorporated into the local culture of Asian regions [25]. Countries in South America, however, showed signs of potential resurgence or a second wave of COVID-19 cases at the end of the study period. It is very concerning that containment measures did not appear to be effective in flattening the epidemic curve for individual countries in the region or that the health literacy of the public regarding COVID-19 infection should be improved. These concerns appear to be valid in light of the several subsequent spikes of percent increase in daily new cases identified across several South American countries (such as Argentina, Colombia, and Peru) following the lifting of containment measures before the disease was brought under control.

In addition to substantially lowering the percent increase in daily new cases in countries and regions with severe surges (≥100%), the investigated containment measures could also help less prevalent areas (<10%) to maintain a stable low percent increase in daily new cases. Accordingly, these interventions can be initiated at an early stage to flatten the curve and protect local health services from overwhelming demand. In view of the considerable economic cost associated with restriction of population movement, early public health interventions may also help alleviate the impact of prolonged recession and massive disruption of economic activities posed by mounting crises of newly confirmed cases and fatalities.

In this study, the relationships between the number of COVID-19 tests performed per 1000 people and the number of confirmed cases per 1 million people were examined by country and type of containment measures. These relationships were more obvious when the number of confirmed cases per million people was relatively small (around 1000 confirmed cases per 1 million people), which indicates the importance of leadership and coordination by local governments to make rapid expansion of PCR testing feasible and achievable in areas affected by the COVID-19 pandemic. Once the number of confirmed patients exceeded 1000, the effect of the same testing-to-population ratio almost disappeared; this may have occurred because health care facilities were overwhelmed by symptomatic patients, leaving no capacity to contact potential cases and conduct tests to identify an even larger number of carriers. Additionally, a country may fail to expand its testing capacity if the availability of testing agents becomes a rate-determining factor. Meanwhile, adoption of curfews and stay-at-home orders in countries did not appear to control the rise of new confirmed cases; hence, it can be postulated that testing must be coupled with other nonpharmaceutical public health control measures, such as quarantine and isolation, to exert effects on disease control.

The COVID-19 pandemic revealed how wealth inequity is associated with differential outcomes among countries. In our study, it was observed that high-income countries achieved a greater reduction in percent increase in daily new cases following the implementation of a stay-at-home order, curfew, or lockdown. This decline was more gradual among low-income and lower-middle-income countries, potentially because the crowded living conditions in these countries pose inherent risk of disease transmission. Moreover, other public health measures such as quarantine of exposed individuals may not be as effective in these countries because low-income employees who rely on their wages to meet financial obligations may avoid COVID-19 testing and subsequent forced quarantine. Therefore, development of a comprehensive public health intervention should be part of the strategic plan to manage the outbreak of infectious diseases alongside the promotion of health literacy of individuals, particularly in low-income countries and among people of lower social classes.

Limitations

The current study has several limitations. First, in an attempt to control the disease outbreak, many governments implemented multiple public health interventions simultaneously or within a short timeframe [26]; thus, individual strategies could not be evaluated independently. Second, the effects of different containment measures on disease transmission can vary depending on the patterns of social contact at home, schools, and workplaces, which would likely be influenced by local cultures and the enforcement of public regulations. Third, the observational design of this study precluded causal inference. However, clinical trials are neither feasible nor ethical during the current public health emergency, and compelling evidence comparing the experience of one outbreak area with another in which different policies were adopted is still lacking.

During the pandemic, crowdsourced data collection has played an increasingly important role in providing timely and accurate data for disease surveillance [27]. A number of local crowdsourcing platforms, such as *Ushahidi* [28], have been working together to deliver real-time reports to the public to raise public awareness and facilitate the development of local contingency measures. This collaboration represents a potential advancement in the publication of timely epidemiological analyses and reports; however, the accuracy, ethics, and confidentiality of research data should not be undermined [29].

Conclusion

National containment measures are essential to controlling the COVID-19 pandemic. In this study using crowdsourced data, countries that implemented stay-at-home orders, curfews, and lockdowns managed to curb the percent increase in daily new cases within a month; however, a resurgence in cases was observed in several South American countries two weeks after the commencement of containment measures. While no vaccine or effective treatment is yet available, the findings of our study can shed light on the impact of different containment measures and the priority of their implementation during this pandemic.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank *Our World in Data* for providing open-source data on confirmed COVID-19 cases and tests and Mr Owen Chu for statistical assistance. There was no funding source for this study.

Authors' Contributions

CW, JW, and AW contributed equally to all aspects of the study design and conduction, data interpretation, and writing of the manuscript. CW, ET, CHA, and KL contributed to the data extraction, analysis, and interpretation. AW contributed to the data validation. All authors approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Income group and date of stay-at-home order, curfew and lockdown by countries.
[DOCX File , 70 KB - jmir_v22i7e19904_app1.docx]

References

1. Azhar EI, Hui DSC, Memish ZA, Drosten C, Zumla A. The Middle East Respiratory Syndrome (MERS). Infect Dis Clin North Am 2019 Dec;33(4):891-905 [FREE Full text] [doi: 10.1016/j.idc.2019.08.001] [Medline: 31668197]
2. Hui DSC, Zumla A. Severe Acute Respiratory Syndrome: Historical, Epidemiologic, and Clinical Features. Infect Dis Clin North Am 2019 Dec;33(4):869-889 [FREE Full text] [doi: 10.1016/j.idc.2019.07.001] [Medline: 31668196]
3. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? Int J Epidemiol 2020 Feb 22:dyaa033 [FREE Full text] [doi: 10.1093/ije/dyaa033] [Medline: 32086938]
4. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses 2020 Feb 22;12(2):244 [FREE Full text] [doi: 10.3390/v12020244] [Medline: 32098422]
5. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
6. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis 2020 Jul;20(7):793-802 [FREE Full text] [doi: 10.1016/S1473-3099(20)30230-9] [Medline: 32247326]
7. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University & Medicine. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-07-02]
8. Country & Technical Guidance - Coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance [accessed 2020-07-16]
9. Updated WHO recommendations for international traffic in relation to COVID-19 outbreak. World Health Organization. 2020 Feb 29. URL: https://www.who.int/news-room/articles-detail/updated-who-recommendations-for-international-traffic-in-relation-to-covid-19-outbreak [accessed 2020-04-16]
10. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar;395(10228):931-934 [FREE Full text] [doi: 10.1016/s0140-6736(20)30567-5]
11. Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 2020 Mar 13;27(2):taaa020 [FREE Full text] [doi: 10.1093/jtm/taaa020] [Medline: 32052841]
12. Schmidt B, Davids EL, Malinga T. Quarantine alone or in combination with other public health measures to control COVID-19: A rapid Cochrane review. S Afr Med J 2020 Apr 30;110(6):476-477. [doi: 10.7196/samj.2020.v110i6.14847]
13. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 2020 May 15;368(6492):742-746 [FREE Full text] [doi: 10.1126/science.abb4557] [Medline: 32269067]
14. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, CMMID COVID-19 working group, Klepac P, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med 2020 May 07;18(1):124 [FREE Full text] [doi: 10.1186/s12916-020-01597-8] [Medline: 32375776]
15. Flaxman S, Mishra S, Gandy A, Unwin H, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Jun 08:online. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
16. Krieger M. Rt COVID-19. URL: https://rt.live/ [accessed 2020-07-02]

XSL•FO
RenderX

17. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
18. Ritchie H. Coronavirus Source Data. Our World in Data. URL: https://ourworldindata.org/coronavirus-source-data [accessed 2020-04-17]
19. World Bank Country and Lending Groups. World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-04-17]
20. Waller JL, Addy CL, Jackson KL, Garrison CZ. Confidence intervals for weighted proportions. Stat Med 1994 May 30;13(10):1071-1082. [doi: 10.1002/sim.4780131009] [Medline: 8073202]
21. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis 2020 Apr;93:339-344 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.031] [Medline: 32198088]
22. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. JAMA Intern Med 2020 May 01:online [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]
23. Chan YD, Flasche S, Lam TT, Leung MJ, Wong M, Lam H, et al. Transmission dynamics, serial interval and epidemiology of COVID-19 diseases in Hong Kong under different control measures. Wellcome Open Res 2020 May 11;5:91. [doi: 10.12688/wellcomeopenres.15896.1]
24. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288 [FREE Full text] [doi: 10.1016/S2468-2667(20)30090-6] [Medline: 32311320]
25. Wong T. Coronavirus: Why some countries wear face masks and others don't. BBC. 2020 May 12. URL: https://www.bbc.com/news/world-52015486 [accessed 2020-07-15]
26. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 Apr 10:1-9 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
27. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digit Health 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: 10.1016/S2589-7500(20)30026-1] [Medline: 32309796]
28. Coming Together to Improve Data Collection in Uncertain Times. Ushahidi. URL: https://www.ushahidi.com/covid [accessed 2020-04-19]
29. Leung GM, Leung K. Crowdsourcing data to mitigate epidemics. Lancet Digital Health 2020 Apr;2(4):e156-e157. [doi: 10.1016/s2589-7500(20)30055-8]

Abbreviations

COVID-19: coronavirus disease
MERS: Middle East respiratory syndrome
PCR: polymerase chain reaction
R0: initial transmission rate
Rt: effective transmission rate
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
SURGE: Systematic Urgent Review Group Effort

Edited by G Eysenbach, G Fagherazzi; submitted 06.05.20; peer-reviewed by J Yang, C Wang, C Mavrot; comments to author 18.06.20; revised version received 02.07.20; accepted 13.07.20; published 22.07.20.

Please cite as:
Wong CKH, Wong JYH, Tang EHM, Au CH, Lau KTK, Wai AKC
Impact of National Containment Measures on Decelerating the Increase in Daily New Cases of COVID-19 in 54 Countries and 4 Epicenters of the Pandemic: Comparative Observational Study
J Med Internet Res 2020;22(7):e19904
URL: http://www.jmir.org/2020/7/e19904/
doi:10.2196/19904
PMID:32658858

XSL•FO
RenderX

©Carlos K H Wong, Janet Y H Wong, Eric H M Tang, Chi Ho Au, Kristy T K Lau, Abraham K C Wai. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns

Henry C Cousins[1], MPhil; Clara C Cousins[2,3,4], BA; Alon Harris[5], MS, PhD; Louis R Pasquale[5], MD

[1]Department of Genetics, Stanford School of Medicine, Stanford, CA, United States

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Department of Data Sciences, Dana-Farber Cancer Institute, Harvard TH Chan School of Public Health, Boston, MA, United States

[4]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, United States

[5]Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Louis R Pasquale, MD
Department of Ophthalmology
Icahn School of Medicine at Mount Sinai
One Gustave L Levy Place
Box 1183
New York, NY, 10029
United States
Phone: 1 212 241 6752
Email: louis.pasquale@mssm.edu

Abstract

Background: Timely allocation of medical resources for coronavirus disease (COVID-19) requires early detection of regional outbreaks. Internet browsing data may predict case outbreaks in local populations that are yet to be confirmed.

Objective: We investigated whether search-engine query patterns can help to predict COVID-19 case rates at the state and metropolitan area levels in the United States.

Methods: We used regional confirmed case data from the New York Times and Google Trends results from 50 states and 166 county-based designated market areas (DMA). We identified search terms whose activity precedes and correlates with confirmed case rates at the national level. We used univariate regression to construct a composite explanatory variable based on best-fitting search queries offset by temporal lags. We measured the raw and z-transformed Pearson correlation and root-mean-square error (RMSE) of the explanatory variable with out-of-sample case rate data at the state and DMA levels.

Results: Predictions were highly correlated with confirmed case rates at the state (mean r=0.69, 95% CI 0.51-0.81; median RMSE 1.27, IQR 1.48) and DMA levels (mean r=0.51, 95% CI 0.39-0.61; median RMSE 4.38, IQR 1.80), using search data available up to 10 days prior to confirmed case rates. They fit case-rate activity in 49 of 50 states and in 103 of 166 DMA at a significance level of .05.

Conclusions: Identifiable patterns in search query activity may help to predict emerging regional outbreaks of COVID-19, although they remain vulnerable to stochastic changes in search intensity.

(J Med Internet Res 2020;22(7):e19483) doi:10.2196/19483

KEYWORDS

epidemiology; infoveillance; COVID-19; internet activity; Google Trends; infectious disease; surveillance; public health

Introduction

Early detection of regional coronavirus disease (COVID-19) outbreaks is essential for efficient medical resource allocation, public health messaging, and implementation of infection prevention and control strategies [1]. It is particularly important given the probability of future waves of COVID-19 cases and the difficulty of applying traditional epidemiological forecasting models in areas with low case levels [2,3]. However, laboratory testing capacity is limited, and confirmed case reports lag behind underlying infections, decreasing their predictive capacity in the early days of an outbreak or resurgence.

Internet browsing data, such as search-engine query results, can provide a real-time indication of symptoms in a population and

XSL•FO
RenderX

have been used extensively to predict and model outbreaks like influenza and dengue [4-7]. Such methods generally assume that specific and detectable patterns in internet behavior, such as search trends or social media postings, reflect health-seeking behavior in real time at the population level. Forecasting models based on search queries, such as Google Flu Trends, have shown predictive value without direct reliance on formal case reports, although historical inaccuracies mean that they can only supplement, not replace, traditional forecasting methodologies based on confirmed cases [8-10].

COVID-19 case rates display significant regional heterogeneity that requires locally tailored containment strategies. Google search trends, encompassing a majority of internet queries in the United States and publicly available through Google Trends (GT), provide a powerful resource for systematic comparison of browsing behavior between US regions. We hypothesized that keyword libraries could be screened for specific terms whose aggregate activity would reflect regional differences in COVID-19 case rates, as has been demonstrated for influenza [4]. While several studies have previously attempted to model the COVID-19 pandemic using search query data, such attempts have largely focused on specific regions, like Taiwan and Iran, and a limited number of individually selected search terms [11-14]. We explored the potential of large-scale, publicly accessible search query data to signal new COVID-19 cases at the state and metropolitan-area levels in the United States.

Methods

Data Collection and Processing

We obtained confirmed case data for US states and counties from the New York Times (NYT) data set from January 21, the date of the first confirmed US case, to April 2, 2020, comprising county-specific, lab-confirmed COVID-19 case reports compiled daily from local and state health authorities [15]. We used the NYT data set because of both its inclusion of county-level case geotags and its strong correlation with other case tracking sources [16]. Next, we used GT to compile a library of 463 unique search queries and their associated daily activity levels over the same time period. Library terms were automatically retrieved based on the likelihood of user association with a set of prespecified coronavirus-related seed terms (Multimedia Appendix 1), using the GT "Related Queries" function.

We compared the z-transformed correlation of each query's search activity with an in-sample data set comprising daily confirmed national cases rates for days through March 10 with >100 new cases per day. Each query's search activity was offset by temporal lags of between 0 and 14 days, generating a list of best-fitting queries and their associated optimal lag times. To focus on terms with early predictive power, we excluded queries whose optimal lag was less than 9 days. We selected the five best-fitting queries and constructed a single explanatory variable by summing the lag-adjusted, relative activity levels of each query. Finally, we linearly fit the explanatory variable to national data through March 10 to generate a single scalar coefficient.

Data Analysis

We measured the correlation of state-specific activity levels for our explanatory variable with daily reported case levels in individual states using out-of-sample data from March 11 through April 2. We also measured how well the explanatory variable explained out-of-sample case rates in 166 designated market areas (DMA), which are collections of approximately 15 counties each constituting the highest-resolution regional data available on GT. Means and confidence intervals for correlation coefficients were calculated using the inverse z-transformation of the averaged z-transformed coefficients. The strength of model predictions over time was measured using a partial correlation of first confirmed case dates with z-transformed correlation coefficients in all regions with >100 cases, controlling for regional population. We used root-mean-square error (RMSE) as an additional measure of model performance. Model predictions were adjusted for regional population and internet access [17]. All data were anonymous, and the study protocol was approved by the institutional review board of the Icahn School of Medicine at Mount Sinai.

Results

Search Query Characteristics

Queries incorporated into the final explanatory variable were highly correlated with national case data, with correlation coefficients ranging from 0.996 to 0.999 on the in-sample data. The optimal temporal lags for incorporated queries were from 11 to 12 days, indicating a prediction horizon of up to 10 days (assuming that a day's full GT query results become available on the subsequent day). The final variable, the linear sum of weighted, lag-adjusted activity levels for the five best-fitting terms from the 463-term library, fit the in-sample data with a correlation of 0.998.

Characteristics of additional screened queries validated our methodology. For instance, acute topics like medical care and testing had smaller associated lag times with confirmed case rates, as would be expected for urgent inquiries (Table 1). Queries unrelated to COVID-19 had correspondingly weaker correlations with the observed data. The best-fitting category of queries was "COVID-19 guidance," which included terms related to coronavirus-specific medical advice from health authorities. Relative levels of search activity had no significant effect on fit with case data.

XSL•FO
RenderX

Table 1. Characteristics of query topics screened for fit with coronavirus disease (COVID-19) case data.

Search query category	Unique queries[a], n (%)	Correlation with national case rate[b], mean	Associated lag time (days)[c], mean	Activity weighting[d]
COVID-19 guidance	32 (6.9)	0.96	9.1	0.38
COVID-19 news	57 (12.3)	0.96	8.3	1.00
COVID-19 symptoms	91 (19.7)	0.94	8.9	0.41
Medical treatments	34 (7.3)	0.93	10.1	0.31
COVID-19 testing	58 (12.5)	0.89	5.4	0.11
Medical care	33 (7.1)	0.89	7.2	0.60
Nonspecific symptoms	62 (13.4)	0.89	6.8	0.57
Economic effects	28 (6.0)	0.86	5.9	0.12
Unrelated to illness	51 (11.0)	0.86	6.6	0.76
Symptoms of other illnesses	17 (3.7)	0.84	8.3	0.77

[a]Number of queries of each type in the query library (eg, the category "COVID-19 testing" would include the specific query "coronavirus test near me," and the category "nonspecific symptoms" would include the query "cough").

[b]Expressed as the inverse z-transformation of the averaged z-transformed correlations with in-sample national data.

[c]Mean lag time between best-fitting query activity and confirmed case rate, in days.

[d]Relative mean search activity levels, normalized.

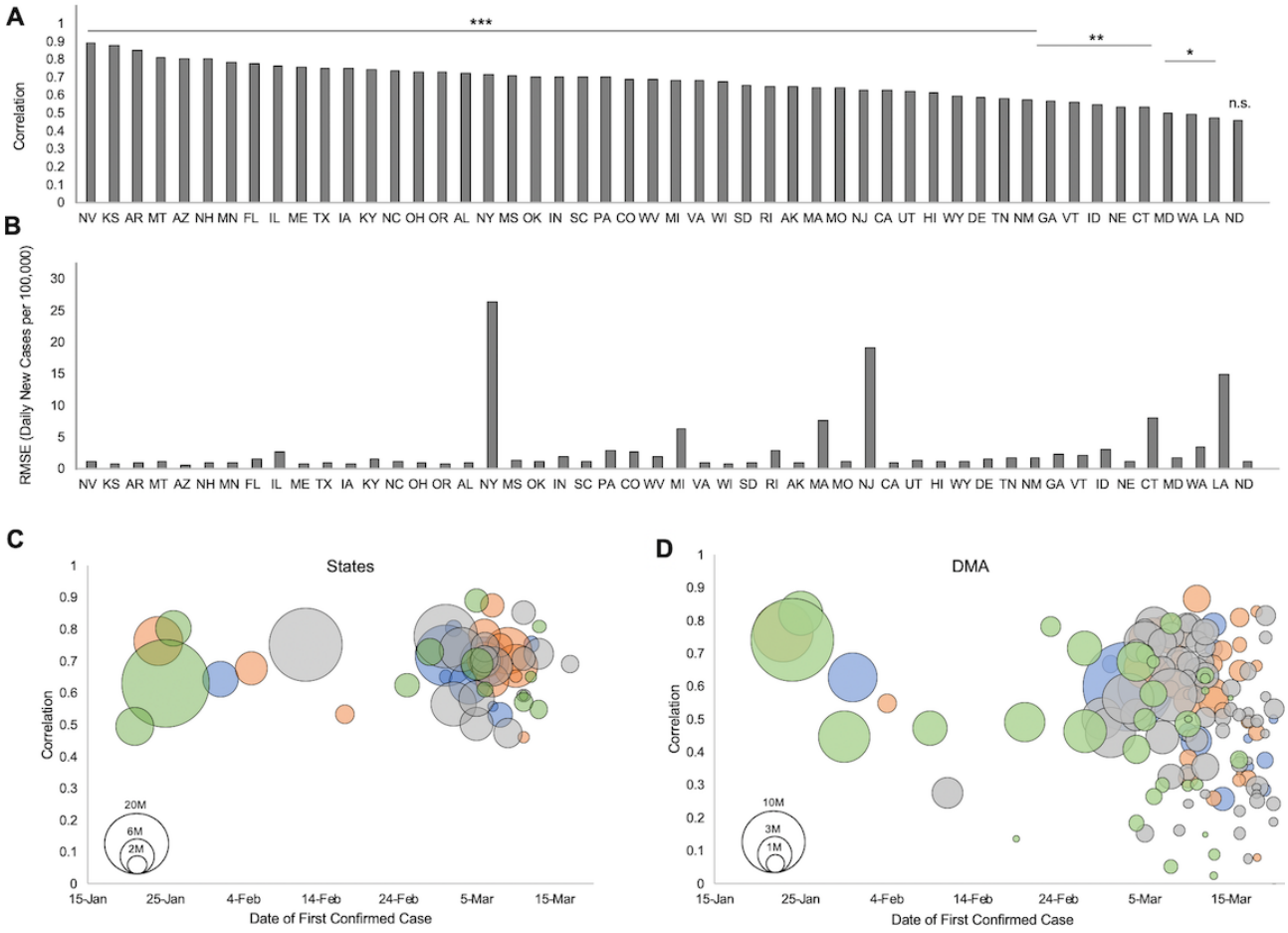
Regional Case-Rate Predictions

The query-based predictions fit well with out-of-sample case rate data at the national level, with a correlation of 0.84 (P<.001) for out-of-sample data and 0.83 (P<.001) for all available data. The predictions were also well correlated at the state level in nearly all cases, fitting case data in 49 of 50 states at a significance level of α=.05 and 41 of 50 states at α=.005 (mean r=0.69; 95% CI 0.51-0.81; Figure 1A; Multimedia Appendix 2). RMSE was less than 4 cases per 100,000 residents for model predictions in 44 of 50 states (median 1.27; IQR 1.48; Figure 1B).

At the DMA level, the query-based predictions fit with daily case data for 62% (103/166) of regions at α=.05, or 79% (84/107) excluding DMA with fewer than 100 cases (mean r for all DMA=0.51; 95% CI 0.39-0.61; Multimedia Appendix 3). RMSE was slightly higher for DMA-level compared to state-level predictions but was less than 7 for 92% (152/166) of DMA (median 4.38; IQR 1.80). Furthermore, at both the state and DMA levels, the strength of the correlation was not significantly associated with the date of a region's first confirmed case (P=.51 for states and P=.71 for DMA for partial correlations in regions with >100 cases, controlling for population), suggesting that predictive search behaviors may precede new cases regardless of the timing of a regional outbreak (Figure 1C and D). The explanatory variable consistently produced well-fitting predictions with data available 10 days in advance of predicted new case activity (Figure 2), even in regions where fewer than 100 new cases were confirmed per day.

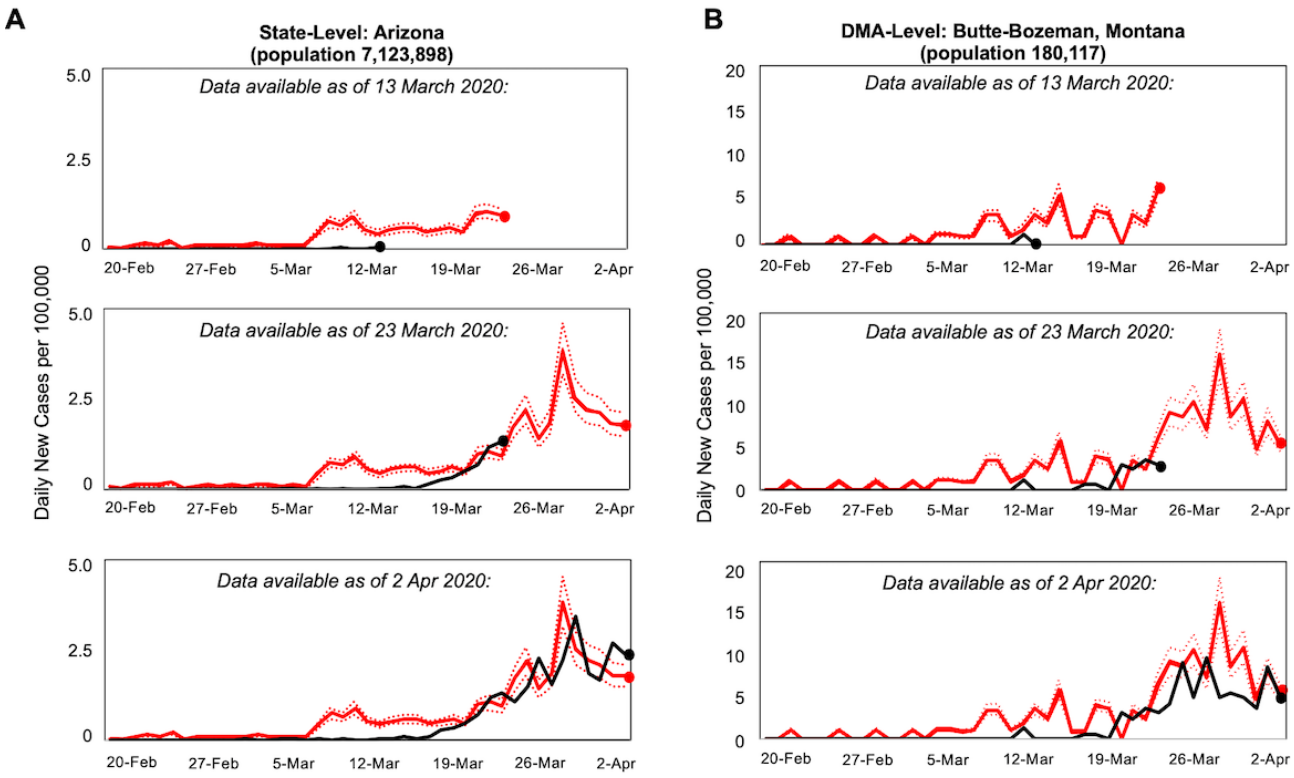
XSL•FO
RenderX

Figure 1. Correlation of query predictions with regional coronavirus disease (COVID-19) confirmed case rates. (A) Correlation of predicted case rates with actual case rates for the 50 states. Values are Pearson correlation coefficients. * indicates significance at α=.05; ** at α=.01; *** at α=.005. (B) Root-mean-square error (RMSE) between predicted case rates and actual case rates for the 50 states, in units of daily new cases per 100,000 population. (C) Prediction correlations at the state level do not depend on outbreak timing, as measured by the date of the first confirmed case. Circle size indicates the relative population of the state. Color indicates US census-designated region (blue: Northeast; orange: Midwest; gray: South; green: West). (D) Prediction correlations at the designated market area (DMA) level do not depend on outbreak timing, as measured by the date of the first confirmed case. Circle size indicates the relative population of DMA. Color indicates the US census-designated region, as described. n.s.: not significant.



XSL•FO
RenderX

Figure 2. Correlation of query predictions (red) with regional coronavirus disease (COVID-19) case rates (black) at the state and designated market area (DMA) levels, February 20 to April 2, 2020. (A) Comparison of predicted case rates (red) with actual case rates (black) at the state level, with Arizona shown as an example. Dashed lines indicate 95% CIs. (B) Comparison at the DMA level, with the Butte-Bozeman area shown as an example of predictions in a low-population region.



Discussion

These data suggest that specific patterns of internet search behavior, which can be curated automatically from libraries of search terms, precede and correlate with regional case rates of COVID-19. Such patterns, which we captured using a single explanatory variable, remain correlated with case rates in regions with a broad range of populations, locations, and outbreak times, making aggregate search trends a useful tool for estimating regional COVID-19 outbreaks in the days preceding confirmed case reports. Correlation strengths were not significantly associated with the date of onset of regional outbreaks, making it unlikely that a single national event, such as a press release, could explain the strength of model predictions in all regions. Furthermore, search queries explicitly related to COVID-19 have more predictive power than unrelated keywords, and acute queries, such as those concerning testing or medical care, have smaller associated lag times.

Taken together, these results suggest that systematic screening of key term libraries can identify search queries reflecting real-time health-seeking behaviors at the regional level, expanding the suite of “infoveillance” methods that may assist in monitoring COVID-19 cases. This type of approach does not directly depend on either regional testing capacity or local media reports, making it particularly relevant in areas with small populations, limited medical infrastructure, or low case numbers. Such information can supplement traditional epidemiological approaches, such as estimates based on a compartmental framework, to guide community health interventions in the early days of an outbreak.

Several aspects of query-based approaches to case estimation, such as this work, must be further characterized for COVID-19. First, while correlations were statistically strong across most US regions, elevated RMSE indicated lower accuracy for predictions in the New York City and New Orleans areas, both regions with major outbreaks. However, comparable losses in accuracy were not observed for other major outbreak sites, such as Philadelphia, Los Angeles, or Chicago. This may reflect region-specific differences in both internet browsing behavior and patterns of community infection and may be a limitation of query-based models using fixed terms. As evidenced by previous attempts to predict influenza outbreaks based on search data, browsing behavior will also likely change as public understanding evolves over the course of disease spread [18]. Therefore, search-term relevance is likely to vary with time, which may require continuous supplementation or reselection of query terms to ensure representativeness of current population behaviors. Furthermore, although we generally observed strong historical correlations, query-based models must also be monitored for sudden changes in COVID-related query activity due to external events, such as unrelated news reports. Such distortions would be particularly important in regions with limited internet access. Future models incorporating learning and real-time updating of region-specific search terms may improve query-based prediction efforts for future COVID-19 outbreaks.

XSL•FO
RenderX

Acknowledgments

HCC, CCC, and LRP conceptualized the study and designed the methodology. HCC and CCC conducted the data analysis. HCC, CCC, LRP, and AH wrote the manuscript. The corresponding author had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HCC is supported by a graduate fellowship award from Knight-Hennessy Scholars at Stanford University.

Conflicts of Interest

Unrelated to this work, AH would like to disclose that he receives remuneration from AdOM for serving as a consultant and a board member, and from Thea for a speaking engagement. AH also holds an ownership interest in AdOM, Luseed, Oxymap, and QuLent. Unrelated to this work, LRP is a consultant to Eyenovia, Bausch+Lomb, Nicox, Emerald Bioscience, and Verily. No relevant financial relationship exists for any of the other authors.

Multimedia Appendix 1
Seed query terms.
[DOCX File , 14 KB - jmir_v22i7e19483_app1.docx]

Multimedia Appendix 2
Performance of model predictions for individual states.
[DOCX File , 17 KB - jmir_v22i7e19483_app2.docx]

Multimedia Appendix 3
Performance of model predictions for individual designated market areas (DMA).
[DOCX File , 33 KB - jmir_v22i7e19483_app3.docx]

References

1. Heymann DL, Shindo N. COVID-19: what is next for public health? The Lancet 2020 Feb;395(10224):542-545. [doi: 10.1016/s0140-6736(20)30374-3]
2. Gander K. CDC director says there may be another coronavirus wave in late fall and a "substantial portion of Americans" will be susceptible. Newsweek. 2020 Apr 1. URL: https://www.newsweek.com/cdc-director-coronavirus-wave-late-fall-substantial-portion-americans-will-susceptible-1495401 [accessed 2020-04-06]
3. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. Proc Natl Acad Sci 2020 Jul 02. [doi: 10.1073/pnas.2006520117] [Medline: 32616574]
4. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
5. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea. J Med Internet Res 2016 Jul 04;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]
6. Yang S, Kou SC, Lu F, Brownstein JS, Brooke N, Santillana M. Advances in using Internet searches to track dengue. PLoS Comput Biol 2017 Jul;13(7):e1005607 [FREE Full text] [doi: 10.1371/journal.pcbi.1005607] [Medline: 28727821]
7. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. J Med Internet Res 2005 Jul 1;7(3):e36 [FREE Full text] [doi: 10.2196/jmir.7.3.e36] [Medline: 15998627]
8. Pervaiz F, Pervaiz M, Abdur RN, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012;14(5):e125 [FREE Full text] [doi: 10.2196/jmir.2102] [Medline: 23037553]
9. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
10. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013;9(10):e1003256 [FREE Full text] [doi: 10.1371/journal.pcbi.1003256] [Medline: 24146603]
11. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
12. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]

XSL•FO
RenderX

13. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):e [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
14. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul 11;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
15. We're Sharing Coronavirus Case Data for Every U.S. County. The New York Times. 2020 Mar 28. URL: https://www.nytimes.com/article/coronavirus-county-data-us.html [accessed 2020-04-06]
16. Smith J. Comparison of COVID-19 case and death counts in the United States reported by four online trackers: January 22-May 31, 2020. medRxiv 2020 Jun 25 [FREE Full text] [doi: 10.1101/2020.06.20.20135764]
17. Usage and Population Statistics. Internet World Stats. URL: https://www.internetworldstats.com/ [accessed 2020-04-06]
18. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

Abbreviations

COVID-19: coronavirus disease
DMA: designated market area
GT: Google Trends
NYT: New York Times
RMSE: root-mean-square error

Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by JM Ocampo, A Daughton, S Pesälä; comments to author 20.06.20; revised version received 13.07.20; accepted 19.07.20; published 30.07.20.

Please cite as:
Cousins HC, Cousins CC, Harris A, Pasquale LR
Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns
J Med Internet Res 2020;22(7):e19483
URL: http://www.jmir.org/2020/7/e19483/
doi:10.2196/19483
PMID:32692691

©Henry C Cousins, Clara C Cousins, Alon Harris, Louis R Pasquale. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Multiple Epidemic Wave Model of the COVID-19 Pandemic: Modeling Study

Efthimios Kaxiras[1], PhD; Georgios Neofotistos[2], PhD

[1]Department of Physics, Harvard University, Cambridge, MA, United States

[2]Institute of Advanced Computational Science, J.A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States

Corresponding Author:
Efthimios Kaxiras, PhD
Department of Physics
Harvard University
Lyman Lab 339
17 Oxford Street
Cambridge, MA, 02138
United States
Phone: 1 617 495 7977
Email: kaxiras@physics.harvard.edu

Abstract

Background: Intervention measures have been implemented around the world to mitigate the spread of the coronavirus disease (COVID-19) pandemic. Understanding the dynamics of the disease spread and the effectiveness of the interventions is essential in predicting its future evolution.

Objective: The aim of this study is to simulate the effect of different social distancing interventions and investigate whether their timing and stringency can lead to multiple waves (subepidemics), which can provide a better fit to the wavy behavior observed in the infected population curve in the majority of countries.

Methods: We have designed and run agent-based simulations and a multiple wave model to fit the infected population data for many countries. We have also developed a novel Pandemic Response Index to provide a quantitative and objective way of ranking countries according to their COVID-19 response performance.

Results: We have analyzed data from 18 countries based on the multiple wave (subepidemics) hypothesis and present the relevant parameters. Multiple waves have been identified and were found to describe the data better. The effectiveness of intervention measures can be inferred by the peak intensities of the waves. Countries imposing fast and stringent interventions exhibit multiple waves with declining peak intensities. This result strongly corroborated with agent-based simulations outcomes. We also provided an estimate of how much lower the number of infections could have been if early and strict intervention measures had been taken to stop the spread at the first wave, as actually happened for a handful of countries. A novel index, the Pandemic Response Index, was constructed, and based on the model's results, an index value was assigned to each country, quantifying in an objective manner the country's response to the pandemic.

Conclusions: Our results support the hypothesis that the COVID-19 pandemic can be successfully modeled as a series of epidemic waves (subepidemics) and that it is possible to infer to what extent the imposition of early intervention measures can slow the spread of the disease.

(J Med Internet Res 2020;22(7):e20912) doi:10.2196/20912

KEYWORDS

COVID-19; multiple waves; transmission; intervention measures; simulations; modeling; pandemic response index; pandemic; virus; intervention

XSL•FO
RenderX

Introduction

The coronavirus disease (COVID-19) pandemic has produced a great number of studies that aim to understand the dynamics of the disease spread and predict its future evolution (see, for example, [1-8]). Different types of models can be assumed to describe this dynamic evolution. In studying past epidemics, scientists have systematically applied "random mixing" models that assume that an infectious individual may spread the disease to any susceptible member of the population, as originally proposed by Kermack and McKendrick [9]. Recent approaches consider mobility and contact networks [10,11], epidemic waves attributable to community networks [12], subepidemic modeling [13], Bayesian modeling and inference [14], models of spatial contacts in large-scale artificial cities [15], and power-law models of infectious disease spread [16], to name but a few representative examples.

At present, there is a trove of data from different countries that can serve to put strict limits on plausible models. In this paper we use simulations from agent-based models and simple analytical solutions to fit reported data from a range of countries. This provides a comprehensive picture of likely scenarios of how the disease evolved in various countries. These scenarios can be useful in predicting the future spread of the disease and provide insight on how the imposition of social-distancing measures can be effective in containing or slowing its spread.

Our work is based on two premises. First, the apparent regular features in the reported infections in many countries are not random but rather contain useful information, as their persistence and regularity suggest. Second, there is a general underlying dynamic of the spread of the disease, *in spirit* similar to the original Kermack-McKendrick model of three populations [9], the "susceptible population" $S(t)$, the "infected/infectious population" $I(t)$, and the "removed/recovered population" $R(t)$, which are related by $S(t) + I(t) + R(t) = N$, where N is the total population. The time evolution of the *susceptible-infectious-removed* (SIR) populations is described by the equations:

The SIR model involves two positive parameters, β and γ, which have the following meanings:

- β describes the effective contact rate of the disease; an infected individual comes into contact with β other individuals per unit time (the fraction that are susceptible to contracting the disease is S/N).
- γ is the mean removal (recovery) rate, that is, $1/\gamma$ is the mean period of time during which an infected individual can pass it on before being removed from the group of the infected individuals.

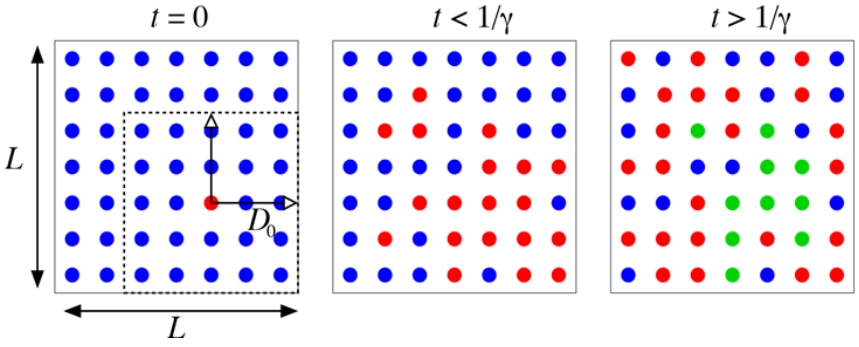
However, the time evolution of the SIR populations, as captured by the linear first-order differential equations of the Kermack-McKendrick model, produce behavior that is much simpler than the actual reported data of infections. Therefore, more detailed (microscopic) models of how the disease is spread from one infected individual to others are required to produce features that can emulate real data. In this study, we consider the simplest possible microscopic model to motivate the reasons that underlie the common features of real data. We then use these results to propose a simple analytical model for fitting the data with a few parameters. Finally, we use the results of the fitting to draw some insights on the actual evolution of the disease in representative countries.

Methods

The Microscopic Agent-Based Model

To understand the dynamics of the epidemic in more detail, we use a more detailed model based on individual agents, which are also classified as susceptible S, infected/infectious I, and removed/recovered R, that exist on a 2D regular grid of points. Each of these agents starts as susceptible and can be infected by another infectious with probability β per unit time (which we take here to be 1 day), and once infected can infect other individuals within a range $\pm D_0$, as illustrated in Figure 1.

Figure 1. Illustration of the agent-based simulation model on a 2D grid of size $L \times L$; in this illustration L=7, while in our simulations we considered L=1000. At t=0 (left panel) all the population is susceptible (blue dots), except for one infected individual (red dot). For a period $t < 1/\gamma$ (middle panel) the original infected individual can infect others in a neighborhood within a range $\pm D_0$ in each direction (shown by the dashed square), and each of those infected individuals also infects others within a corresponding range. For $t > 1/\gamma$ (right panel) some of the infected individuals have recovered (green dots), depending on when each was infected.



The size of this range turns out to be a crucial quantity, justifying the quest for social distancing measures to contain or slow down the spread of the disease. After being infected, an agent remains infectious for a period of $1/\gamma$ days, at which point the agent is removed (recovered) from the infectious population and can neither be infected again (has acquired immunity) or infect

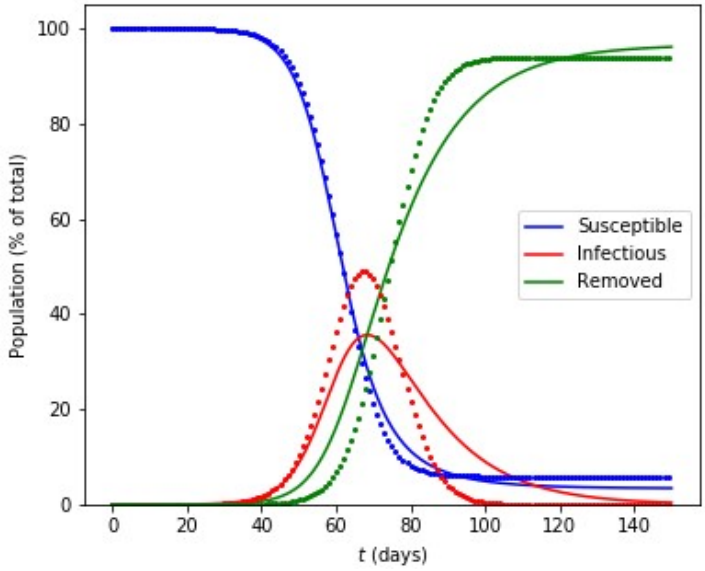
XSL•FO
RenderX

others. We suggest that the "microscopic" model is likely to be closer to the actual evolution of the disease than the continuous populations model represented by equation (1). The size of the 2D model, a square of length L in the example discussed below, corresponds to a small, uniformly populated "virtual country" of population $N = L \times L$.

We first provided a comparison of the numerical solution of the continuous SIR populations, represented by equation (1) and the agent-based simulations. We chose a total population of $N=10^6$ for both cases and a number of $I(0)=4$ infected agents, distributed randomly on the 2D grid in the case of the agent-based simulations. In Figure 2, we give a comparison of the evolution of the three populations as a function of time for a total period of 150 days, by which time the infectious population has been reduced to zero and the susceptible and removed populations have reached their long-term asymptotic values in both models. Although the overall behavior of the three populations in the two models is similar, the tail of the $I(t)$ population is quite "fatter" for the solution of the differential equations. The behavior of the tail is important, as it determines the rate at which the total number of infections (cumulative) grows with time, a subject of active research [17]. The value of the range D_0 in the simulations can be chosen at will up to $D_0=L/2$. The continuous SIR model *contains no information* on this range, which must be somehow included in the effective value of β. Keeping the value of β the same and adjusting the range D_0 and the initial condition for $I(0)$ in the continuous model, we can obtain reasonable agreement between the two models, as shown by an example in Figure 1. In this example, the evolution of the $S(t)$ population is captured well through the entire time range, except for the asymptotic value. This value is important, because it corresponds to the portion of the population that has not been infected at the end of the epidemic and is given by $N - R_{tot}$, where R_{tot} is the total number of removed; this is also equal to the time integral of the infected population divided by the mean period of infection, $1/\gamma$, as can be easily derived from equation (1c):



Figure 2. Comparison of the numerical solution to the susceptible-infectious-removed equations (solid lines) and agent-based simulations (points) for a population N. In both cases, we use $N=10^6$, β=.25 per day, and $1/\gamma = 14$ days. For the simulations, the range parameter is $D_0=50$ (see text for details).



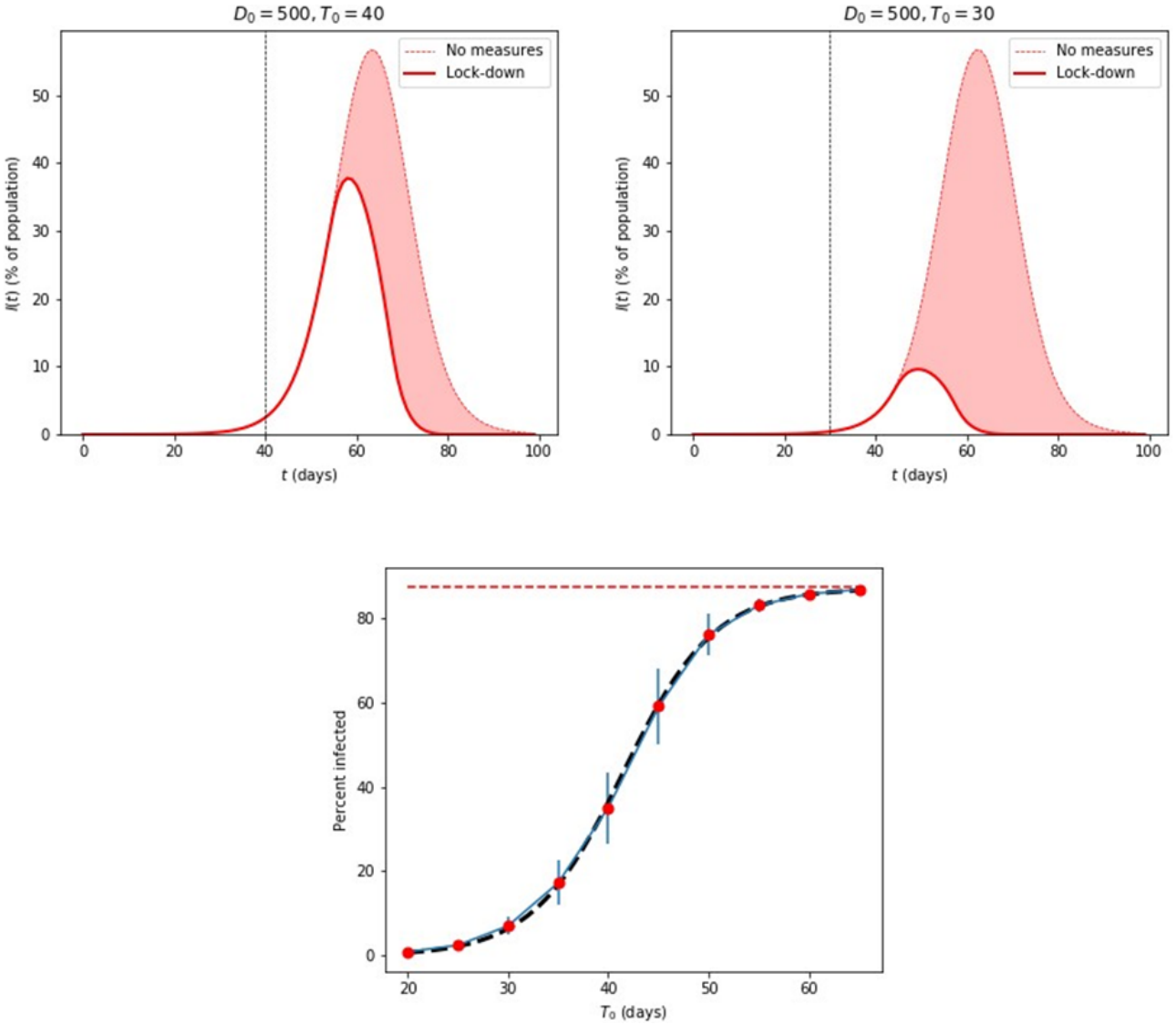
An important consideration in the dynamics of the disease is the effect of measures that restrict the movement of individuals in a population. This can easily be captured in the agent-based simulation model by taking a time-dependent value for the range that each infected individual has, namely:



where T_0 is the time at which the measures are imposed; both t and T_0 are measured from the time of the first infections, defined as $t=0$. Since $D(t) \to 0$ for $t \to \infty$ (assuming $\lambda>0$) the behavior of the range corresponds to "lock-down" measures in which individuals are restricted to a small range and eventually cannot infect anyone else (they are in "quarantine"). We use $\lambda=2.5$ days in our simulations, which means that from the moment that the measures are imposed ($t=T_0$), the initial range is reduced by a factor of ~20 for each week that passes. The effect of lockdown measures is quite dramatic, as shown in Figure 3. To provide a quantitative measure of this effect, we first let D_0 be the largest possible, $D_0 = L/2$ (half the size of the grid on which the agents live) and then consider several values of T_0, from 20 to 65, the last value being the time where the maximum of $I(t)$ occurs in the case of no imposition of restrictions, such as lockdown. A useful measure to quantify this effect is the total population of infected individuals over the course of the epidemic scaled by the mean period of infection $1/\gamma$, which is the same number as the total population recovered, see equation (2). This quantity, given as a percent of the total population, is shown in Figure 3 for the whole range of T_0 values we considered.

XSL•FO
RenderX

Figure 3. Top: Epidemic disease simulations of the infected population, $I(t)$, as a percentage of the total population, using 1 million individual agents on a 2D grid, with different lockdown dates, T_0, after the initial cases at t=0. The shaded curve in the background corresponds to $I(t)$ with no lockdown measures. For these simulations, β=.25, $1/\gamma$ = 14 days, and λ=2.5 days. Bottom: The percent of the total infected population, $R_{tot}(T_0)$ for different values of T_0, ranging from 0.93%, for T_0=20 days, to 86.7%, for T_0=65 days. The error bars represent standard deviations from averaging over 30 samples in each case (for the largest and smallest values, the error bars are too small to be visible). The black dashed curve is the fit from equation (4). The red dashed line represents the value for no lockdown measures, which is 87.4%.



The asymptotic value, reached for T_0=65 is 86.7% (±0.33%), was almost equal to the value when no lockdown measures were imposed, 87.4% (±0.06%); this last value corresponds to the "herd immunity" limit for the parameters we have used in the present simulation. The smaller T_0, the lower R_{tot} is, reaching the value of 0.93% (±0.34%) for T_0=20. The behavior of these values is well approximated by the expression:



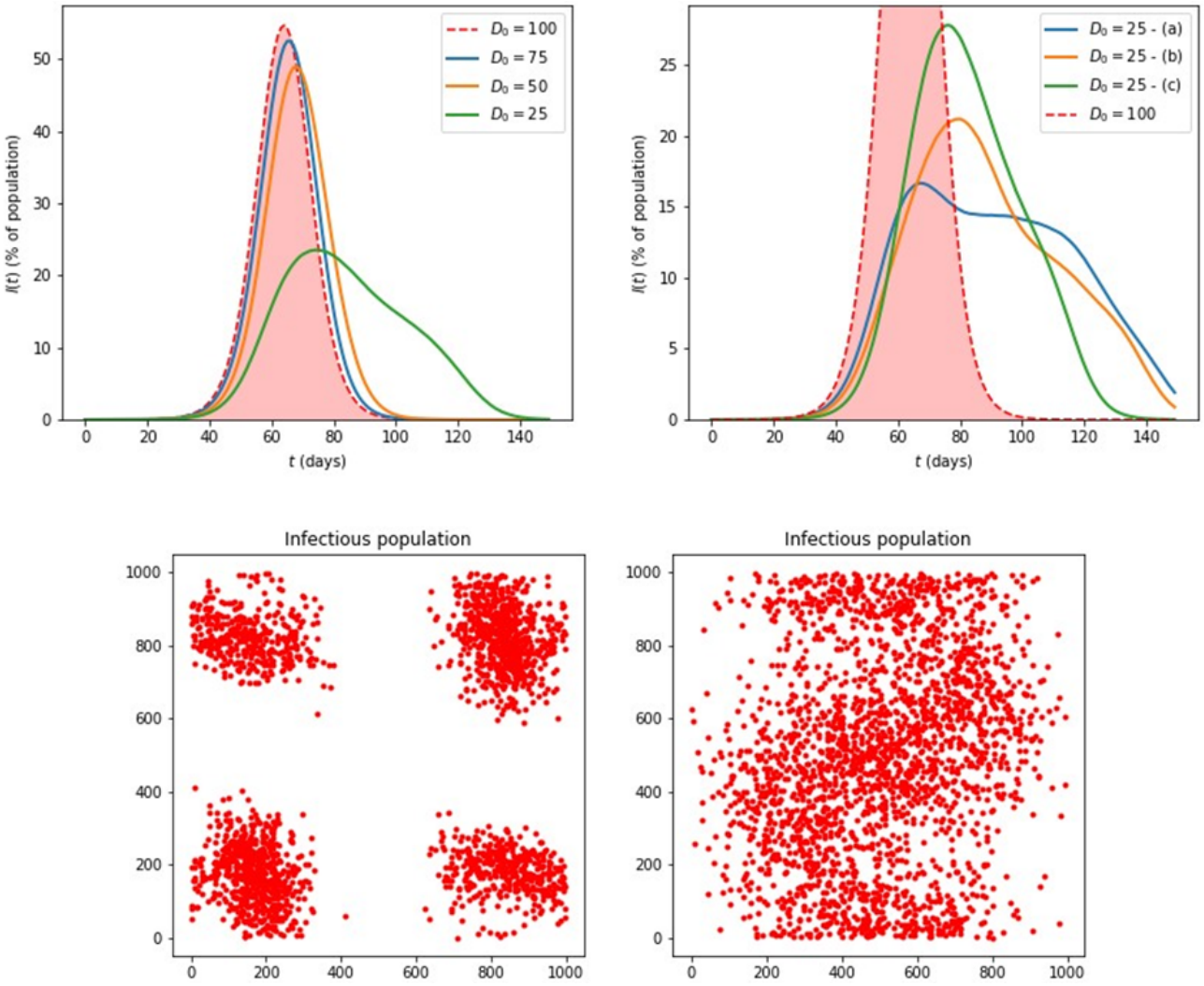
with I_0=43.7, τ=9.1 days, and T_1=41.5 days. A clear conclusion from this set of results is that the early imposition of measures makes a significant difference in the total infected population; for example, in a country with a total population of 10 million, the imposition of measures 20 days after the first few reported cases can reduce the total number of infected from 8.74 million to 93,000. The assumptions in this example involve allowing free movement of all persons for the entire period of the disease in the worst-case scenario to full quarantine within 2 weeks after imposing lockdown measures, which reduces the initial range spanning the entire country by a factor of 270, enough to essentially stop any disease transmission.

An interesting exercise is to consider what are the effects of *finite* D_0 much smaller than the size of the system (our "virtual country"). We give some examples of such simulations in Figure 4. For D_0=100 or larger, the result is essentially the same as that of the limiting case of free motion throughout the entire system, which was discussed in Figure 3 (background curves with no lockdown measures). For $D_0 \leq 75$, the curves start deviating from this behavior and even exhibit more interesting behavior, with additional "bumps" in the descending part and long tails, as the case D_0=25 shows. In fact, for D_0=25 the curves are not unique and depend on the initial random distribution of the infected individuals at t=0; we give an example in Figure 4. Within our simple microscopic model, this behavior arising from small

XSL•FO
RenderX

values of D_0 (like the case D_0=50 in Figure 4) can be easily explained; it corresponds to several small clusters of infections, which spread through the country in waves, as one cluster eventually becomes all removed, but before reaching this point, some infected individuals have moved to a region where there were no infections at all, starting a second wave of infections. We emphasize that, independent of the initial conditions that determine the number and time of the subsequent waves' occurrence, for a given value of small D_0, the asymptotic value of R_{tot} for $t\to\infty$ is always the same and of course depends on the specific value of D_0. This simply means that the total number of infections is determined by the effective range of interactions and, given enough time, the disease will infect the same total number of agents no matter how it proceeds from one cluster to others. This behavior of disease evolution may be actually close to what is seen in reported data, as described in the following section.

Figure 4. Top row: Simulations based on individual agents with a finite range of spreading the disease, D_0. Left: various values of D_0; the case D_0=100 already approaches the limit of free motion throughout the system, D_0=L/2, see also Figure 3. Right: several instances of D_0=25, labeled (a), (b), (c), exhibiting multiple bumps and long tails. Bottom row: the actual infected populations at the same time moment t=30 days, for D_0=50 (left) and D_0=100 (right).



In contrast to the situation for small values of D_0, for larger values of D_0 all the curves for $I(t)$ are identical and the behavior is that of one wave smoothly spreading through the entire system, as shown also by an example (D_0=100) in Figure 4. The critical value separating these two regimes is between 100 and 75 in our simulations. It would be interesting to study what this critical value is and what is its relation to the system size L in a more systematic manner using methods from statistical physics.

The Multiple Wave Forced-SIR Model

Recently, the forced-SIR (FSIR) model was proposed by the authors and used to describe the evolution of the COVID-19 pandemic in a representative set of countries [18]. This model treated the evolution of the infected population as a single wave (single peak wave). It contains three adjustable parameters that are estimated for each country by fitting actual data. However, the single-wave assumption cannot explain the entire incidence curve (infected population curve) in each country. Wavy patterns ("bumps") are evident in the actual data for many countries, which cannot be attributed to simply random fluctuations, due

XSL•FO
RenderX

to their regularity and similarity among several countries in which, at first glance, the disease was at different levels of severity. Here, we expand the analytical FSIR model to capture the multiple waves (subepidemics), underlying a country's overall incidence curve. This is akin to the case of finite D_0 discussed previously in the agent-based model. We applied this multiple wave analytical model to a representative set of 18 countries, in all of which the behavior of the infection as a function of time is accurately represented by our model.

In its original version, FSIR applies to a single epidemic wave, in which the infected population is given by the expression:

$$I(t) = N - S(t) - R(t) \quad \textbf{(5)}$$

with the approximate solution given by:



where N', α_1, α_2, t_1, t_2 are treated as adjustable parameters, with t_1 and t_2 representing the times at which the  and  populations reach their sigmoid midpoint values, respectively. Here, we extend this model to allow for multiple waves that capture the subepidemics in the infected population of a country. As argued in the agent-based microscopic model previously presented, several clusters of infections can appear in a country in waves; each cluster eventually becomes all removed, but before reaching this point, some infected individual has moved to a region where there were no infections at all, starting a second local wave of infections. In the extended model, we assume each wave is captured by a function described by equation (6), with different values of the parameters involved.

We apply this extended model to fit the multiple wave behavior of infected populations in different countries, as obtained from the European Centre for Disease Prevention and Control [19], for a period ending on May 16, 2020, which corresponds to 120 days from the onset of the exponential growth of reported cases in China. To obtain a meaningful fit, we had to consider data for each country that showed a monotonic increase at the beginning. This means that a few data points in each case were excluded, as they corresponded to sporadic reports of a few isolated cases, typically 1-10 in a given day, interspersed by several days of zero cases. In practice this means that the fitting begins at a certain cutoff day denoted as t_0.

As in the case of the original FSIR model, to make the fit more robust and simpler, we chose the α_1 and α_2 parameters to have the same value $\alpha_1=\alpha_2=\alpha=0.25$. We have found this to be the optimal value for the countries we considered. Moreover, a common value for the exponential decrease of the susceptible population, which is captured by the value of α_1, and for the exponential increase of the removed population, which is captured by the value of α_2, is actually more consistent with the agent-based simulations, as described in Section II.

Finally, instead of using t_1 and t_2 for each wave as independent parameters, we elected to use as independent parameters t_1 and $\Delta t = t_2 - t_1$. To make the multiple-wave fit more robust, simpler, and systematic, we chose Δt =14 (days) for all waves, which is a reasonable choice, as it corresponds to a common mean time period of 14 days before the infected individual is removed (Δt = 1/γ in the agent-based microscopic model). This mean time-period has been imposed as a quarantine measure for the majority of countries imposing measures (interventions) and is consistent with a reported estimated median time of approximately 2 weeks from onset to clinical recovery for mild cases [20]. This condition leaves two adjustable parameters per subepidemic that can be varied to obtain the best fit to the data, namely the onset time t_1, which corresponds to the midpoint of the sigmoid representing the decline of the susceptible population, and N', which is a parameter representative of the number of daily cases near the peak of the infected population curve in the given wave. N_T, the total number of infected in the given wave, can readily be obtained. The best fit here is defined in the Root-Mean-Square (RMS) sense. The model parameters were determined by employing the Levenberg-Marquardt algorithm.

Results

Application to Representative Countries

We were able to obtain reasonable fits for over 30 countries from the entire database [19], primarily selecting countries for which the temporal COVID-19 evolution had reached peak intensity of the infection. Rather than including over 30 countries in the following discussion, we have chosen to focus on three groups, a total of 18 countries, that span the whole range of parameter values and could hopefully provide some insight to the multiple wave behavior of the pandemic. The choice of the 18 countries also aimed to represent parts of the world more heavily or less heavily impacted by the disease, as well as more typical cases. Here we defined the impact as the total number N_T of infected individuals during the first 120 days of the pandemic, as predicted by the FSIR model [18]; this number is scaled by the population of the country (N_P). In particular, we have included six countries in which the impact was small, China, Australia, Greece, Cyprus, Tunisia, and Japan for which (N_T/N_P)<1000 infected per million; six countries in which the impact was moderate, Israel, Denmark, Germany, France, Canada, and Portugal for which 1000<(N_T/N_P)<3000 infected per million; and six countries in which the impact was large, Sweden, Switzerland, United Kingdom, Italy, the United States, and Spain for which (N_T/N_P)>3000 infected per million.

We fit 7-day running averages of the daily data, for all countries, with data up to May 16, 2020. For each country, we estimated the number of waves (subepidemics) in which the infected population curve could be analyzed, the model parameters of each subepidemic, and the expected number of cases for the first major wave ($N_T^{(1)}$) and for all waves (N_T). Table 1 presents the model parameters for the countries in our set.

XSL•FO
RenderX

Table 1. The values of the various parameters that enter in the multi-wave forced-susceptible-infectious-removed (FSIR) model of equation (6), for the representative countries considered.[a]

Code	Country	$t_1^{(1)}$ (days)	$N'^{(1)}$	$t_1^{(2)}$ (days)	$N'^{(2)}$	$t_1^{(3)}$ (days)	$N'^{(3)}$	$t_1^{(4)}$ (days)	$N'^{(4)}$	$N_T^{(1)}$	N_T[b]
CHN	China	17.3	5868	N/A[c]	N/A	N/A	N/A	N/A	N/A	N/A	81,991
AUS	Australia	21.5	484	N/A	N/A	N/A	N/A	N/A	N/A	N/A	6770
GRC	Greece	6.21	44	21.5	112	44.5	39	N/A	N/A	2173	2707
CYP	Cyprus	9.0	49	30.7	11	N/A	N/A	N/A	N/A	677	821
TUN	Tunisia	9.7	50	21.7	23	N/A	N/A	N/A	N/A	686	1014
JPN	Japan	68.4	562	74.5	262	N/A	N/A	N/A	N/A	7870	14,966
ISR	Israel	21.6	799	40.3	377	N/A	N/A	N/A	N/A	11,179	16,453
DNK	Denmark	7.4	106	30.1	404	50.4	195	65.9	118	7076	11,467
FRA	France	26.8	5531	41.4	3309	62.2	1436	N/A	N/A	77,439	141,057
DEU	Germany	23.5	4937	33.5	4411	49.3	2328	69.1	1140	69,078	179,379
CHE	Switzerland	21.0	1368	34.4	615	52.5	186	N/A	N/A	19,128	30,365
PRT	Portugal	22.3	1066	39.6	716	60.9	381	N/A	N/A	14,911	30,281
CAN	Canada	30.6	1662	48.2	1921	63.3	1891	N/A	N/A	23,267	76,617
SWE	Sweden	12.6	143	32.4	650	50.5	776	69.3	754	11,081	32,476
ITA	Italy	22.9	6875	37.1	4062	51.5	3404	67.4	1583	96,243	222,918
GBR	United Kingdom	30.0	4744	43.6	5879	61.7	6630	N/A	N/A	66,414	241,489
ESP	Spain	25.1	10,389	40.5	3824	52.9	1639	74.1	1737	145,378	246,080
USA	United States	31.3	39,205	47.8	33,059	64.0	32,818	N/A	N/A	548,817	1,470,776

[a]The ordering of the countries is discussed in Table 2.

[b]The last column includes the values for the *expected* total number of cases N_T when the number of infections has dropped to near zero and is an *extrapolated* value.

[c]N/A: Not applicable, as there is no relevant wave (subepidemic) for the respective parameter value to be obtained.

In the following, we present results obtained by the multiple-wave FSIR model for selected countries that can be accurately fitted by 4 waves (Italy, Sweden), 3 waves (United States, Portugal, Greece), and a single wave (China). The countries selected fall in two distinct classes: the first class comprises countries, which implemented stringent intervention measures rather fast; the second class comprises classes that implemented measures at rather later times and not at a high stringency level. The stringency of the measures is tracked daily by the Oxford COVID-19 Government Response Tracker (OxCGRT) [21], which systematically collects information on several different common policy responses governments have taken, scores the stringency of such measures, and aggregates these scores into a common Stringency Index. OxCGRT collects publicly available information on 17 indicators of government responses, that is, eight policy indicators recording information on containment and closure policies such as school closures and restrictions in movement, four indicators recording economic policies such as income support to citizens or provision of foreign aid, and five indicators recording health system policies. Italy, Portugal, Greece, and China had implemented high stringency measures rather fast, whereas Sweden, the United Kingdom, and the United States had not done so at that level.

Figure 5 presents the multiple wave fit for Italy, which was one of the most heavily impacted countries by COVID-19. An initial large subepidemic was followed by 3 declining subepidemics. The use of the term “declining” (or its opposite, “increasing”) refers to the peak intensity of the subepidemic. Italy has taken strong intervention measures, since the country’s maximum stringency level was 94.29 [21]. The shape of the curve is reminiscent of the shape of the curve produced by the agent-based microscopic model, Figure 4 for D_0=25. There is an excellent agreement between the 4-wave fit and the actual data in both daily and cumulative data. As can be seen, the single wave fit of the data, depicted by the green dashed lines, significantly underfits the data.

XSL•FO
RenderX

Figure 5. Results for ITA and PRT, obtained by fitting the multiple-wave forced-susceptible-infectious-removed (FSIR) model with data up to May 16, 2020. Top row: Red dots are the daily data reported by the European Centre for Disease Prevention and Control. The blue dots are 7-day running averages of the daily data. The green dashed line is the fit by the single-wave FSIR model. The black solid line is the 4-wave fit by the multiple-wave FSIR model. Middle row: Decomposition of the 7-day running average data (blue dots) in 4 waves for ITA and 3 waves for PRT. The black line represents the superposition of the multiple waves. The fit is in excellent agreement with the actual data. Bottom row: Blue dots are cumulative daily data (7-day running averages). The black line is the fit by the multiple-wave FSIR model, and it is essentially indistinguishable from the actual data. The green dashed line is the fit of the single-wave FSIR model, which clearly underfits the actual data. ITA: Italy; PRT: Portugal.

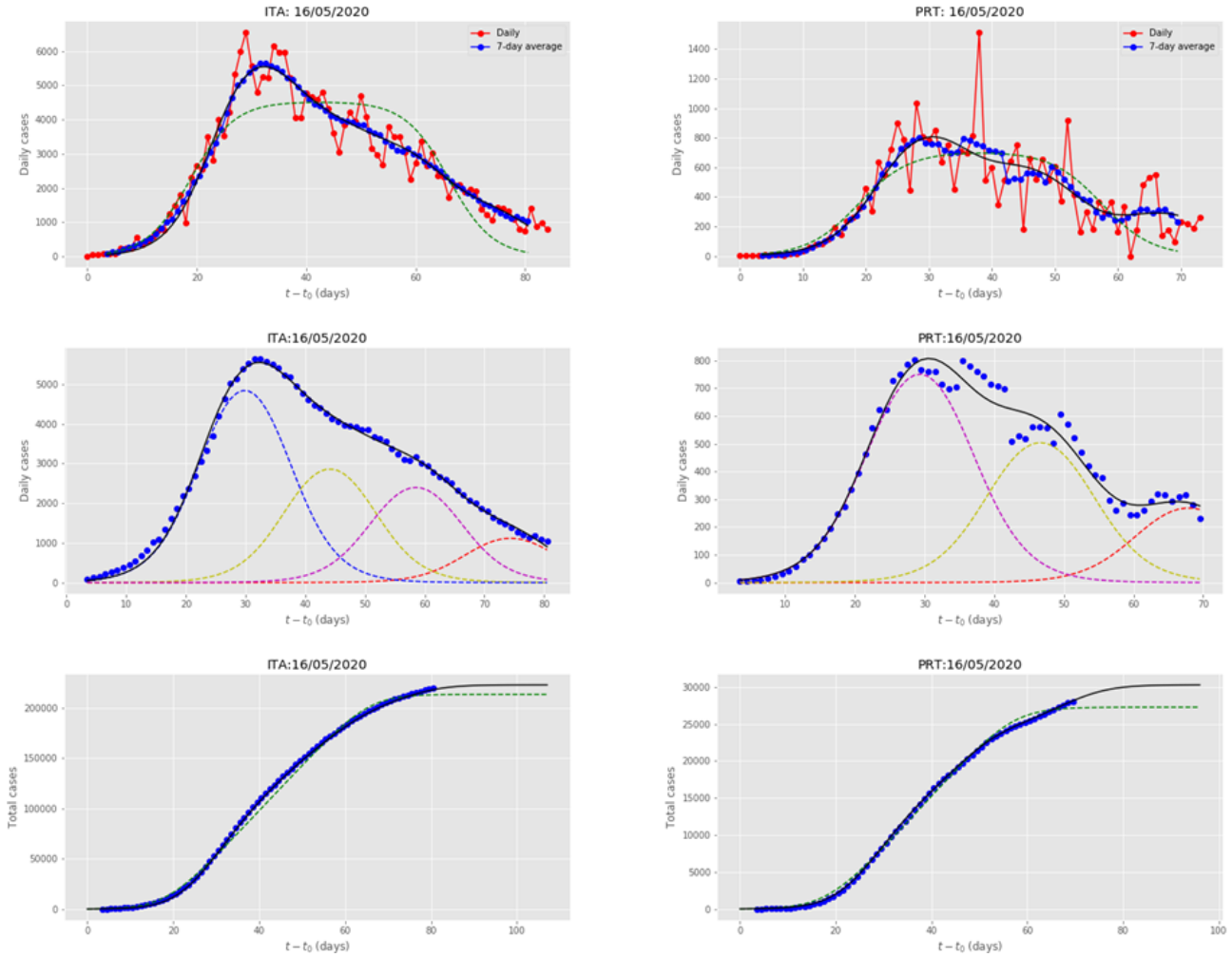


Figure 5 also presents the multiple wave fit for Portugal, which is a country experiencing a heavy impact by COVID-19. The country's government has implemented stringent measures, with the highest stringency level at 89.52 [21]. The country's incidence curve was fitted by 3 waves. An initial large subepidemic was followed by 2 declining subepidemics. Here too, the shape of the curve is reminiscent of the shape of the curve produced by the agent-based microscopic model, Figure 4 for D_0=25. There is an excellent agreement between the 3-wave fit and the actual data in both daily and cumulative data. As can be seen, the 1-wave fit of the data significantly underfits the actual data. Italy and Portugal are representative examples of countries where the initial major wave is followed by several waves of *declining* strength, suggesting that, despite the initial large impact, the countries were successful in eventually containing the epidemic. Germany, France, Spain, Switzerland, Denmark, and Spain are exhibiting similar behavior, namely that of a major initial wave followed by several of declining strength.

Figure 6 presents the multiple wave fit for the United States, which appears to be the hardest hit country by COVID-19, in terms of total number of cases. The country implemented a series of intervention measures to stop the disease's transmission and impact, which were deemed not to be taken aggressively enough, with highest stringency level 73.57 [21]. The country's incidence curve was fitted by 3 waves. An initial large subepidemic was followed by subepidemics, with seemingly declining strength. However, the United States is a country comprising of more than 50 states and territories, and it is not clear if additional waves, possibly of strength comparable to the original ones, may materialize or not at later times. A study of decomposing the United States infected population curve per state is currently under way by the authors.

XSL•FO
RenderX

Figure 6. Results for the USA and SWE, obtained by fitting the multiple-wave forced-susceptible-infectious-removed model with data up to May 16, 2020. The meaning of symbols is the same as in Figure 5. SWE: Sweden; USA: United States.

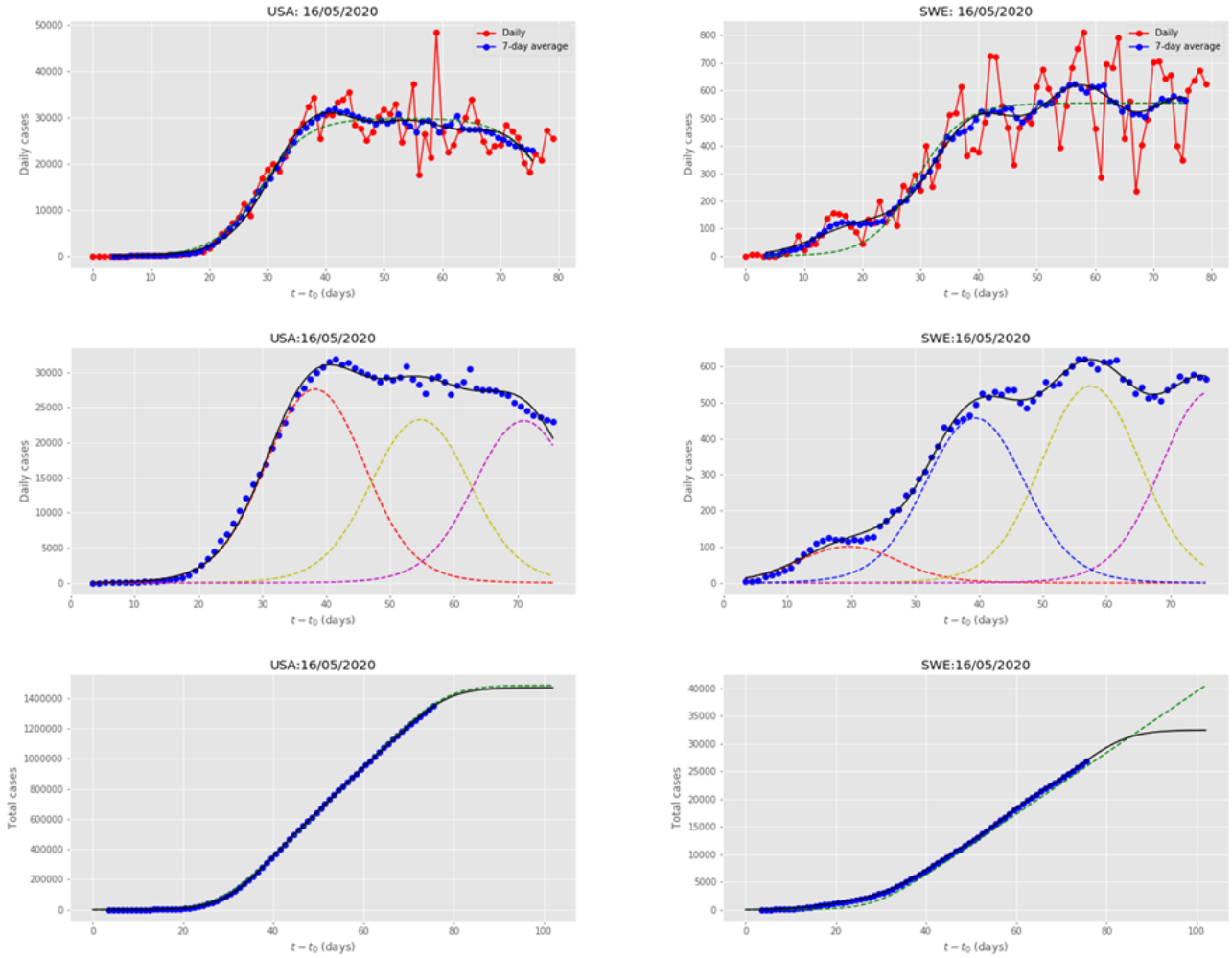


Figure 6 also presents the multiple wave fit for Sweden, which was also one of the hardest hit countries by the disease. However, the Swedish Government decided not to impose strict intervention measures but to inform the citizens to adopt certain precautionary measures, in a mostly individualistic capacity; the country's maximum stringency level was 58.10 [21]. Thus, the country followed a different mitigation policy, with respect to the rest of Europe and most of the world. Sweden's incidence curve was fitted by 4 waves. An initial small subepidemic was followed by 3 increasing subepidemics. It seems that the disease is spreading in waves; once a cluster of infected people is all removed, another bigger one is getting infected. Thus, the adoption of voluntary policy makes multiple nondeclining subepidemics of the disease get hold of the country. Since there was no clear trend as of May 16, 2020, of the country getting over the intensity peak, the 1-wave fit predicts a linear increase of the total number of expected cases. The 4-wave fit estimates a plateau of the total number of cases after the fourth wave, assuming that more waves do not materialize. According to the taxonomy of epidemic waves [13], COVID-19 in Sweden has generated an endemic wave; it remains to be determined if this is stationary or temporary.

Canada and the United Kingdom are among the countries exhibiting a similar subepidemics pattern, that of being impacted by a major wave followed by several waves of nondeclining strength.

Figure 7 presents the single wave evolution of the disease in China, which was the first country to be hit by COVID-19, and the Chinese Government implemented a series of rather fast and strict intervention measures to stop the disease's transmission and impact. The country's incidence curve was fitted by a single wave.

XSL•FO
RenderX

Figure 7. Results for CHN and GRC, obtained by fitting the multiple-wave forced-susceptible-infectious-removed (FSIR) model with data up to May 16, 2020. The meaning of symbols is the same as in Figure 5. CHN: China; GRC: Greece.

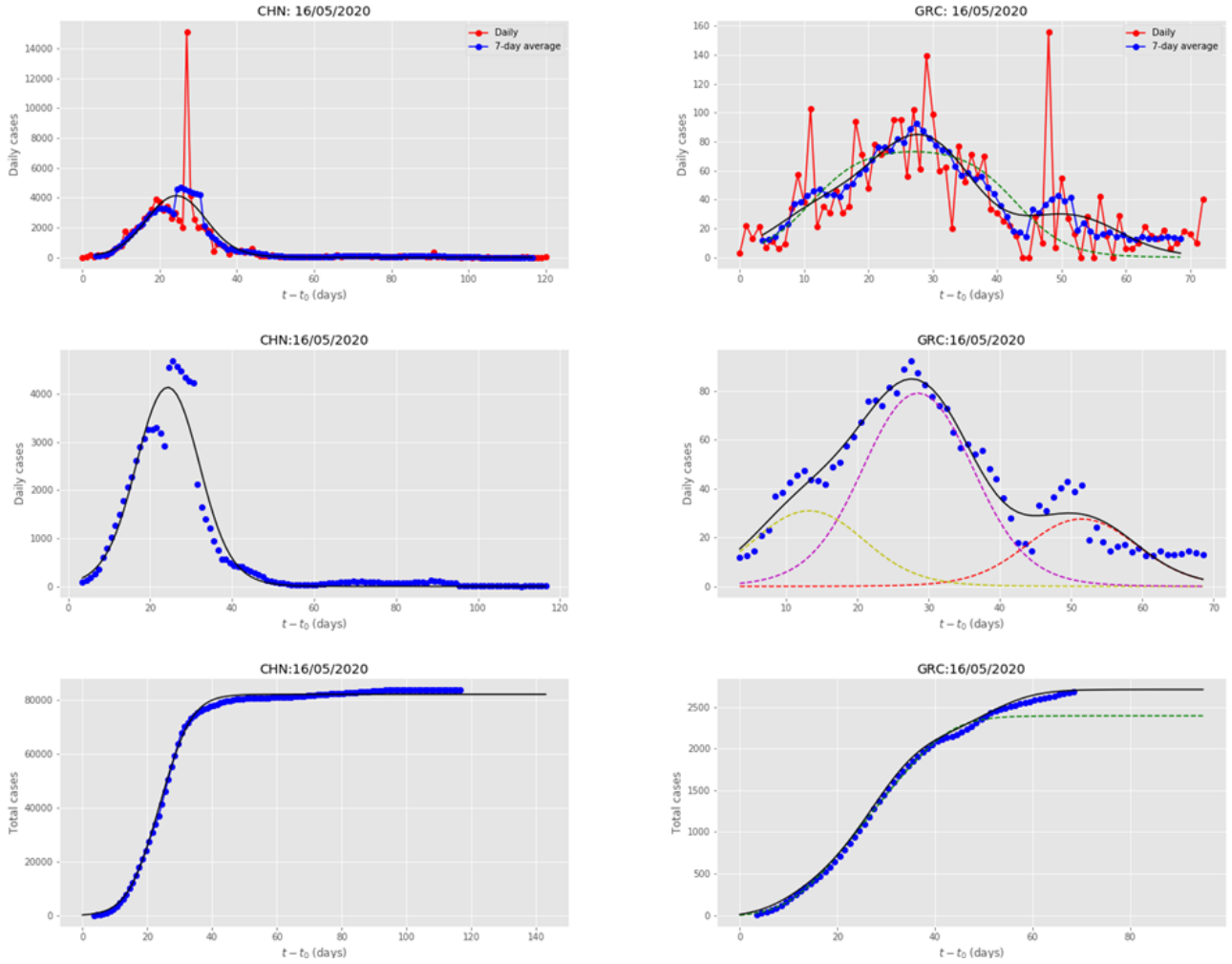


Figure 7 also presents the multiple-wave fit for Greece, whose government quickly implemented a series of intervention measures to stop the disease's transmission and impact, with highest stringency level 85.95 [21], keeping the total number of cases at very low levels. The country's incidence curve was fitted by 3 waves. An initial small subepidemic was followed by a larger one, which was followed by one of declining strength, suggesting that it contained the epidemic efficiently and fast. Greece is a representative example of countries, such as Sweden and Denmark, which also exhibit an initial small wave followed by a larger one. However, Greece and Denmark countered the epidemic before reaching high levels of cases. In terms of declining subepidemics, Greece follows the pattern of countries such as Japan, Israel, Cyprus, and Tunisia, in which the initial major wave was followed by a single subepidemic of *declining* strength.

Similar to China, the imposition of strict measures in countries such as Australia and New Zealand, shows that countries were able to reduce the disease's impact to a single wave. A quantitative estimate of the gain obtained is presented in the next-to-last column of Table 1, which presents the number of cases that could had been saved if countries had reduced the epidemic to a single wave rather than experiencing multiple ones; up to two-thirds of the total number of infections, as in the case of the United States, could have been avoided. Recent findings on the differential effects of intervention timing on COVID-19 spread in the United States [22] strongly corroborate this picture.

The Pandemic Response Index

Countries respond to the pandemic in varied ways. It is an interesting question to quantify their varied response and make comparisons, which may be useful for contributing to the evaluation of the different policies followed. Based on the results of our model, it is possible to construct an index, the Pandemic Response Index (PRI), and assign a value to each country depending on its response to the pandemic.

To do this in an objective manner, we took into account two factors. First, the total number of infections as given by the quantity N_T of Table 1, divided by the population of the country N_P. The range of this quantity when multiplied by 100 is between 0 and 0.5 approximately. This is a measure of the overall impact of the pandemic on the population of the country, and as argued by our microscopic model, it is a reflection of how early measures to contain the epidemic were imposed. The second quantity we considered is ΔN_T, the number of cases that correspond to all the waves except for the first major one, which in some cases includes the earliest small wave (see Table 2).

XSL•FO
RenderX

Table 2. Ranking of the various countries according to the PRI, defined in equation (7).

Code	Country	ΔN_T (/million)[a]	N_T (/million)[b]	PRI[c]
CHN	China	0	59	9.94
AUS	Australia	0	271	9.73
GRC	Greece	50	252	8.76
CYP	Cyprus	121	690	8.44
TUN	Tunisia	28	88	8.32
JPN	Japan	56	118	7.51
ISR	Israel	594	1852	6.55
DNK	Denmark	757	1978	6.11
FRA	France	950	2106	5.64
DEU	Germany	1330	2163	4.76
CHE	Switzerland	1319	3565	4.59
PRT	Portugal	1495	2945	4.58
CAN	Canada	1440	2067	4.45
SWE	Sweden	2101	3189	3.52
ITA	Italy	2096	3689	3.47
GBR	United Kingdom	2633	3632	2.74
ESP	Spain	2155	5267	2.69
USA	United States	2818	4495	2.37

[a]ΔN_T is the difference between N_T and the total number of cases that were infected by the first major subepidemic (for Sweden, Denmark, and Greece, both the small initial wave and the second wave have been taken into account). ΔN_T has also been normalized by the country's population in millions.

[b]N_T is the asymptotic value after all waves have decayed, given in Table 1, normalized here by the country's population in millions.

[c]PRI: Pandemic Response Index.

Arguably, this number of cases could have been avoided had the country imposed early and strict measures after the first wave of the epidemic was plainly evident; this was the case of single wave countries, for instance, China and Australia for which ΔN_T=0. The larger this number is, the worse the performance of the country. This number, divided by $2N_T$, lies in the range 0-0.5. With these two quantities, we then define the "Pandemic Response Index" as:



a quantity that lies in the range from 0-10, the higher values corresponding to better performance. This provides a quantitative and objective way of ranking the countries according to their performance. The results of this comparison and the relevant numbers that enter in the evaluation of the PRI are given in Table 2. We note that the classification is consistent with our initial selection of the countries considered here, as being those on which the disease had greater impact as measured by the number of infections per million, with arbitrarily chosen cutoffs in the ranges (smaller than 1000, between 1000 and 3000, and larger than 3000 per million). The only country that changes category based on the PRI value is Switzerland, which is raised to higher performance (average, see Table 2); this is a result of the fact that, although Switzerland had a relatively large number of cases per million (3565 cases per million), most of those occurred in the first wave, leaving a rather small percentage for subsequent waves. This remark suggests that the PRI is indeed a finer tool for evaluating performance, rather than relying on crude categorizations like the one based on the number of infections per million with arbitrary cutoff values between categories.

Discussion

Principal Findings

Reported cases of COVID-19 infections in various countries show features that are both common and regular, which we interpret as successive waves of transmission. We present evidence for this interpretation, using both agent-based simulations and a multi-wave model to fit the infected population data for many countries and give representative examples. This evidence supports the hypothesis that the COVID-19 pandemic can be successfully modeled as a series of epidemic waves (subepidemics). We analyzed the data from 18 countries based on this hypothesis and present the relevant parameters of a simple analytical model that accurately represents the data. Based on this analysis, it is possible to infer to what extent the imposition of early social distancing measures has slowed the spread of the disease. This analysis provides an estimate of how much lower the number of infections could have been, if early and strict intervention measures had been

XSL•FO
RenderX

taken to stop the spread at the first wave, as actually happened for a handful of countries.

Comparison With Prior Work

Recent works have emphasized more realistic approaches of human behavior and mobility involving larger transmission jumps by incorporating power-law decay of spatial interaction among human contacts [16], punctuated outbreaks as the disease progresses from one community to the next [12], and border effects [23]. In this study, our agent-based simulations start with 4 country-wide initial seeds for the disease onset in the "virtual country," thus approximating in a reasonable way the effect of longer jumps before the imposition of the intervention measures such as social distancing and lockdown. Nevertheless, the microscopic model can be extended to encompass network and community structures as well as border effects by incorporating weighted interactions among the agents in the simulation grid.

The multiple-wave FSIR model can identify multiple waves (subepidemics), specifying *each one* by only three parameters, t_1, Δt, and N', all of which are obtained by directly fitting the reported data of daily populations of infected individuals. Each of these parameters can be assigned a physical meaning, which help quantify certain generally held views; a detailed discussion of the meaning of these parameters can be found in [18]. Moreover, the quantitative picture that emerges from the values of these parameters produces a rather accurate picture of the severity of the epidemic in the various countries, and the effect of the intervention measures if and when any were taken.

A limitation of the original FSIR model is that it provides the extrapolation to future cases of infection as only a *lower limit*; this point has been discussed in an elegant mathematical analysis of the data by Fokas et al [17], highlighting the need for the inclusion of nonlinear terms in the underlying differential equations to capture the slow rate of the infected population decay. This is evident in the countries that have long passed the peak of the reported cases; the tail does not asymptote to a constant value, as the sigmoid (logistic) model predicts, but the number actually keeps growing at a slow rate. The multiple-wave FSIR mitigates this limitation of the original FSIR model; by modeling more accurately the wavy behavior of the infected population curve, it can provide a better fit to the daily data and to the cumulative actual data, and a better estimate to the cumulative number of cases (N_T), as can be seen in all the cases we examined, see Figures 5 and 6.

Limitations

The multiple-wave FSIR model may suffer from the fact that the number of infections dies off exponentially as the last wave does, a feature that appears unrealistic according to several other models that attempt to capture the long-term behavior [24-29]. Another limitation relates to the fact that in many cases, when Δt is estimated as an adjustable parameter, it tends to provide an aggregate fit, that is, an initial large subepidemic tends to be followed by a longer in time and smaller in peak intensity averaged wave, which is the sum of smaller subepidemics. This wave can be characterized as a temporary endemic wave according to the taxonomy of [13]. To improve the resolution of the model and enable it to specify the underlying smaller subepidemics, an epidemiologically reasonable value for Δt is necessary. Furthermore, caution must be exercised in interpreting the subepidemics because they may constitute a superposition of even smaller ones, as in the case of the United States, a country comprising 50 states with varied responses to the epidemic.

Conclusions

Multiple waves of transmission during infectious disease epidemics represent a major public health challenge. Our agent-based simulations encompassing strong social distancing measures show epidemics with multiple wave structures. The analysis of reported data from 18 countries supports the hypothesis that the COVID-19 pandemic can be successfully modeled as a series of epidemic waves (subepidemics). The main strength of the simulations and the models developed and used in this work is the simplicity and the insight they offer on how the disease is transmitted in a country and on quantifying the effect of the intervention measures of the disease dynamics. Based on the model's results, the construction of a PRI provides a finer tool for evaluating each country's performance, instead of relying on crude categorizations like the one based on the number of infections per million with arbitrary cutoff values between categories.

Conflicts of Interest

None declared.

References

1. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]
2. Haushofer J, Metcalf CJE. Which interventions work best in a pandemic? Science 2020 Jun 05;368(6495):1063-1065. [doi: 10.1126/science.abb6144] [Medline: 32439658]
3. Leung K, Wu J, Liu D, Leung G. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. Lancet 2020 Apr;395(10233):1382-1393. [doi: 10.1016/s0140-6736(20)30746-7]

XSL•FO
RenderX

4. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477. [doi: 10.3201/eid2607.200282] [Medline: 32255761]
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207. [doi: 10.1056/nejmoa2001316]
6. Imai N, Dorigatti I, Cori A, Riley S, Ferguson NM. Estimating the potential total number of novel coronavirus cases in Wuhan City, China. Imperial London College 2020 Jan 17:A.
7. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971. [doi: 10.1056/nejmc2001468]
8. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
9. Kermack WG, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Roy Soc A 1997 Jan;115(772):700-721. [doi: 10.1098/rspa.1927.0118]
10. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
11. Klein B, LaRock T, McCabe S, Torres L, Friedland L, Privitera F, et al. Reshaping a nation: mobility, commuting, and contact patterns during the COVID-19 outbreak. Northeastern University - Network Science Institute Report 2020 May 11 [FREE Full text]
12. Hoen AG, Hladish TJ, Eggo RM, Lenczner M, Brownstein JS, Meyers LA. Epidemic wave dynamics attributable to urban community structure: a theoretical characterization of disease transmission in a large network. J Med Internet Res 2015 Jul 08;17(7):e169 [FREE Full text] [doi: 10.2196/jmir.3720] [Medline: 26156032]
13. Chowell G, Tariq A, Hyman J. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. BMC Med 2019 Aug 22;17(1):164 [FREE Full text] [doi: 10.1186/s12916-019-1406-6] [Medline: 31438953]
14. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science 2020 Jul 10;369(6500):eabb9789 [FREE Full text] [doi: 10.1126/science.abb9789] [Medline: 32414780]
15. Zhang M, Verbraeck A, Meng R, Chen B, Qiu X. Modeling spatial contacts for epidemic prediction in a large-scale artificial city. JASSS 2016;19(4):3 [FREE Full text] [doi: 10.18564/jasss.3148]
16. Meyer S, Held L. Power-law models for infectious disease spread. Ann Appl Stat 2014 Sep;8(3):1612-1639 [FREE Full text] [doi: 10.1214/14-aoas743]
17. Fokas AS, Dikaios N, Kastis GA. Predictive mathematical models for the number of individuals infected with COVID-19. medRxiv 2020:A. [doi: 10.1101/2020.05.02.20088591]
18. Kaxiras E, Neofotistos G, Angelaki E. The first 100 days: Modeling the evolution of the COVID-19 pandemic. Chaos, Solitons & Fractals 2020 Sep 30;138:110114. [doi: 10.1016/j.chaos.2020.110114]
19. COVID-19 situation update worldwide. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases [accessed 2020-05-16]
20. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization. 2020 Feb. URL: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf
21. Coronavirus Government Response Tracker. University of Oxford. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker
22. Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the United States. medRxiv 2020 May 20:A. [doi: 10.1101/2020.05.15.20103655] [Medline: 32511526]
23. Zlojutro A, Rey D, Gardner A. A decision-support framework to optimize border control for global outbreak mitigation. Sci Rep 2019 Feb 18;9(1):2216. [doi: 10.1038/s41598-019-38665-w] [Medline: 30778107]
24. COVIDAnalytics. URL: https://www.covidanalytics.io/projections;
25. COVID-19 Modeling. URL: https://covid19.gleamproject.org/
26. LANL COVID-19 Cases and Deaths Forecasts. URL: https://covid-19.bsvgateway.org/
27. Short-term forecasts of COVID-19 deaths in multiple countries. The Imperial College London. URL: https://mrc-ide.github.io/covid19-short-term-forecasts/index.html
28. Institute for Health Metrics and Evaluation. URL: http://www.healthdata.org
29. Forecasts of total deaths. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

Abbreviations

COVID-19: coronavirus disease
FSIR: forced-susceptible-infectious-removed

XSL•FO
RenderX

OxCGRT: Oxford COVID-19 Government Response Tracker
PRI: Pandemic Response Index
SIR: susceptible-infectious-removed

Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by A Fokas; comments to author 26.06.20; revised version received 01.07.20; accepted 19.07.20; published 30.07.20.

Please cite as:
Kaxiras E, Neofotistos G
Multiple Epidemic Wave Model of the COVID-19 Pandemic: Modeling Study
J Med Internet Res 2020;22(7):e20912
URL: http://www.jmir.org/2020/7/e20912/
doi:10.2196/20912
PMID:32692690

©Efthimios Kaxiras, Georgios Neofotistos. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 30.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX